The Caldron of Consciousness

Motivation, affect and self-organization

Edited by
Ralph D. Ellis and
Natika Newton

Advances in Consciousness Research

GEORGIA TECH LIBRARY
5 0671 01192718 0

THE CALDRON OF CONSCIOUSNESS

ADVANCES IN CONSCIOUSNESS RESEARCH

ADVANCES IN CONSCIOUSNESS RESEARCH provides a forum for scholars from different scientific disciplines and fields of knowledge who study consciousness in its multifaceted aspects. Thus the Series will include (but not be limited to) the various areas of cognitive science, including cognitive psychology, linguistics, brain science and philosophy. The orientation of the Series is toward developing new interdisciplinary and integrative approaches for the investigation, description and theory of consciousness, as well as the practical consequences of this research for the individual and society.
Series B: Research in Progress. Experimental, descriptive and clinical research in consciousness.

EDITOR

Maxim I. Stamenov
(*Bulgarian Academy of Sciences*)

EDITORIAL BOARD

David Chalmers (*University of Arizona*)
Gordon G. Globus (*University of California at Irvine*)
Ray Jackendoff (*Brandeis University*)
Christof Koch (*California Institute of Technology*)
Stephen Kosslyn (*Harvard University*)
Earl Mac Cormac (*Duke University*)
George Mandler (*University of California at San Diego*)
John R. Searle (*University of California at Berkeley*)
Petra Stoerig (*Universität Düsseldorf*)
Francisco Varela (*C.R.E.A., Ecole Polytechnique, Paris*)

Volume 16

Ralph D. Ellis and Natika Newton (eds)

The Caldron of Consciousness
Motivation, affect and self-organization – An anthology

THE CALDRON OF CONSCIOUSNESS

MOTIVATION, AFFECT AND SELF-ORGANIZATION – AN ANTHOLOGY

Edited by

RALPH D. ELLIS
Clark Atlanta University

NATIKA NEWTON
Nassau Community College

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

BF
311
.C148
2000

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences — Permanence of Paper for Printed Library Materials, ANSI Z39.48–1984.

Library of Congress Cataloging-in-Publication Data

The caldron of consciousness : motivation, affect and self-organization -- an anthology / edited by Ralph D. Ellis, Natika Newton.

p. cm. -- (Advances in consciousness research, ISSN 1381-589X ; v. 16)

Includes bibliographical references and index.

1. Consciousness. 2. Emotions 3. Motivation (Psychology) . 4. Self-organizing systems. I. Ellis, Ralph D. II. Newton, Natika III. Series.

BF311.C148 2000

153--dc21 00-033720

ISBN 90 272 5136 3 (Eur.) / 1 55619 196 0 (US) (Pbk.)

© 2000 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. • P.O.Box 75577 • 1070 AN Amsterdam • The Netherlands

John Benjamins North America • P.O.Box 27519 • Philadelphia PA 19118-0519 • USA

Table of Contents

List of Contributors vii

Introduction ix

PART I
The Centrality of Emotion

1. Integrating the Physiological and Phenomenological Dimensions of Affect and Motivation 3
Ralph D. Ellis

2. Affective Consciousness and the Instinctual Motor System: The Neural Sources of Sadness and Joy 27
Jaak Panksepp

3. Consciousness, Motivation, and Emotion: Biopsychological Reflections 55
Bill Faw

4. Conscious Emotion in a Dynamic System: How I Can Know How I Feel 91
Natika Newton

PART II
Toward an Ecological Science of the Affective Sphere

5. The 'Mind'/'Body' Problem and First-Person Process 109
Eugene T. Gendlin

6. Dissolving Differences: How to Understand the Competing Approaches to Human Emotion 119
Valerie Gray Hardcastle

7. The Effect of Motivation on the Stream of Consciousness: Generalizing from a Neurocomputational Model of Cingulo-frontal Circuits Controlling Saccadic Eye Movements 133
Marica Bernstein, Samantha Stiehl and John Bickle

8. Motivation and Emotion: An Interactive Process Model 161
Mark H. Bickhard

9. Mind, Brain, and Chaos 179
Nicholas Georgalis

PART III
Emotional Learning and Development

10. Child Development and the Regulation of Affect and Cognition in Consciousness: A View from Object Relations Theory 205
Peter Zachar

11. Emotions: The Fetters of Instincts and the Promise of Dynamic Systems 223
Gary Backhaus

12. Awareness of Emotions: A Neuropsychological Perspective 243
M. Peper

Index 271

List of Contributors

Gary Backhaus
Morgan State University
Department of Philosophy
BALTIMORE, MD 21209
USA

Marica Bernstein
East Carolina University
Brody School of Medicine
Department of Physiology
GREENVILLE, NC 27858–4353
USA

Mark H. Bickhard
Cognitive Science
Lehigh University
17, Memorial Drive East
BETHLEHEM, PA 18015–3068
USA

John Bickle
East Carolina University
Dept. of Philosophy and Program in Neuroscience
GREENVILLE, NC 27858–4353
USA

Ralph D. Ellis
Clark Atlanta University,
Dept. of Religion and Philosophy
223 James P. Brawley Drive at Fair St. SW
P.O. Box 81
ATLANTA, GA 30314
USA

Bill Faw
Brewton Parker College
Psychology Department
MT. VERNON, GA 30445
USA

Eugene T. Gendlin
The University of Chicago
Department of Psychology
5848 S. University Ave.
CHICAGO, IL 60637–1584
USA

Nicholas Georgalis
East Carolina University
Department of Philosophy
GREENVILLE, NC 27858–4353
USA

Valerie Gray Hardcastle
Department of Philosophy
Virginia Polytechnic Inst. and State University
BLACKSBURG, VA 24061–0126
USA

Natika Newton
Nassau Comm. College
Philosophy Department
GARDEN CITY, NY 11530
USA

Jaak Panksepp
Bowling Green State University
Dept. of Psychology
Memorial Foundation for Lost
Children
BOWLING GREEN, OH 43403
USA

Martin Peper
University of Freiburg
Department of Psychology
Niemensstr. 10
D 79085 FREIBURG
Germany

Samantha Stiehl
University of Rochester
Center for Visual Science
Dept. of Brain & Cognitive Science
ROCHESTER, NY 14627
USA

Peter Zachar
Auburn University Montgomery
Department of Psychology
MONTGOMERY, AL 36124
USA

Introduction

Cognitive scientists are just beginning to pay attention to motivation and emotion, a long-neglected dimension of consciousness. The attention began with the work of neurophysiologists such as Damasio, Luria, Posner, Pribram, and LeDoux, and the subject is still strongly associated with the neurosciences (for example, the Association for the Scientific Study of Consciousness has sponsored an electronic seminar on "Consciousness and Emotion," http://www.phil.vt.edu/-ASSC/esem).

This scientific interest in the study of motivation and emotion in consciousness may lead to more efficient and fruitful projects than those associated with the exploration of consciousness as a mere appendage to the more traditionally understood areas of information processing, perception, and intellectual cognition. The role of consciousness was relegated primarily to philosophy for several decades before the seminal work of Crick and Koch in the 1990s (e.g., see Crick and Koch 1990). Now consciousness is the center of a major multidisciplinary program, which has in turn brought emotion into the spotlight.

Our goal in this anthology is to bring together thinkers in various disciplines who are in the vanguard of emotion studies — including physiological, psychological and philosophical approaches to motivation and all the affective processes — with the goal of providing a coherent multidisciplinary framework for understanding emotion and its contribution to the workings of consciousness in general. The organizing premise of this volume is that consciousness and emotion are aspects of the embodied subject as a self-organizing system. Understanding the role of emotion and motivation in a self-organizing process requires approaching it via philosophy, neuroscience, and psychology (both theoretical and clinical). This anthology includes prominent researchers in all of these areas. Because of the common theme of self-organization, the contributors do not talk past each other, but share a general view of consciousness and emotion as processes with central roles in maintaining the well-being of the organism. We believe that such an approach, which coordinates multiple viewpoints within a central framework, is essential to real progress in understanding emotion and consciousness.

1. Emotion — the most basic ingredient in all consciousness?

Would it be possible for a being with no 'emotions' in a broad sense — no motivational impulses — to experience any of the phenomenal aspects of consciousness, the 'what it's like' dimension? Many suggest not (e.g. Needleman 1968; Edelman 1989; Gendlin 1992; Cytowic 1993; Damasio 1994, 1999; Ellis 1995; LeDoux 1995; Panksepp 1998; Watt 1998; Ellis and Newton 1998). A definitive answer to this question would have extensive implications for the neuroscience and philosophy of conscious experience. To begin with, if motivation is a necessary element of all conscious experience, then distinguishing between conscious and non-conscious forms of information processing would hinge on understanding the role of motivating emotions in all modalities of conscious experience and thought, as well as the neurophysiological basis for emotion and the ways these brain processes interact with other brain processes more traditionally associated with cognition and information processing. Understanding these interactions could in turn lead to a better understanding of motivation and emotion themselves.

Recent trends in neuroscience, cognitive theory, and the theory of consciousness have suggested both an increased importance for the role of emotion, and an urgent need for a more sophisticated approach to understanding emotional processes themselves. In the realm of consciousness studies, it becomes increasingly obvious that a mere computational registering and processing of sensory signals (e.g. in the occipital lobe and area V5 for visual information) by itself does *not* result in perceptual *consciousness* of the relevant information (Posner and Rothbart 1992; Mack and Rock 1998). Instead, emotional processes orchestrated in subcortical structures such as the amygdala, the hypothalamus, and brain stem structures are crucial for phenomenal consciousness (Luria 1980; Posner and Rothbart 1992). In cases of complete frontal lobectomy, when attention is severely impaired, some remaining consciousness is made possible via subcortical control of neurotransmitters and by the achievement of various kinds of resonance between such activity and the intact posterior cortical activity; this allows consciousness rather than mere blindsight to result. Emotional processes are thus a necessary component for both agent-directed, attentive consciousness and for more primitive, attention-deficient or agency-deficient forms of consciousness.

The recently developing self-organization approach to mind and brain (e.g. Freeman 1987; Globus 1992; Mac Cormac and Stamenov 1997) and the growing agreement that consciousness and cognition cannot be understood independently of the body and world in which they are embedded and with which they interact (Thelen and Smith 1994; Clark 1996), especially call for a better understanding of emotion and motivation. These approaches suggest that a self-organizing system can actively appropriate, replace and reproduce the physical substratum

elements needed to maintain its multiply-realizable patterns of activity. Self-organization conceptualized in this way leads to a distinction between living and non-living things (Monod 1971; Kauffman 1993), or between 'purposeful' and 'nonpurposeful' phenomena, that does not rely upon question-begging teleological assumptions.

A self-organizational approach to living things is by now noncontroversial; taking the same approach specifically to the conscious states of a living organism is more novel and hence needs more detailed examination. Regardless of the way the details are worked out, this approach has obvious advantages. A self-organization model of consciousness is more consistent with the way we view other aspects of living organisms than is the traditional information-processing or computational model. The information-processing model has attempted to think of consciousness as a detached observer of information inputs, resulting in a difficulty in understanding how the consciousness of an image or idea differs from a mere physical replica of that which is imaged or thought, as in the Medieval 'causal theory of perception.' Information processing theory alone also yields very little help when it comes to understanding the purposeful, affective dimension of consciousness. On the self-organizing model, consciousness results from the action affordances of environmental conditions for a living organism with aims and purposes. Consciousness is a guide to action, and emotional valuation is a part of any conscious experience. The subjectively felt aspect of any conscious experience results from the process through which emotions, including curiosity, motivate the forming of anticipatory imagery as we actively scan the environment looking for items of motivational interest (Ellis 1995; Panksepp 1998; Ellis and Newton 1998).

Viewing consciousness as a type of organismic activity driven by emotion has several philosophical advantages. One is that if consciousness is a motivated activity, the infamous 'Knowledge Argument' (Jackson 1986) evaporates. The Knowledge Argument highlights the apparent puzzle of why, if physicalism is true, conscious experiences cannot be known by an external observer of the brain in the way that they can be known by the subject. But if consciousness is a motivated activity, the mystery disappears. While other people can perceive what I perceive, only I can perform my activities and experience being the agent of those activities.

There is another advantage. While many of the physical substrates of consciousness may involve widely distributed parallel processing, it is often noted that conscious states, as experienced, are serial rather than parallel. If consciousness is viewed as a motivated activity, this phenomenon is easier to explain. Agent-directed bodily actions are performed one step at a time; parts of the body are coordinated to achieve a single end; components of sensory mechanisms must cooperate to bring perceived objects into focus. If we think of consciousness as

involving the organization of parallel-processing subsystems to achieve a unified goal, then the serial nature of conscious experience follows naturally.

Emotions constitute the paradigmatic case of a phenomenon that is paradoxical when we try to fit it into a causal account modelled after inorganic physical causal processes. Nowhere is this paradox more prominent than in the direction of attention, which is difficult to coherently construe as either passive or active. We know that our emotions dictate the direction of our attention, yet we feel as if we are 'free agents,' and part of the action of this free agency is felt to be in the direction of conscious attention toward objects. At the same time we feel that our attention is 'pulled' by the objects, and that our consciousness of them is a passive causal reaction to forces external to us. Hence the traditional paradox emphasized by Kant, Searle (1984), Kim (1992) and many others, that the system of sufficient causes at the physical level, which our experience presents to us, seems inconsistent with the notion that consciousness itself has causal power — a notion that our experience also presents to us. I.e., physical causation seems to exclude the possibility of agency, yet we experience ourselves as agents, and in some respects (e.g. the direction of attention) our agency seems to entail causal power on the part of our own conscious being.

A dynamical systems approach offers a resolution of this paradox: Organismic interests and motivations, which determine the direction of attention, are aspects of a self-organizing tendency on the part of the organism, and the motivated selection for attention has already been established by this self-organizing tendency prior to the actual consciousness of the object (Luria 1980; Posner and Rothbart 1992; Ellis 1995). Thus we actually *feel* as if our attention were passively a reaction to the stimulus, yet also feel that it is within the control of our agency.

This latter causal chain, from motivation through attention to perception, is part of the structure of the perceiving brain. To be conscious of any stimulus, the organism must first be motivated to intensify the processing of certain sensory input at a preconscious stage. This initial motivation is not a part of conscious experience. Unconscious motivation can affect the direction of attention in ways that have evolved because such attentional patterns are geared toward the emotional needs of the organism. Paying attention involves cortico-thalamic loops (Crick and Koch 1990; Bernstein *et al.*, this volume) that first transmit sensory signals from the thalamus to sensory and association cortices; from there, signals return to the thalamus, functioning to enhance its response to selected input deemed important to the organism. It is only at this stage that the organism becomes conscious of the stimulus, which has thus already been categorized and evaluated in a preliminary way (Edelman 1989). If, for example, the stimulus signals danger (an emotional value), the amygdala, a part of the loop, sets in motion a fear response; it is at this stage that the emotional quality becomes part

of the awareness of the stimulus, and in fact largely determines the appearance of objects, not only by motivating attention to aspects of them that are emotionally important for the organism's purposes, but also by permeating the experience of the object with a 'felt' dimension which is inseparable from 'what it is like' consciously to experience that object (LeDoux 1996). Even though the emotion may not be equivalent with the felt quality that we experience in every conscious modality, it is still an indispensable contributor to it.

Our purpose here is to outline an approach to emotion and motivation that can help facilitate an understanding of (1) why the role of emotions, within a self-organizing process in interaction with certain information-processing functions, is crucial in distinguishing between conscious and unconscious processes; (2) how emotions fit with self-organizing models of mind and brain, and thus can account for the role of 'agency,' the active rather than merely reactive nature of organisms' information processing (Gibson 1988; Varela *et al.* 1991); and (3) how a theory of self-organization that includes an adequate emotional-motivational theory shows promise of resolving the traditional mind-body problem by means of a holistic understanding of the conscious, behaving organism.

2. Some basic definitions

What do we mean by the terms 'emotion' and 'motivation'? And how is it that, on the one hand, these phenomena can differ from the mere 'tendencies' that inorganic systems exhibit, while on the other hand they seem capable of occurring on a *less than conscious* basis in many if not all beings that are capable in general of having them also on a *conscious* basis? If we want to distinguish all these terms clearly from each other, there are four kinds of situations that need to be accounted for with them:

1. Self-organizing systems that do not act as *agents* of their own actions can have *tendencies* to sustain themselves by appropriating their own needed material substrata, but these 'tendencies' are not normally elevated to the status of *'intentions'* — i.e., intentional motivations or emotions. The lack of 'intentionality' seems to correlate with the organism's lack of representations (perceptual and imaginative) associated with any *aims or objects* to which its behavioral tendencies relate. E.g., *plants* seem to fall into this category (see Faw, this volume).
2. We need to distinguish these non-agent beings that have tendencies from the kinds of self-organizing systems that do act as *'agents'* in appropriating their own needed material substrata; the sense of 'agency' meant here is simply that the organism acts in a unified effort to meet its self-organizational needs, but not *consciously:* We say that these organisms are *motivated,* but not consciously

motivated. Their systems can execute plans of action, e.g., to move from point A to point B if there is more food at B, but they are not conscious of intending to do it for this reason. And there is some sort of rudimentary 'representation' or 'perception' associated with the motivated action, even if not a conscious representation (Newton, this volume and 1996). E.g., *insects,* and perhaps even *frogs,* might fall into this category (Merleau-Ponty 1942/1963; Faw this volume).

Georgalis (this volume) makes the important argument that unconscious nervous states are only *potentially* intentional, in the sense that a mere physical replica or isomorphism to the object, manifested in unconscious nervous activity, does not yet know what it is 'about' until we do bring it into awareness; it does not yet literally *refer* to its intentional object. However, such unconscious states are 'preconsciously' intentional in that they behave analogously to conscious intentional states, and their intentionality can be explored when we do consciously attend to them.

3. The difference between 'motivation' and 'emotion' seems in many usages to hinge on whether intentional objects are accessible to conscious awareness; 'emotion' is something that in principle we *can* feel, even if we do not actually feel it on every occasion. Thus the question also arises as to whether insects, humans, etc., can have *unconscious* 'emotions,' or whether humans can but insects cannot, or whether we completely reserve the term 'emotion' to refer to *conscious* motivations; we prefer not to do the latter, because we humans apparently can have emotions and be completely unconscious of what they are really 'about,' i.e. their aims and objects, and also what the quality of the emotion really is: E.g., I may think I am 'angry' about my son's not taking out the garbage, but it is not the case that if he had done it then I would feel fine; what may really be happening is that I am 'frustrated,' not angry — and not because of what my son did, but about the way my career or love life is going. So we want to leave room for *unconscious* emotions in this sense, and these are very important in understanding motivation, since more often than not our first kneejerk reaction about what our emotions really want us to *do* (their aims in relation to their objects) are very inaccurate, and to get more accurate requires getting more clear on 'what we really feel' and what the feeling really wants us to *do* about it.

For this reason, we want to reserve the right to talk about 'unconscious emotions' and not just define an emotion as a conscious motivation. We then need to clarify why there is a difference between an unconscious *motivation* and an unconscious *emotion,* in some such way as the following: A motivation that is incapable of becoming conscious in the particular organism in which it occurs is not an 'unconscious emotion.' A motivation can be an emotion only in an organism which, in principle, could be capable of being conscious that it feels so-and-so. This requires both (a) that the organism be a type of organism that is

capable of consciousness, and (b) that the particular motivation is of a kind that in principle could be consciously experienced.

So, for example, we would not say that a plant or even an insect has even 'unconscious' *emotions,* although the insect can have *motivations* (whereas the plant cannot). And we would not say that I have an 'unconscious emotion' that wants to regulate my blood sugar level, even though I am *motivated* to do so (because my system acts as a whole to regulate it, and will readjust its different systems and its behavioral activity as a whole in relation to perceived environmental conditions, for example food, if that is needed to facilitate the process). But, if I am frustrated and do not know it or cannot feel it, we *do* want to say that I have an 'unconscious emotion' of frustration. Or, for those who are really averse to speaking of emotions as unconscious, an alternative term in this context might be 'preconscious,' since the latter term implies that the emotion *could* become conscious if we were to engage in some kind of conscious process aimed to bring it into awareness. But it is still important here not to say that the conscious emotion is just the unconscious emotion *plus* a direction of conscious attention *to* it, because when we become conscious of something, we execute a vast complexity of neurophysiological processes, and this will undoubtedly *change* the physiological structure of the emotion itself. This is important for avoiding what Natsoulas (1993) calls an "appendage theory" of consciousness. I.e., consciousness is not just an extra layer that is tacked on to processes that could have occurred on an unconscious basis without their basic structure's being affected by whether they are conscious or unconscious.

4. Finally, if a self-organizing system is motivated to act in a unified way to appropriate the needed material substrata for its motivated activity, and if there is also some degree of accuracy in the organism's intentional representations of what the aims and the objects of the motivation are, then we say that this motivation is a *conscious* emotion. There is not just a perceptual representation *alongside* a motivational tendency, but also an awareness or feeling *that* some represented state of affairs is what the motivational tendency is 'about.' As Newton (this volume and 1996) discusses, we entertain proprioceptive and sensorimotor imagery of what it would be like to act in the way intended by a motivation. "Imagery of the bodily response to the most positive action plan affects the current bodily state, enhancing the activity of the subsystems that would be involved in carrying out the plan, and inhibiting others.... The entire organism mobilizes around the chosen plan of action" (Newton, this volume). But of course, even in this case, it is still possible for the subject to get still more clear on what it really wants than the initial emotion would lead us to believe. E.g., I may feel myself being 'all churned up,' but may not know at first whether my system is preparing for fight or for flight — and then if I realize that I only forgot to take my blood pressure medication, I may realize that there is no

emotion of either fear or anger, but merely an objectless and aimless physiological event, which really should not even be called an 'emotion' in the strict sense at all (since an emotion takes an aim and an object, as discussed in more detail in the Ellis contribution).

So there seem to be four levels that are possible in self-organizing systems: (1) Unified *tendencies* that are neither emotions nor motivations; (2) Motivations that are not associated with any emotions; (3) Motivations that are associated with unconscious emotions; and (4) Motivations that are associated with conscious emotions.

There is also another terminological problem: Some people reserve 'emotion' to refer only to *instinctual and hardwired* motivational affective feelings, and not to motivational affective feelings in which the aims and objects are not hardwired; whereas other people just lump them all under the term 'emotion.' So, e.g., suppose I feel a little 'disgusted' with a baseball team's manager because he apparently worked his ace relief pitcher so hard that he developed arm trouble: It seems obvious that I could not have been *hardwired* to instinctually feel *that* particular emotion about *that* particular issue (and when I reflect on the feeling, I might also find surprises — e.g., that I am projecting feelings toward my father onto the team's manager, etc.). So some people would call this feeling an 'emotion,' and others would not. Musicians routinely talk about 'emotional expression' in music, but the 'emotions' in question are hardly ever, and maybe never, the instinctual, hardwired ones. (One of the contributors to this volume, Gary Backhaus, will argue that, for adult humans, there are *no* specific feelings that are instinctually hardwired to inevitably go with specific given objects and aims.) Philosophers also are prone to use 'emotion' in the broader sense, as in A.J. Ayer's 'emotivist theory of ethics.'

Whether we limit 'emotion' to the narrower class or extend it to the broader class seems to be essentially a matter of terminological convention, and there is no way to resolve that difference of usage here; so the only thing we can do is to make sure and be clear as to which sense of the term we mean. The contributors to this anthology generally agree that it would be too artificial and simplistic to make a sharp distinction between instinctual, hardwired 'emotions' on the one hand, and more complex and social 'affects' on the other. Instead, the emotional life must be considered as a whole, with a holistic balance as an ultimate guiding principle that can be affected in various ways by different kinds of environmental conditions, whether consciously or unconsciously, and whether the conditions are specific or whether they are amorphous and difficult to attribute to specific objects.

3. The specific contributions

Part I — The Centrality of Emotion. The approach here is multidisciplinary. Our first goal is to clarify the complex intentionality of emotional experience, so that all disciplines studying affective experience will appreciate its subtlety. Thus Part I, "The Centrality of Emotion," presents evidence that all forms of consciousness are distinguished from unconscious information processing by virtue of the role of emotion in them. The contribution by **Ralph Ellis** stresses that, if the intentionality of emotional experience is conceptualized in a too simplistic way, then no physiological account, no matter how sophisticated, can be effectively correlated with such an oversimplified phenomenological account. The paper sets the task of elucidating this question of intentionality in a way that integrates the phenomenological *experiencing* of the affective dimension, in all its subtlety and richness, with the scientific study of the correlated biological processes. It is suggested that an 'enactive' approach to emotion, viewing the organism as a self-organizing system geared toward appropriating and replacing components needed for its own self-regulated patterns of activity, is a good basis on which to understand both the physiology and the phenomenology of the affective dimension. Ellis outlines a theme that is followed up with more detailed analyses by other contributors later in the volume: The concept of self-organization, as developed in biochemistry by Kauffman (1993) and Monod (1971), in philosophy and psychology by Merleau-Ponty (1942/1963) and in neuroscience by Walter Freeman (1987) and others, provides the key to integrating the phenomenological and physiological dimensions of affective experience, and thus of bridging the 'explanatory gap' between these domains.

The stage is then set for an integration of this perspective with neurophysiology. The contribution by neurophysiologist **Jaak Panksepp** outlines a much richer picture of the aims of emotions than the simplistic behaviorisms and consummatory reductionisms of the recent past. In Panksepp's view, many endogenous emotional tendencies are not learned through reinforcements associated with experiences of consummatory pleasure. Rather, evolution has selected for organisms such as mammals that also have innate tendencies toward exploration, play, nurturance, and social bonding, independently of any consummatory drive reducing value of such activities. In self-organizational terms, we could say that complex organisms seek not only homeostasis through their consummatory drives, but also to avoid entropy by means of curiosity, exploration, and other action tendencies which are just as primary as the homeostatic needs for food, water, and sex.

The paper by neuropsychologist **Bill Faw** is an important cornerstone of the volume, laying out the empirical 'state of the art' with respect to neurophysiological research that has revealed correlations and connections between conscious

experience and motivational-emotional processes in the brain on the one hand, and on the other the correlation of these subcortical areas with 'higher' brain processes that are more traditionally associated with consciousness. The upshot is that emotion is integral to all forms of consciousness, including the most abstract, intellectual, and seemingly passive forms.

Philosopher **Natika Newton** then pulls together all that has been said in Part I of the volume, to work toward a comprehensive theory of emotion that elucidates its intentionality as the determinative difference between conscious and unconscious affective processes, integrating this theory with neurophysiology by means of the concept of self-organization. Newton explains the way dynamical systems emerge from chemical systems at dynamic equilibrium, and shows how such systems can be organized around motivational purposes, arguing that only in systems organized in this way can the phenomenon of conscious awareness, or a 'what it's like' dimension of experiencing, arise. Consciousness is seen as an activity which an organism executes, rather than as a stimulus 'input' that it passively receives. If so, then the structure of information processing in conscious beings (i.e., those whose conscious activities are motivated by the emotional purposes of the organism) would be completely different from the structure of information processing in nonconscious systems; this would be an important implication from the standpoint of modelling artificial intelligence systems if such systems are to serve as a guide to the way humans process information. This would also explain the apparent empirical unobservability of conscious qualities by means of empirical-scientific methods of inquiry — i.e., the problem posed by Jackson's 'knowledge argument' as to why a neuroscientist who knows everything physical about my brain still would not know what my experiencing feels like to me. Newton argues that a process model can answer these questions by showing that one must be in the location of the organism that is executing an emotionally motivated conscious process in order to know what that process feels like to that organism; since no one can execute anyone else's organismic processes, each of us has a realm of 'private' access to our own consciousness that is not available to an external observer.

Part II — Toward an Ecological Science of the Affective Sphere. Part II of the volume concerns the complexity and the dynamical system or process nature of emotion, in several respects — the intricacy of its intentionality, the complex and multiple neural pathways involved in the interconnections between emotion and other cognitive processes, and the interwoven feedback loops and shunt mechanisms that are the earmarks of the brain as a dynamical system. Philosopher and clinical psychologist **Eugene Gendlin** opens this part by showing how we must develop a new kind of science capable of including subjective experiences such as the affective dimension presents — a 'first person science' that can be

interconnected with third person science by virtue of a new discipline akin to ecology, which views complex processes as fundamental, since the identity of their constituents is dependent on the role of each constituent in the overall process, and the identity of the process is defined by the way it replaces constituents to maintain its pattern, rather than by the constituents themselves. First person phenomena are especially in need of being elucidated in terms of complex processes in this sense. If we study the constituents as if they could exist in isolation from the process that appropriates them and gives them their structure, we are breaking up Humpty Dumpty in a way that makes it impossible to fit him together again, and this is a serious impediment to our current understanding of the role of affective processes in experience.

Valerie Hardcastle, a specialist in the philosophy of neuroscience, argues that understanding emotional processes requires a compromise between 'social constructivism' and 'physiological reductionism.' The key to this compromise involves the fluid way that dynamical systems interact with their environment, maintaining stability across perturbations. According to Hardcastle, the brain is an actively self-organizing system that is continually in a state of dynamic equilibrium at the edge of chaos; perceptual stimuli do not mechanically cause the brain to 'respond' in certain ways, but only disturb its already ongoing dynamic activity, allowing it to create a new 'basin of attraction' associated with the newly learned stimulus pattern. This process is the basis for learning and memory.

Marica Bernstein, Samantha Stiehl, and **John Bickle**, two biologists and a cognitive theorist working with the Focused Research Program in Computational Neuroscience at the University of Cincinnati, report results of their efforts, using computer modelling, to look specifically at the way corticothalamic loops allow the organism's motivational purposes to control the direction of attention toward emotionally important items in the subject's perceptual field, thus providing a basis for understanding how conscious processing, by contrast to nonconscious processing, always has an emotional dimension. They interpret the question, 'how do emotions affect consciousness?' in terms of the way parallel processing in brain areas associated with emotion and motivation (i.e., limbic areas) interact with areas associated with 'higher' cognitive abilities. 'Voluntary' visual attention is their model system; thus they are asking questions about how the saccade-generating system of frontal eye fields and lateral intraparietal area are connected to, and are influenced by, processing in cingulate gyrus, amygdala, etc., and they approach the question from a neurocomputational perspective. This leads them to a new hypothesis about how brains really could get a serial stream of consciousness, in part 'controlled' by emotion, from parallel processing.

Mark Bickhard is a process theorist, and sets as his task to develop a coherent process-oriented account of emotion and consciousness which is capable of resolving questions about the intentionality of relations between a dynamical

organism and its environment. The question of the intentional relationships in an emotional and representing system are important for understanding the implications of a self-organizational approach for the questions of cognitive scientists about the psychophysical reducibility of consciousness, and the multiple realizability of conscious processes. For Bickhard, motivation arises from representation of *potentialities for action* in relation to the environment, and emotions are complex, dynamical processes that intend success or failure in reducing *uncertainty* with regard to those interactions. Specific motivations do not explain why organisms do something rather than nothing, as if the organism would be completely inert if not for some external energy source. On the contrary, living organisms *cannot* be inactive, by their very nature, so what specific motivations explain is only why they choose this action rather than that.

Nick Georgalis, a philosopher and linguist, continues the line of thought Bickhard has initiated. Georgalis is interested in the intentionality of consciousness from the perspective of dynamical systems, arguing that the difference between conscious and nonconscious processing is an extreme one, and that no nonconscious system can genuinely manifest the property of intentionality. In this paper, Georgalis integrates this analysis of intentionality into a view of the brain as a dynamical system.

Part III — Emotional Development and Learning. Part III explores the problem of emotional learning from the perspectives of three different disciplines. **Peter Zachar**, a professor of clinical psychology, looks at psychological theories of emotion, especially object relations theories, that emphasize a crucial, active role for emotions in directing the stream of consciousness and development of behavior patterns. Many psychodynamic approaches in a broad sense (including but not limited to Freudian approaches to the unconscious) have emphasized that 'the passions' are not 'passive,' that emotions are active driving forces that mold our experience and motivate the direction of our attention. Zachar explores the contributions of object relations theory for an understanding of both normal and abnormal emotional maturation.

Gary Backhaus, a phenomenologist, carries the self-organization line of thought a step further by showing that no emotions in human adults are genetically 'hardwired' to specific intentional objects. I.e., the 'instincts' of adults cannot be thought of as the same as the 'instincts' with which infants or young children are equipped when they come into the world. Thus it is misleading to think of the maturational development of emotional experience in terms of the suppression of 'natural' or 'instinctual' motivational tendencies which then must be 'modified' or 'repressed' by adult learning and experience. Instead, the organism is structured in such a way that the dynamics of its own development lead it naturally to change its organizational patterns, and motivational impulses

change in the process. This introduces a helpful way of thinking about the way 'nature' and 'nurture' interact to form personality patterns. 'Learning' is not the only way for an organism to change its patterns of consciousness.

To conclude Part III, **Martin Peper**, neuroscientist at the University of Freiberg, reports on methods of empirically researching aspects of emotional development. This paper is an especially good model for those who are interested in doing carefully operationalized research on emotion and emotional development.

The key to an interdisciplinary approach to emotion and consciousness seems to be the idea that biological systems are self-organizing, that the brain is such a system, and that conscious experience arises from the interplay of such a self-organizing system with its environment, not just by means of intentional representational activity, but through the combination of this intentional activity with the organism's own dynamic purpose-directedness. This seems to provide a promising strategy for understanding how conscious information processing systems are different from nonconscious ones.

References

Clark, Andy. 1997. *Being There: Putting Brain, Body and World Together Again.* Cambridge: MIT Press.

Crick, F. and C. Koch. 1990. "Some Reflections on Visual Awareness". *Cold Spring Harbor Symposium on Quantitative Biology* 15: 953 62.

Cytowic, Richard. 1993. *The Man Who Tasted Shapes.* New York: Warner.

Damasio, Antonio. 1994. *Descartes' error.* New York: Putnam.

Damasio, Antonio. 1999. *The Feeling of What Happens.* New York: Harcourt, Brace & Co.

Edelman, Gerald. 1989. *The Remembered Present.* New York: Basic Books.

Ellis, Ralph D. 1995. *Questioning Consciousness: The Interplay of Imagery, Cognition and Emotion in the Human Brain.* Amsterdam: John Benjamins.

Ellis, Ralph D. and Newton, Natika. 1998. "Three Paradoxes of Phenomenal Consciousness: Bridging the Explanatory Gap". *Journal of Consciousness Studies* 5: 419–42.

Freeman, Walter. 1987. "Simulation of Chaotic EEG Patterns with a Dynamic Model of the Olfactory System". *Biological Cybernetics* 56: 139–150.

Gendlin, Eugene. 1992. "Thinking Beyond Patterns". In B. den Ouden and M. Moen (eds), *The Presence of Feeling in Thought.* New York: Peter Lang.

Gibson, Eleanor J. 1988. "Exploratory Behavior in the Development of Perceiving, Acting, and the Acquiring of Knowledge". *Annual Review of Psychology* 39: 1–41.

Globus, Gordon. 1992. "Towards a Noncomputational Cognitive Neuroscience". *Journal of Cognitive Neuroscience* 4: 299–310.

Jackson, Frank. 1986. "What Mary Didn't Know". *Journal of Philosophy 83:* 291–295.

Kauffman, Stuart. 1993. *The Origins of Order.* Oxford: Oxford University Press.

Kim, Jaegwon. 1992. "Multiple Realization and the Metaphysics of Reduction". *Philosophy and Phenomenological Research* 52: 1–26.

LeDoux, Joseph. 1996. *The Emotional Brain.* New York: Simon and Schuster.

Luria, Alexander R. 1980. *Higher Cortical Functions in Man,* 2nd ed. New York: Basic Books.

Mac Cormack, Earl and Staminov, Maxim (eds). 1996. *Fractals of Brain, Fractals of Mind.* Amsterdam: John Benjamins.

Mack, Arien, and Irvin Rock. 1998. *Inattentional Blindness*. Cambridge: MIT/Bradford.

Merleau-Ponty, Maurice. 1942/1963. *The Structure of Behavior.* A. Fischer (trans.). Boston: Beacon.

Monod, Jacques. 1971. *Chance and Necessity.* New York: Random House.

Needleman, Jacob. 1968. *Being in the World: Selected Papers of Ludwig Binswanger.* New York: Harper & Row.

Panksepp, Jaak. 1998. *Affective Neuroscience.* New York: Oxford University Press.

Posner, Michael I. and Mary K. Rothbart. 1992. "Attentional Mechanisms and Conscious Experience". In A.D. Milner and M.D. Rugg (eds), *The Neuropsychology of Consciousness.* London: Academic Press.

Searle, John. 1984. *Minds, Brains and Science.* Cambridge: Harvard University Press.

Thelen, Esther, and Smith, Linda. 1994. *A Dynamic Systems Approach to the Development of Cognition and Action.* Cambridge: MIT Bradford.

Varela, Francisco, Evan Thompson, and Eleanor Rosch. 1991. *The Embodied Mind.* Cambridge: The MIT Press.

Watt, Douglas. 1998. "Affect and the 'Hard Problem': Neurodevelopmental and Cortico-limbic Network Issues". *Consciousness Research Abstracts: Toward a Science of Consciousness, Tucson 1998:* 91–92.

Part I

The Centrality of Emotion

CHAPTER 1

Integrating the Physiological and Phenomenological Dimensions of Affect and Motivation

Ralph D. Ellis
Clark Atlanta University

A neglected but recently resurgent tradition in psychology and neuroscience emphasizes that affect, emotion, and motivation are central components in all conscious processes, and in fact are crucial for the differences between the way conscious and nonconscious beings process information (Freeman 1975, 1987; Pribram 1980; Macrides *et al.* 1982; Ellis 1986; Cytowic 1993; Varela *et al.* 1993; Alexander and Globus 1996; Watt 1998, In press; Ellis and Newton 1998, In press). In this tradition, beings with affective feelings are members of a larger class of beings that are living rather than nonliving. 'Living' in this context does not mean 'composed of carbon, hydrogen, oxygen, and nitrogen rather than silicon or steel.' Rather, it refers to beings that 'organize and maintain' themselves by actively appropriating, replacing, and reproducing the material components needed to maintain a continuity of organizational structure (Merleau-Ponty 1963; Monod 1971). Recently, such structures have come to be understood as types of 'dynamical systems' (Kauffman 1993; Thelen and Smith 1994; Mac Cormack and Staminov 1996; Ellis 1999a, b, c, 2000).

One of the exciting aspects of the renewal of this tradition is that it offers possibilities for bridging the gap between the neurosciences and a phenomenological understanding of the conscious experiencing process, which for so long have seemed incommensurable. Self-organizing structures, or dynamical systems, are understood as open thermodynamic systems, i.e., as systems that exchange energy and constituent materials with their environment, yet are organized so as to maintain continuity in their tendency to settle into complex but homeostatic patterns of activity called 'attractors' or 'basins of attraction.' Biological organisms are the prime examples of such complex but stable systems. The identity of the organism is constituted by the continuity of its pattern of activity, not by a continuity of specific material components. Organisms can learn and remember by creating new basins of attraction structurally related to the learned

stimulus pattern (Freeman 1987, 1988; Alexander and Globus 1996).

Pursuing this approach could make two crucial differences in our understanding of the role of consciousness in cognitive processes. (1) It would lead to a completely different assessment of the *structure* of information processing in humans, which almost certainly would turn out to be very different from the structures of information processing that can be constructed in nonconscious machines, or in flow charts based on abstract conceptions of information *content* relationships, or even as an 'appendage' (as Natsoulas 1993 puts it) to better-understood *nonconscious* processes in the brain. And (2) it could facilitate a meaningful response to Chalmers's (1995) 'hard problem.' Even if we can explain completely the physical mechanisms that result in information processing, based on knowledge gleaned from empirical science, there still seems to be a leftover element that has not been explained: We still would not have explained why those particular mechanisms should result in *consciousness* rather than nonconscious or zombieistic information processing and behavior. This incommensurability between phenomenology and physiology has been an impediment to neuroscientists' attempts to correlate physiological processes with the conscious processes that we can subjectively experience. Particularly with emotion and affect, if we begin with an oversimplified or superficial phenomenological account, we are unlikely to successfully correlate elements in this faulty account with elements of *any* physiological story, however sophisticated. Also, the realm of *conscious* processes includes many *cognitive* processes; thus a bad phenomenological account can also obstruct attempts to understand the workings of human cognition.

How could a self-organizational theory of emotion and motivation bridge between the neuroscience and the phenomenology of conscious experience? First, affects and emotions in conscious organisms can be thought of in terms of patterns of self-organization that tend toward the 'basins of attraction' discussed in dynamical systems theory, so that processes that we actually experience could also be understood in terms of perspicuous scientific accounts of 'embodiment' (in a sense that must be explicated in self-organizational terms). Secondly, conscious, living organisms are 'purpose-oriented'; in some sense, they 'act' on their environment rather than merely 'reacting' to it, and form 'intentions' toward it rather than merely 'tendencies'. Their behavior is 'motivated' rather than merely 'moved.' The twentieth century sciences have had a great deal of difficulty in making sense of these distinctions. The notion of self-organization provides resources with which such distinctions can meaningfully be drawn. The difference between living and nonliving can be formulated in terms of it; we can hope to explain why an organism is 'more than the mere sum of its parts,' since self-organizing systems actively appropriate and continually replace the needed 'parts' to facilitate the survival of the patterns of activity that define their

identity. They do not merely 'respond' to 'stimuli.' To elaborate such a notion in physiological and phenomenological detail would constitute what I shall call an 'enactive' rather than merely 'reactive' view of the emotions. This usage is meant to be consistent with the way Varela *et al.* (1993) use the term 'enactive.'

To understand consciousness and cognition as features of self-organizing systems requires understanding what is so special about *some* self-organizing systems, e.g. humans, that leads to the familiar experience of phenomenal consciousness; one exciting possibility is that, as Cytowic (1993) suggests, *emotion and affect* are among the properties of certain living systems that ground their peculiar ability to sustain what we call phenomenal consciousness. In Cytowic's view, "Consciousness *is* a type of emotion" (1993: 194, italics added). Affective tendencies, or motivations, are the key to bridging between the biological self-organizing system, which can be understood scientifically, and the experience of consciousness, which can be accessed through careful phenomenological reflection on our own ongoing experiencing process. Thus if we can understand emotion and motivation as features of dynamical systems, and then understand the central role of emotion and motivation in constituting the difference between conscious and nonconscious processing, then we can understand human cognition and consciousness in scientific as well as phenomenological terms which at that point have become commensurable.

For example, it is well known that, in response to a visual stimulus, the occipital lobe can be fully activated, corresponding to the processing of the visual data — the lines, angles, colors, etc. — yet *consciousness* of the visual object can still be lacking (Posner and Rothbart 1992; Mack and Rock 1998). *Emotional and motivational* brain processes must become involved, so that attention to these visual stimuli has already been motivated by the emotionally significant purposes of the organism, before there can be conscious (as opposed to nonconscious) processing of the data, in the form of enhanced corticothalamic loops (Bernstein *et al.*, this volume); increased activation of neurotransmitters that correlate with consciousness (Faw, this volume); and anterior cingulate and other *efferent* activity, which first 'looks for' emotionally important data, and then, when the corresponding stimuli are found, resonates with the *afferent* responses going on in the occipital lobe (Posner and Rothbart 1992; Ellis 1995; Faw, this volume and 1997). All these motivational activities are aspects of the organism's ongoing dynamical activity, which then is merely 'disturbed' by perceptual input (Hardcastle, this volume and 1996).

There is no danger here that to understand consciousness as grounded by emotion and motivation is circular given that emotion and motivation are already conscious phenomena — so that our 'explanations' would become question-begging and would already assume an understanding of what they are supposed to be explaining. Emotion and motivation can be understood as phenomena that

lend themselves to *either* a conscious *or* an nonconscious status. I.e., they have to be combined with *something else* in order to be phenomenally conscious data of which we can be 'aware.' What that something else is must be worked out carefully and in detail, but a good candidate seems to be imagery or representation (Newton, this volume; Ellis and Newton 1998); the aims and objects of emotions must be 'represented' in some sensory or proprioceptive modality in order for them to become conscious. Otherwise, they are unconscious, are structured quite differently from the way they would be if they were conscious, and give rise to very different behaviors with nonconscious processing, which is more piecemeal and less complete, as Section 3 below will discuss more fully.

I use the terms 'motivation,' 'unconscious emotion,' and 'conscious emotion' in the ways summarized in the introduction to this volume. I.e., four different levels are possible in self-organizing systems: (1) Unified *tendencies* that are neither emotions nor motivations; (2) Motivations that are not associated with any emotions; (3) Motivations that are associated with unconscious emotions; and (4) Motivations associated with conscious emotions.

The next section will discuss some important contrasts between two different ways of approaching the way emotions are caused — the 'reactive' and the 'enactive.' My contention is that the reactive approach has been dominant throughout the twentieth century, but that it is now rapidly giving way to an enactive approach. Then Section 2 will suggest some physiological advantages of the enactive approach, and Section 3 will discuss the complexities of the phenomenological experiencing of the intentionality of motivational feelings, with a view to connecting the phenomenological and physiological domains.

1. The enactive and reactive approaches

In most theories, emotions are thought of as intentional or potentially intentional processes that can involve both phenomenal and physical elements, even if emotions *per se* are not necessarily conscious. I.e., when we do pay conscious attention to our emotions, they seem to present themselves as being 'about' something with respect to which we want to do something or avoid something. Even disinterested 'blissful' states aim to avoid distracting thoughts and prolong the state.

Of course, aims may be general rather than specific: a 'restless' mood may not desire a particular outcome, but it does generally intend some sort of change, some instance of a wide class of activities that could relieve the 'restlessness.' Some may question whether all emotions have *objects,* but there is much less controversy as to whether they all have *aims,* at least of a general kind. Since aims also involve object-relatedness, arguments against the aim-relatedness and

object-relatedness of emotions really amount to arguing that we are not always consciously *aware* of the aims and objects of our feelings.

So we can say that emotions are intentional *or potentially intentional* states of an organism that lead it to desire or avert some activity of the organism in relation to some object or situation (where 'doing' something can also include engaging in some further experiencing process). Emotions are thus thought of as having *objects* and *aims,* but with considerable breadth and open-endedness in the classes of objects and aims intended by any particular emotion. The environmental situation intended in relation to the aim of the emotion is the 'object,' and the activity of the organism intended in relation to that object is the 'aim.' It is true that we may 'disinterestedly contemplate' a mood or feeling, as in meditation, but we would not do so if we did not want or intend to focus our attention on it in the first place (Panksepp 1998: Chapter 8). So even disinterested contemplation involves an aim (focusing attention) and an object (the feeling on which we want to focus).

Another area of agreement for most contemporary approaches to emotion is a rejection of dualism. In a dualistic theory, the phenomenologically experienced dimension would have to be nonphysical, but the bodily dimension would have to be physical — which would seem to make the phenomenon of conscious emotion or affect impossible in principle. For example, in a dualistic theory, there is no way to understand how a conscious choice to raise my hand can be a necessary part of the physical causal force that makes the hand go up. Dualistic approaches to emotion thus have few serious contemporary advocates because they cannot explain mental causation — how the *conscious experience* of these 'passions' can *causally affect the physical realm,* especially our own bodies, and also how the supposedly physical 'passion' can cause effects at the conscious level. Whether or not we consciously experience an affective feeling *does* often seem to affect the way we execute physical actions, but dualism entails that it should not be able to do so. Attempts to explain emotions as purely physical passions — as merely the passive conscious registering *of* a causally-sufficient physical event — thus lead to the same insurmountable problems facing mind/body 'interactionism' in general. Such approaches fail to explain how there can be any true causation between physical and conscious levels, since to suppose that a *mental* cause is necessary would lead to corresponding ruptures in the sufficiency of *physical* causal chains (Ellis 1986, 1995). As Kim (1993) points out, if physical events have purely physical causal explanations, then conscious events should be causally irrelevant to them. So conceiving of mind and body as different kinds of entities leads to irresolvable problems regarding the causal relationships involved.

Some commentators see recent *panpsychist* theories (e.g., Chalmers 1995) as attempts to revive dualism, but this would be a hasty and possibly unwarranted

charge. According to panpsychism, every event has both a physical and a conscious dimension. Although this entails that the conscious (or psychic) dimension cannot be *explained by* the physical dimension (Muller 1998), panpsychists may or may not be dualists for this reason. The crucial question is whether only the physical aspect of things can have causal power, or whether consciousness also has it. Answers could be dualistic, epiphenomenalist, reductionist, etc. To my knowledge, Chalmers has not yet committed himself with respect to these questions.

In any event, the two main types of theories currently taken seriously are not straightforwardly dualistic: They account for mind-body interaction in terms of either a 'reactive' or an 'enactive' approach to the causation of emotion. In *reactive* theories, emotions are *caused by* discrete chemical reactions or physical changes. In such theories, it is natural to conceptualize the *aims* of emotions in terms of a causal physical process such as the completion of an oxidation-reduction reaction somewhere in the nervous system, and the phenomenal emotion itself is simply the conscious correlate of a state of chemical imbalance or restoration of balance (such as the reduction of electrostatic charges). These theories also think of all emotions as *drive reductive.* A reductive 'drive' is defined as a physiological deficit external to the nervous system, perceived by the nervous system as noxious. When the 'deficit' is neutralized, the discomfort goes away, and the drive has been 'reduced' (White 1965).

The twentieth century was dominated by a *dispositionalist-behaviorist* form of reactivism. In these, emotion is only a disposition to behave in a certain way, strengthened or weakened by reinforcement — with the reinforcement consisting of an electrochemical change. Thus the aim of every emotion is to achieve electrochemical neutrality and to reduce the valences of all the atoms in the nervous system. This reduction of electrochemical imbalances (drives) happens to serve other organismic purposes only because natural selection has made it so (e.g., Eysenck 1957; Spence and Spence 1964; the young Freud of course pioneered this view, but later in *Beyond the Pleasure Principle* he realized that if electrochemical homeostasis were the aim of all emotions, then the aim of life would be death — see Freud 1959). Interestingly, many connectionists *could* subscribe to a dispositionalist-behaviorist theory, if they were to apply connectionism to emotions, since the disposition to behave in a certain way could be thought of as 'all in the weights' (Tienson 1987).

The behaviorist type of reactivism is currently going out of fashion, essentially because it is less an attempt to explain emotions and other conscious and intentional phenomena than it is an attempt to *avoid* explaining them. There was a time when *epistemological* behaviorism was attractive, because so little was known about the brain that it was impossible to discover the reliable correlations between neurophysiological and phenomenal processes that now

make measurement of the subjective states much more feasible. EEG patterns, fMRIs, CT scans and other measures of neural activity in various parts of the brain have now been extensively correlated with conscious acts such as remembering (Damasio *et al.* 1985; Damasio 1989); attention (Hernandez-Peon *et al.* 1963; Cohen *et al.* 1988; Posner and Rothbart 1992); the integration of sensory and memory mechanisms *via* frontal lobe activity (Nauta 1971); obsessional thought patterns (Gibson and Kennedy 1960); hysterical conditions (Flor-Henry 1979); feelings of elation and depression (Gainotti 1973; Damasio and Van Hoesen 1983; Ahern and Schwartz 1985); the activity of listening to music (Miller 1990: 79) — which apparently involves very different brain areas for trained musicians (more left-lateralized); word recognition (Petersen *et al.* 1990); language acquisition (Dore *et al.* 1976); and many other such consciousness/brain-electrical correlations. Clinicians now are therefore less impressed with the idea that they should ignore subjective conscious states on the grounds that they are not systematically observable or measurable. It is true that the measurements are indirect, and depend on some prior knowledge or subjective reports of the phenomenal states to be linked to the physiological measurements; but, in principle, the results to be achieved by investigating phenomenal states are now too important to warrant simply ignoring them.

A different kind of reactivism, an *epiphenomenalist* form of reactivism, is currently more popular than dispositionalist-behaviorist reactivism. In epiphenomenalist theories, conscious emotions are the phenomenal states that merely *register in consciousness* certain aspects of what is going on in the physiological-drive system. What goes on at the phenomenal level is therefore literally a causal reaction to what goes on at the physical level. Thus, for Searle (1984), we can accommodate the causal sufficiency of physical mechanisms by attributing no causal power to any phenomenal state, and then explaining the phenomenal state as a causal *byproduct* of the physical mechanism.

Epiphenomenalist reactivism, whether for epistemological or ontological reasons, tends to explain emotional experience in terms of what Natsoulas (1993) calls an 'appendage' theory of consciousness. In appendage theories, an emotion experienced at the conscious level is viewed simply as the corresponding unconscious physiological process, plus an extra layer of consciousness superadded to it, as if the conscious emotion, X, were nothing but the unconscious emotion X plus our awareness *of* X. This ignores the phenomenological complexities involved in the relations among conscious, preconscious, and unconscious emotions. The process through which X becomes explicated as a conscious phenomenon is complex; as we experience the emotion consciously, it *changes* its structure and its assessment of its goals and intentions. An *unconscious* frustration at the way my career is going may at first be experienced merely as anger at my son for failing to take out the garbage; further *conscious* attention to

the emotion reveals further and further complexity as to what the feelings are 'about' (Gendlin 1992). Both the intentional object and the felt quality of it change as we consciously explore its 'felt sense.' Thus if X is a conscious emotion, X is not simply the *same* bodily response that could just as well occur unconsciously (Ellis 1986, 1995, 1996; Jackendoff 1996).

A more general problem with epiphenomenalist reactivism is a corollary of criticisms of epiphenomenalism *per se.* It is very difficult for this type of approach to account for the *causal power* of consciousness. It is true that a sophisticated epiphenomenalist, such as Jackendoff (1996), can say that conscious states only *seem* to have causal power because their underlying physiological substrates have causal power. But the problem with this response is that, if we say that consciousness is a causal byproduct of a physical process, then the question arises as to whether the consciousness itself is physical or not. If it is physical, then it should have *causal power,* like all other known physical events; but if it is *not* physical, then it reduces to a form of metaphysical dualism in which the ontological status of the conscious event is not only mysterious and occult, but more importantly falls victim to all the criticisms of dualism itself, which are formidable.

Many people who feel intuitively drawn to epiphenomenalist reactivism, after considering these kinds of arguments, find that their views are better represented by a third type, *psychophysical identity* reactivism, which does have a way to account for the causal power of consciousness, as discussed next:

In psychophysical identity reactivism, emotions just *are* physical processes, which just *are* both conscious and physical. The *reactive* type of psychophysical identity treats emotions simply as aggregates of discrete chemical-physical reactions that operate not self-organizationally, but in *partes extra partes* mechanical fashion. (Advocates of this approach include Kim 1992, 1993; Goldman 1969, 1970; Davidson 1970.) Part of the attraction of this view is that it offers a more coherent ontological story than either behaviorism or epiphenomenalism. If we *equate* a state of emotional consciousness with its physiological correlates, then we can attribute causal power to the conscious emotion, since to do so is equivalent with attributing that same causal power to the physical correlates of the emotion. Thus, for example, we can *equate* the conscious choice to raise our hand with the physical brain events that cause the hand to go up.

Problems with this approach are of two types: First, a given conscious process often seems to be 'multiply realizable' by means of various combinations of physiological substrates. E.g., relief from depression, for some individuals, may feel the same whether the underlying mechanism for the relief is obtained by the action of zoloft or of prozac, or by resolving some important life problem, which in turn changes the subject's self-regulated brain chemistry in ways different from the ways zoloft or prozac change it. If the same feeling can be

obtained by a multiplicity of physiological processes, then it seems strange to say that some given physiological process can be necessary and sufficient for the feeling in question; it seems that the feeling could occur in the absence of some specific substrate, provided that another is available. The problem is that a self-organizing relation cannot be reduced to any particular collection of physical substrata.

This multiple realizability might be an acceptable consequence for sophisticated versions of psychophysical identity, if it were not for the *second* type of problem with such theories — the apparent empirical unobservability of the relation that constitutes a state of consciousness. I.e., psychophysical identity theories have trouble responding to sophisticated versions of the 'knowledge argument' (Jackson 1986; Horgan 1984; Ellis 1986, 1995): If experiencing were literally equivalent with the brain states that empirical neurophysiological methods can observe, then complete empirical knowledge of brain states should constitute knowledge of everything about my experiencing; but complete empirical knowledge of the brain states would *not* constitute knowledge of everything about the experiencing (the observations alone would not reveal 'what it is like' to have that experience); therefore, there is something about experiencing that is not equivalent with empirically accessible components of the process. This argument shows, not that consciousness is non-*physical* (since some physical processes *might* be observationally inaccessible from an external perspective), but rather that consciousness goes beyond what is directly empirically observable. It is true that an empirical observer of the brain can *infer* what the subject might be feeling, but she cannot do so from the observations alone; what the feelings 'are like' is known through memory of her own *subjective* experiences. The problem is this: It is difficult to see how a phenomenon that is not empirically observable could be exactly equivalent with one that *is* empirically observable.

Of course, the other side of this coin is that, if consciousness were *non*-physical, then there should not be such systematic correlations between the conscious and physical realms. Some theorists are thus led to search for an account in which consciousness can be *physically* instantiated, yet not exactly equivalent with anything that is empirically observable (from an external perspective). One hope for such an account arises from the recent theory of self-organization (Monod 1971; Kauffman 1993), in which emotions and motivations are aspects of the tendency of self-organizing processes (e.g., biological organisms) to achieve and maintain certain patterns of organization, and in which the self-organizing process as a whole acts to appropriate and replace the physiological substrates needed to maintain the overall pattern of activity. To experience is to execute those self-organizing activities in a unified agent-directed way. Since no one else can execute my actions, it is not surprising that they cannot experience

my feelings, even though they can empirically observe their components. I shall refer to such approaches to the role of emotion and motivation as 'enactive' theories. On the enactive view, behavior is enacted by an organism unified by fundamental biological purposes, using environmental elements as demanded by the self-maintaining agenda of those purposes, rather than passively resulting from piecemeal inputs and micro-events. The total organismic blueprint continually readjusts both micro-constituents and environmental conditions as needed to meet its purposes.

This 'enactive' approach sharply contrasts with the reactive approach: On the enactive conceptualization, organisms have endogenous, holistic structural properties, and can rearrange their own micro-constituents and use environmental inputs for the purposes of maintaining structure. Thus an organism is not angry because of the release of norepinephrine; on the contrary, it releases norepinephrine because it senses a threat to its ongoing, holistic balance, and is mobilizing a search for a way to restore this self-maintaining pattern. Emotions are aspects of self-organizing processes, where the overall processes actually work to organize their own discrete substrata, such as particular chemical reactions.

This approach differs from reactive approaches because the enactive theorists do not treat the process merely as caused by the actions of its current constituents; instead, the self-organizing processes that ground emotions and motivations have the power to *appropriate and replace* the substratum elements needed to maintain the desired pattern. It accomplishes this result by means of a complex network of overcausation, using shunt mechanisms and other redundant causal systems typical in biological contexts, and by continually rearranging the background conditions needed for any given causal micro-sequence to obtain. The structural process is holistic in the sense that the system tends to rearrange some of its parts if needed to restore the balance in other parts. The structure of the pattern is caused not by current micro-level events, but by past events, for example natural selection guiding genetic changes long ago. The holistic pattern tends to incorporate current elements as they become available, as in the development of an embryo. This does not mean that human behaviors are not caused, but rather that the causation is complex, dynamic and self-maintaining. The process massively overdetermines, not uniquely specific outcomes, but structurally defined overall patterns of outcomes. Advocates of this enactive approach include Maslow (1968); Freeman (1988); Gendlin (1992); Bickhard (1993); Globus (1992, 1995); Neisser (1994); Mac Cormac (1996); and Stamenov (1996). Currently, the most prevalent way of explaining the causal power of self-organizing processes to rearrange and replace their own substratum elements, without at the same time violating the sufficiency of causal chains at the substratum level (e.g., efficient chemical reactions) is to be found in 'dynamical

systems theory' (e.g., Kauffman 1993; Globus 1992; Thelen and Smith 1994; Mac Cormack and Stamenov 1996).

This theory is still in its infancy, and part of our purpose here is to insist that such an approach, if it is to become a viable account of consciousness, must allow a place for emotion and motivation as key components of a self-organizing system capable of accommodating the above problems about supervenience, causation, and the apparent empirical unobservability of the 'what it's like' dimension of phenomenal consciousness.

Enactive theories regard the aims and objects of emotions as holistic-situational in relation to the total organism's dynamical functioning, rather than as object-specific. I.e., they regard as overly simplistic the view that some particular stimulus event could be the object of an emotion — e.g., a piece of food, or someone's insulting me. Such a simple stimulus-response approach to intentionality does not distinguish between the specific *trigger* of an emotion and the more general realm of issues that an emotion may be *'about.'* For example, my son's failure to take out the garbage may *trigger* anger, when in fact the anger may involve a much broader category of situations in my life.

In holistic-situational approaches, the objects of emotions can be very general facts about broad environmental situations — that an area is not a good place in which to thrive (perhaps *triggered* by the inadequacy of a specific piece of food), or that my overall environment makes it difficult to maintain self-esteem (perhaps *triggered* by a particular person's insult). The remainder of this paper will explore an account of emotion and motivation of this last type — an enactive, holistic-situational approach. Such an account is consistent with a view of the organism as a self-organizing process or dynamical system. As a first step, I shall consider some practical differences it makes whether we apply an enactive or a reactive approach. Then the concluding section will connect these concepts to a phenomenological characterization of the *intentionality* of emotion and affect in a way that can fit with what emotions actually *feel* like when we carefully reflect on them phenomenologically; then we can hope to correlate a physiological account with these feelings.

2. Physiological advantages of the enactive approach

A simple example will illustrate the difference between the enactive and reactive approaches to understanding the workings of emotion. Suppose I hear a noise outside the house I am in the process of burglarizing, followed by a sudden fear response. In the *reactive* approach, the incoming signal from the stimulus splits when it reaches the thalamus, into a slow path through sensory cortices to imaging and intellectual brain areas, and a fast path directly to the amygdala,

which in turn causes the bodily responses we feel as the emotion of fear. The organism's total response to the noise is indirectly caused by the stimulus. I.e., in the slower pathway, the stimulus activates mechanisms in the primary sensory area for sound (in the temporal cortex), which in turn activates the secondary and tertiary sensory areas (which encode parameters of the sound such as its volume, pitch, rhythms, timbres, etc.); meanwhile, the incoming signal has already taken the *fast* path through the thalamus directly to the amygdala, which is caused to undergo chemical changes leading to a variety of bodily responses. If the perceptual and intellectual processing (which takes longer) then assesses the noise as unthreatening, the brain *inhibits* the fear response that has already begun (LeDoux 1996; Goleman 1997).

Even with such a simple example, the reactive approach encounters difficulties. Why is it that the noise outside elicits an extreme fear response while I am burglarizing the house, but none at all if the exact same noise occurs while I am simply a guest in the house? LeDoux's (1996) answer to this question is that the fear response is susceptible to classical conditioning. We can learn through repeated pairing of stimuli to respond with fear to a stimulus in some contexts but not others. One problem with this response is that it does not explain how single-trial learning is possible. I.e., the amygdala can activate an extreme fear response to a stimulus under certain circumstances, but not under others that are different in very complex, subtle, and often socially defined ways, *without* any conditioning intervening between the two instances.

The enactive approach, while accepting much of the above description of the 'slow' and 'fast' pathways, would give a different answer to this question. In the instance where the noise occurs during the social visit, the enactive approach would *not* say that the amygdala first activates a strong fear response, which is then inhibited *after* perceptual and intellectual processing reveals that the noise indicates no danger. Instead, it would say that the amygdala has *already* been primed, through input from other adjacent brain areas, to be either hypervigilant or relaxed. The brain as a whole self-organizing system is geared toward maintaining certain balances, and it actively and sensitively seeks out environmental resources that are screened and detected with the full sophistication of all the brain's perceptual, imaginative, and thinking resources.

On the reactive view, emotions are primitive and stupid, but faster than intellectual and perceptual processing. On the enactive view, the very first activation of an emotion — prior to full perceptual and intellectual processing — can show intelligence, sophistication, and subtle discrimination, because emotional responses are part of a total process that already integrates the perceptual and intellectual information into the total pre-response set of the organism.

Of course, emotions *can* be stupid and primitive, analogously to the ways in which superstitious, neurotic, or self-deceptive thought can be. But emotion is

not by its very nature a simplistic, brutish, course-grained system in need of refinement and inhibition by subsequent intellection.

Consider another example. Suppose an extremely neurotic, dysfunctional kind of character who obviously poses no physical danger approaches me in a safe environment (e.g., in a crowded lecture hall) with an insulting demeanor. On the reactive view, my amygdala would first be caused by this stimulus pattern to begin executing an extreme anger response; then after my intellectual capacities determined a second later that the person was neurotic and dysfunctional, posed no real threat, and did not warrant initiating any verbal altercation, this intellectual decision would override and inhibit an amygdala reaction that had already begun. On the enactive view, by contrast, it would be possible for a person to have already decided that neurotic, dysfunctional kinds of characters in physically safe environments are not to be gotten into altercations with, and thus the amygdala would already be pre-tuned in such a way that the stimulus would not even elicit a full anger response in the first place. There would be no need for subsequent intellectual functions to *inhibit* the anger response, because the anger response would already have been minimized by the subject's total pre-perceptual set. On this view, the amygdala is not just biologically determined to react instinctively to certain kinds of hardwired stimuli, susceptible to crude modification through classical conditioning. Instead, the amygdala on each specific occasion is informed by sophisticated thought that has already taken place, and its instantaneous responses to stimuli are pre-tuned according to these sophisticated sources of complex information. This can occur, for example, by virtue of the fact that the amygdala receives both direct and indirect inputs not just from the thalamus, but also from the anterior cingulate, the temporal lobes, the frontal lobes, the hippocampus, and other brain areas. As a part of a holistic organism, the amygdala becomes a fine-tuned, sensitive instrument, as opposed to a brute in need of inhibition by the other fine-tuned instruments of the brain.

Whether we take a reactive or enactive view affects our understanding of the intentionality of emotions — what they are about. On the *reactive* view, the intentional object of an emotion is just whatever object was *perceived* at the point when the input from that object triggered the activation of the emotional brain areas, and thus the aims of the emotion must be conceived of as directly involving that object. On the *enactive* view, the aim of all emotions is to preserve and enhance the self-organizational integrity of the whole organism, and the emotion is one part of a total system that includes different resources for monitoring what is needed to do so. If the neurotic but innocuous person who insults me triggers anger, the enactive approach does not assume that this innocuous person is what the anger is directed toward. On the contrary, it assumes that the amygdala is already pre-tuned by enough information about neurotic but innocuous people to be *capable* of avoiding hypervigilance toward

such people. The question would then become: What is the subject *really* angry about, that would cause her to be hypervigilant toward any innocuous opportunity to symbolize the anger by 'taking it out' on someone?

In the enactive approach, no assumption is made that the perceptual object at the moment of amygdala activation is the object of the emotion, let alone that the aim of the emotion has anything to do with that perceptual object. One of the goals of psychotherapy is to move a person to the point where the initial feeling that would accompany such an emotional trigger would not be "I want to beat the hell out of this sonofabitch, but I guess I shouldn't," but instead something more like "I wonder why I feel this tendency to agitation in response to an innocuous, neurotic person's insult? What is this feeling *really* about?"

A physiologist guided by the *reactive* approach will ask the question, "What is the underlying physical mechanism that can explain why Bill's perception of John's insulting behavior causes Bill to be angry at John?" The answer will involve determining the physiological mechanisms of the perception, and of the emotion, and asking how the perceptual mechanisms cause the emotional ones. By contrast, a physiologist guided by the enactive approach will ask, "What is out of balance about Bill's whole orientation to his environment, that is preventing the holistic equilibrium of his system, and which Bill's hypervigilance toward things like John's insulting behavior offers him an opportunity to explore?" In the former case, we allow ourselves to falsely assume that the perception causes the emotion. In the latter case, we leave open the possibility that the emotion may also cause the hypervigilance and other pre-perceptual tunings that lead Bill to look for and thus have the kinds of perceptions that he has, of which the perception of John's insulting behavior is an instance. In effect, cause and effect may well be the opposite of what the reactive theorist is led to assume, and certainly is more complex.

For these reasons, an enactive approach leaves room for a more subtle and intricate account of the intentionality of the emotions we feel — what they are 'about.' The next section explores the phenomenology of this question in relation to an enactive physiological account.

3. Understanding the phenomenological dimension of emotions

Before trying to correlate variables, we should try to have as clear as possible an idea of what the phenomena are that we want to correlate. If we begin with bad psychological variables, then we should expect to end up with poor correlations between them and the physiological processes we can observe. This means that a physiological account is not likely to be adequate to explain emotional processes if we have misconstrued the phenomenological dimension of emotional

processes to begin with. We then might end up with a mechanism that would be adequate to explain *something or other,* but which does not explain *emotion.* Emotion, we must remember, is distinguished from mere tendencies in physical systems by having a kind of 'intentionality' in a phenomenological sense — i.e., there is something in the environment that the emotion is 'about' or something that it 'means.'

In the following account, I follow closely the phenomenological approach to affective experiencing developed by Eugene Gendlin (1962/1997, 1971, 1973, 1992a, 1992b). I see this account as connecting well with a self-organizational approach to the physiological substrates of the way affective processes relate to their intentional aims and objects.

In forming questions about what kind of *intentionality* emotions have, we should be careful not to assume that the organism, simply by feeling an emotion, automatically *knows* what the emotion is about. This kind of assumption can cause trouble in two ways: The simplest way is to unwarrantedly assume that whatever 'triggers' an emotion (e.g., an insult) is what the emotion (e.g., anger) is 'about.' The trigger may be only an *instance* of a broader category, or even a reminder of a set of issues that it does not instance at all.

The second way this assumption can occur is by thinking that some homunculus within the organism would 'unconsciously' have the information already 'stored' somewhere about exactly what might be going on in the environment that the emotion is 'about,' while the organism's 'conscious mind' would merely have to perform some sort of search to find it. The intentional object of an emotion is often murky and difficult to pin down: Often it turns out that what the affective feeling is 'about' is completely different from what we initially supposed. And, in many cases, what the feeling 'wants' or 'wants us to do' is a more general category of outcomes than the specific 'trigger' for the emotion might lead us to believe. At the moment when someone insults me (the trigger for my emotional response), an accurate assessment as to what the feelings are really 'about' may not yet be present; to discover their real intentionality may require further reflection, which in turn would correlate with further physiological changes and with changes in the felt sense of the affect. The information is not yet present *in* the organism until these changes occur.

If we are to avoid confusing the *trigger* for emotion with its *intentional objects,* we must begin by appreciating that what emotions intend is never as simple as a trigger, or a simple behavior in relation to the trigger. Instead, the organism already has a pattern of activity that it is trying to maintain through the mechanisms of self-organization. When the organism is failing to engage in the pattern of activity that it is trying to engage in, it initially senses that 'something' isn't right. But, until something triggers us to pay attention to the felt sense of this 'wrongness,' we lack an explicit thematization of what that 'something' is.

The felt sense of the privation of the desirable pattern may *implicitly* contain the information about what is wrong, because there is a correspondence between what is wrong in the organism and what is missing from the environment that could fix it if present. But the organism cannot know exactly *what is* missing until it goes through a whole series of *conscious* processes which will be discussed below (or in some cases *'derivatively preconscious'* processes, in the sense of 'derivatively' in which contents may be habituated, sedimented, etc., from earlier conscious ones, as when we no longer pay attention to the preconscious process of reading music — see Ellis 1995: 18–20).

The initial felt sense of an emotion, prior to the elaboration of these conscious and preconscious processes involving refinement of the 'aboutness' aspect of the emotion, does have certain 'information' about what is wrong or missing or what it would like, because it knows very generally what it wants to *do* – i.e., engage in the total self-organizing process that the organism is trying to maintain. We seldom know much *consciously* about what that overall process *is,* literally; that would be too complex, consisting strictly speaking of millions of chemical equations and vast accurate knowledge of the environment. What we always *do* know, merely by having an emotion, is either that we are not *in* the desired pattern, or that we are. But we are unable to spell out exactly at which points we are falling short — as for example when we know that someone has played a piece of music badly, but cannot pinpoint exactly what they did wrong. We just know that it either feels right or not.

Initially, in feeling an emotion prior to conscious elaboration of intentional objects, we know already, 'preconsciously,' *that* something is wrong or missing (or that it isn't). I.e., we may not be *conscious* of the fact that something is wrong or missing, even though it is, and even though the fact that *something* is wrong is already information that is available to us through attention to bodily felt senses which we *can* pay attention to if we choose to (i.e., 'preconscious' bodily felt senses). But often, we do not choose to pay attention, just as one may have a sore neck, but not pay attention to it until at some point one realizes that it is sore — and we know that it was sore long before the person consciously realized that it was.

Here we come to the important role of *'triggers'* for emotional responses, which are often too hastily identified with the intentional objects or even the causes of the emotions. Often, we do not pay attention to the felt sense that the self-organizing pattern is 'off' — i.e., failing to maintain its desired pattern of activity — until something 'triggers' our paying attention to it. E.g., an insult may trigger us to attend to a preconscious sense that something is a little 'off' in our career or love life or interpersonal relations generally, and that it was *these* problems, not the insult per se, that our emotional response was really 'about' (i.e., that determined its real aims and objects). The 'trigger' for paying

attention to the felt sense comes to our attention in just the way that intentional objects normally do — through perception, imaginative acts, or logical-verbal thoughts (e.g., when we decide for philosophical reasons that there is no God, and therefore that death is a more disturbing problem than it had earlier seemed.) In many instances, the function of the trigger is not (unless misinterpreted through hasty identification with an intentional object or cause) to tell us *what is* missing or wrong in our overall self-organizing process, but only to get us to attend to the fact *that* something is missing or wrong. Only *after* the trigger has called our attention to the felt sense of the situation can we begin to formulate a sense of *what* is wrong or missing with the situation, in terms of objects in the environment, by paying attention to the bodily felt sense that the self-organizing process is not optimally maintaining the pattern that it is trying to maintain. Of course, this is not to deny that we may sometimes be so agitated as to take a general cataclysmic action (e.g., fight or flight) prior to any understanding of what the feeling is about, and this tendency may sometimes be adaptive.

But only after paying attention to the felt sense can we begin to determine *what* is wrong. So suppose we do pay attention to the bodily felt sense. If we then attempt to specify an intentional object for the emotion, we are consciously or (more often) preconsciously posing questions to ourselves: First, we ask ourselves what the emotion 'is like,' in Nagel's (1974) sense. At this point, we come up with some kind of 'representation' of what it 'is like.' But we are aware (or can be aware) that the representation is only a metaphor (Gendlin 1997), because the best we can do is to know what the emotion is 'like,' not exactly what it *is* in minute biochemical detail and with precise knowledge of how the process is 'off' from the pattern which is trying to maintain itself, and what aspects of the environment are throwing it off in each little respect; that would be too complicated. So the best we can do is to say "It feels as if I were choking" or "I feel as if I can't breathe" or "It feels the way I felt when I went to bat in the 9th inning of the championship game with the game on the line." The metaphorical description is directing our attention to the way parts of our body feel (a proprioceptive sense), but also 'representing' the way it feels, metaphorically, by focusing on certain repeatable *aspects* of it (e.g., it is 'like' the way I felt on some other occasion or would feel in some type of situation, such as when somebody was choking me — because it is not *literally true* that I can't breathe or that I'm choking, but only that it feels *as if* I were choking, in certain respects). So by representing the felt sense metaphorically, we are selecting some repeatable aspect of it, which is to 'intuit an essence' in Husserl's (1931) oft misunderstood sense *(Wesenschau).* (Husserl does not speak of intuiting *the* essence of an experience, but *an* essence, i.e., an 'essential aspect' of it.)

After having selected certain essential aspects of the felt sense to pay attention to, the *way in which* we pay attention to them is *by means of* representing those

aspects *via* metaphorical imagery of some sort, sometimes visual, sometimes proprioceptive, sometimes verbal. This is how we become conscious of phenomenal properties, or the 'what it's like' which was already implicitly in the body; we are not aware of this phenomenal quality until we direct our attention to it — e.g., the way your shoe feels on your left foot at this moment. When we do attend to it, it presents itself as 'sort of' having already been there, but preconsciously. I say 'sort of,' because the way it would have felt before we paid attention is not exactly the same as the way it does feel when we do pay attention, because the very act of paying attention to something begins to change it a little (Schües 1994). E.g., suppose I have writer's cramp, and I attend to the feeling by 'inhabiting' the hand. As soon as I do, I notice that the tension is not just in the hand, but also in the wrist, and my attention naturally begins to inhabit the wrist, and then it becomes apparent that the tension is up in the shoulder, and the neck, etc., and by then I may realize that now the hand *is no longer* tense — that the very act of attending to the tension has changed what it 'was like'; I then realize that it was the tension in the neck that *produced* the tension in the hand, and that when I let the body be aware of its whole holistic way of holding itself, it spontaneously found *another* way of holding itself that did not involve as much tension.

We become conscious of phenomenal qualities, then, by entertaining metaphorical imagistic representations of our felt sense of the patterns that our body is engaging in, and the aspects of them that feel 'off.' But usually, we also ask ourselves a completely different kind of question: What kinds of elements of the *environment* are connected with that 'offness'; i.e., what elements are there such that, if they could be changed, would allow the organism to optimally engage in the self-organizing pattern that it is trying to engage in? At this point, of course, the tendency of human beings to confuse causal stories with *superstitions* about what causes what may come into play. Just as primitive peoples thought that, if the healing of a wart was accompanied by an eclipse, then it must have been the eclipse that healed it; so modern people are still prone to think that, if the coming-to-our-attention of some 'offness' in our self-organizing pattern was accompanied by some particular 'trigger,' then that specific stimulus must be the element of the environment that is connected with the 'offness' in the sense just defined above — the element such that, if it were changed, the 'offness' would go away. But of course this is not the case with triggers. We may kill the person who triggered our anger, but what was wrong that the bodily felt sense wanted fixed would not have been fixed by this removal of the trigger stimulus.

But we should be clear about the ramifications of this distinction between a trigger and an intentional or causal object. A cookie may trigger the realization that the body wants something, but the cookie is not therefore the environmental

stimulus that our body *'wants'*; the cookie only triggers us to pay attention to the *fact that* we are hungry, and then we can find something nutritious to eat which will get the body going in its optimal patterns again (the cookie may not even be a good object for this purpose).

Often what happens in the process of identifying the objects and aims of affects is that it turns out that we have represented the 'what it's like' aspect of the felt sense *too inaccurately* to allow us to attend to what the real intentional objects and aims are. I.e., we have chosen a metaphorical imaginative act that fails to match enough of 'what it's like.' There is a way we can tell, though. The metaphorical image works to 'pull up' the felt sense, in the way that typing the right code pulls up a computer program. Thus, if entertaining the image makes us feel the felt sense more sharply, we know that the metaphor is 'working' for this purpose. But in the case of 'pulling up' the felt sense, there are degrees of vividness (because we are dealing with fractals and holistic patterns, not linear, *partes extra partes* mechanisms), and the degree of vividness with which the image pulls up the felt sense is the measure of the adequacy of the image for representing the felt sense. This is how we know whether we have more or less adequately 'represented' the felt sense.

We then can proceed to identify the objective features in the environment that are connected with the felt sense (e.g., the sense of 'offness'). By paying attention to our bodies in the past, we have learned what kinds of situations it tends to feel certain ways in; so we can now ask ourselves "How *would* I feel different if this or that aspect of the environmental situation *were* changed?" This involves proprioceptive imagery, in the sense of imagining how I *would* feel under different circumstances from the ones I am now in; and it also involves metaphorical imagery, in the sense that in order to imagine how I would feel different, I have to 'represent' *those* different feelings in the same way that I do with this one, i.e., with metaphorical imagery to focus my attention on the 'what it's like' aspect of them. This is the process that Husserl (1962) calls "imaginative variation." We vary in our imagination aspects of the situation, until we hit on the one such that, if it were varied, our holistic balance would be restored. (Of course there usually is not just *one* aspect that would make everything fine, but with experience at this sort of activity, we can become proficient at sorting it all out.)

What causes a particular *object or situation* to *trigger* an emotion in the first place? It certainly must be that some aspect of the image of that object or situation itself serves somewhat well as a metaphorical image that can serve to call up the bodily felt sense. This is also the same thing as serving as a *'symbolization'* of the felt sense, in the sense of 'symbolization' that Gendlin (1962/1997) uses. I.e., I may symbolize the way I feel toward a love object by playing a piece of music; in this case, the love object is the 'object' of the feeling, and the music is an *embodied action* that is used to intensify and actualize the feeling;

one could also say that the music serves as a 'trigger' for the feeling. But the important point is that the trigger is *not* the essential or primary or complete *intentional object* of the felt sense or emotion. In some cases, it may be an *exemplar* of it, as when a particular insult exemplifies a larger and more important life situation about which we feel frustrated.

Notice that a trigger stimulus also functions as an 'affordance' (Newton 1996); i.e., the trigger stimulus is *analogous* to a type of thing that the desired self-organizing pattern of activity *could* be performed in relation to — e.g., the cookie represents the type of thing that we *could* satisfy hunger in relation to; the person who angered us and whom we are tempted to murder represents the *type* of thing in relation to which we *could* activate murderous or angry feelings. The functioning of trigger stimuli as affordances offers an avenue to grounding an explanation of the 'mechanism' of emotion: As Newton (1996) points out, emotions play a role in action planning, and the 'trigger' offers the possibility of the type of action that could be a possible expression of the emotion.

In principle, only after accurately describing the sequence of phenomenological events should we expect to be able to identify their physiological correlates. The intentionality of emotions is too complex to allow simply identifying the object of an emotion with the stimulus that evokes it. The intentionality of an organism's emotions points to the organism's overall relationship to its environment as it seeks opportunities to enact the patterns of organization that its own self-maintaining structure motivates it to enact.

We often think of emotions as being like pain sensations, but in reality the differences between emotions and pain phenomena are vast. Emotions represent holistic organismic conditions, whereas pains represent highly specific disturbances. Correlatively, pains have highly specific receptors, the 'nociceptors,' whereas emotions are sensed by means of whole-body processes. Pains are immediately directed to specific parts of the body and reflect specific environmental effects, whereas emotions are amorphous and ubiquitous.

I have argued that emotion and motivation are keys to understanding consciousness in a way that is accessible to both phenomenological and neurophysiological methods. It is crucial to appreciate the complexity of the phenomenology and intentionality of emotions before moving into the neurophysiological investigation of them. I have suggested that a self-organizational or dynamical systems approach can help us to conceptualize the motivational dimension in a way that is helpful toward an understanding that is equally commensurable with neurophysiological and phenomenological information. The remaining papers in this volume will focus on these motivational and affective processes, and the way they interrelate with cognitive processes, from both experiential and neuroscientific perspectives. Many of the contributions are geared toward integrating the two perspectives.

References

Ahern, G.L. and G.E. Schwartz. 1985. "Differential Lateralization for Positive and Negative Emotion in the Human Brain: EEG Spectral Analysis". *Neuropsychologia* 23: 745–755.

Alexander, David, and Gordon Globus. 1996. "Edge of Chaos Dynamics in Recursively Organized Neural Systems". In Earl MacCormac and Maxim Stamenov (eds), *Fractals of Brain, Fractals of Mind.* Amsterdam: John Benjamins.

Bickhard, Mark. 1993. "Representational Content in Humans and Machines". *Journal of Experimental and Theoretical Artificial Intelligence* 5: 285–333.

Bickle, John. 1992. "Multiple Realizability and Psychophysical Reduction". *Behavior and Philosophy* 20: 47–58.

Center for the Scientific Study of Consciousness. 1998. *Consciousness Research Abstracts: Toward a Science of Consciousness, Tucson III.* Tucson: Center for Consciousness Studies, University of Arizona).

Chalmers, David. 1995. "Facing Up to the Problem of Consciousness". *Journal of Consciousness Studies* 2: 200–219.

Cohen, R.M., W.E. Semple, M. Gross, H.J. Holcomb, S.M. Dowling, and T.E. Nordahl. 1988. "Functional Localization of Sustained Attention". *Neuropsychiatry, Neuropsychology and Behavioral Neurology* 1: 3–20.

Cytowic, Richard. 1993. *The Man Who Tasted Shapes.* NY: Warner.

Damasio, Antonio. 1989. "Time-locked Multiregional Retroactivation: A Systems Level Proposal for the Neural Substrate of Recall and Recognition". *Cognition* 33: 25–62.

Damasio, Antonio, and G.W. Van Hoesen. 1983. "Emotional Disturbances Associated with Focal Lesions of the Limbic Frontal Lobe". In Kenneth Heilman and Paul Satz (eds), *Neuropsychology of Human Emotion.* New York: Guilford Press.

Damasio, Antonio, P.J. Eslinger, H. Damasio, G.W. Van Hoesen, and S. Cornell. 1985. "Multimodal Amnesic Syndrome Following Bilateral Temporal and Basal Forebrain Damage". *Archives of Neurology* 42: 252–259.

Davidson, Donald. 1970. "Mental Events". In Lawrence Foster and Joe W. Swanson (eds), *Experience and Theory.* Amherst: University of Massachusetts Press, 79–102.

Dennett, Daniel. 1995. *Consciousness Explained.* Boston: Little, Brown and Co.

Dennett, Daniel. 1996. *Kinds of Minds*. New York: Basic Books.

Dore, John, Margery Franklin, Robert Miller, and Andrya Ramer. 1976. "Transitional Phenomena in Early Language Acquisition". *Journal of Child Language* 3: 13–27.

Ellis, Ralph D. 1986. *An Ontology of Consciousness.* Dordrecht: Kluwer/Martinus Nijhoff.

Ellis, Ralph D. 1995. *Questioning Consciousness: The Interplay of Imagery, Cognition and Emotion in the Human Brain.* Amsterdam: John Benjamins.

Ellis, Ralph D. 1996. "Ray Jackendoff's Phenomenology of Language as a Refutation of the Appendage Theory of Consciousness". *Pragmatics & Cognition* 4: 125–137.

Ellis, Ralph D. 1999a. "Dynamical Systems as an Approach to Consciousness: Emotion, Self-organization, and the Mind-Body Problem". *New Ideas in Psychology* 17: 237–250.

Ellis, Ralph D. 1999b. "Integrating Neuroscience and Phenomenology in the Study of Consciousness". *Journal of Phenomenological Psychology* 30: 18–47.

Ellis, Ralph D. 1999c. "Why Isn't Consciousness Empirically Observable?" *Journal of Mind and Behavior* 20: 391–402.

Ellis, Ralph D. 2000. "Consciousness, Self-organization, and the Process-Substratum Relation: Rethinking Nonreductive Physicalism". *Philosophical Psychology* 13: 173–190.

Ellis, Ralph D. and Natika Newton. 1998a. *Consciousness and the Brain: An Annotated Bibliography.* http://earthlink.net/~dravita.

Ellis, Ralph D. and Natika Newton. 1998b. "Three Paradoxes of Phenomenal Consciousness: Bridging the Explanatory Gap". *Journal of Consciousness Studies* 5: 419–42.

Ellis, Ralph D. and Natika Newton. In press. "The Interdependence of Consciousness and Emotion". *Consciousness & Emotion.*

Eysenck, Hans J. 1957. *The Dynamics of Anxiety and Hysteria.* New York: Praeger.

Flor-Henry, Pierre, L. T. Yeudall, Z. T. Koles, and B. G. Howarth. 1979. "Neuropsychological and Power Spectral EEG Investigations of the Obsessive-compulsive Syndrome". *Biological Psychiatry* 14: 119–129.

Freeman, Walter. 1975. *Mass Action in the Nervous System.* New York: Academic Press.

Freeman, Walter. 1987. "Simulation of Chaotic EEG Patterns with a Dynamic Model of the Olfactory System". *Biological Cybernetics* 56: 139–150.

Freeman, Walter. 1988. "Why Neural Networks don't Yet Fly: Inquiry into the Neurodynamics of Biological Intelligence". *Proceedings of the IEEE International Conference on Neural Networks* Vol. 2. Dan Diego: IEEE, 1–7.

Freud, Sigmund. 1959. *Beyond the Pleasure Principle*. New York: Bantam Books.

Gainotti, Guido, Carlo Caltagirone, and P. Zoccolotti. 1993. "Left/right and Cortical/subcortical Dichotomies in the Neuropsychological Study of Human Emotions". *Cognition and Emotion* 7: 71–93.

Gendlin, Eugene. 1962/1997. *Experiencing and the Creation of Meaning.* Chicago: University of Chicago Press, 2nd ed.

Gendlin, Eugene. 1971. "A Theory of Personality Change". In Alvin Mahrer (ed), *Creative Developments in Psychotherapy.* Cleveland: Case Western Reserve University Press.

Gendlin, Eugene. 1973. "Experiential Phenomenology". In Maurice Natanson (ed), *Phenomenology and the Social Sciences.* Evanston: NorthwesternUniversity Press.

Gendlin, Eugene. 1992a. "The Primacy of the Body, not the Primacy of Perception". *Man and World* 23: 341–353.

Gendlin, Eugene. 1992b. "Thinking Beyond Patterns: Body, Language, and Situations". In B. den Ouden and M. Moen (eds), *The Presence of Feeling in Thought.* New York: Peter Lang.

Gibson, J. G., and W. A. Kennedy. 1960. "A Clinical-EEG Study in a Case of Obsessional neurosis". *Electroencephalography and Clinical Neurology* 12: 198–201.

Globus, Gordon. 1992. "Towards a Noncomputational Cognitive Neuroscience". *Journal of Cognitive Neuroscience* 4: 299–310.

Globus, Gordon. 1995. *The Postmodern Brain.* Amsterdam: John Benjamins.

Goldman, Alvin. 1969. "The Compatibility of Mechanism and Purpose". *Philosophical Review* 78: 468–482.

Goldman, Alvin. 1970. *A Theory of Human Action*. Englewood Cliffs: Prentice-Hall.

Goleman, Daniel. 1997. *Emotional Intelligence*. New York: Bantam.
Hardcastle, Valerie. 1996. *Locating Consciousness.* Amsterdam: John Benjamins.
Hernandez-Peon, Raul, G. Chavez-Iberra, and E. Aguilar-Figuera. 1963. "Somatic Evoked Potentials in One Case of Hysteric Anesthesia". *Electroencephalography and Clinical Neurophysiology* 15: 889–896.
Horgan, Terrence. 1984. "Jackson on Physical Information and Qualia". *Philosophical Quarterly* 34: 147–152.
Husserl, Edmund. 1931. *Ideas*. W. R. Boyce Gibson (trans.) London: Collier 1931; from "Ideen zu einer Reinen Phänomenologie und Phänomenologischen Philosophie", 1913).
Husserl, Edmund. 1900/1962. *Phänomenologische Psychologie.* Den Haag: Martinus Nijhoff.
Jackendoff, Ray. 1987. *Consciousness and the Computational Mind*. Cambridge: The MIT Press.
Jackendoff, Ray. 1996. "How Language Helps Us Think". *Pragmatics & Cognition* 4: 1–34.
Jackson, Frank. 1986. "What Mary Didn't Know". *Journal of Philosophy* 83: 291–295.
Kauffman, Stuart. 1993. *The Origins of Order*. Oxford: Oxford University Press.
Kim, Jaegwon. 1992. "Multiple Realization and the Metaphysics of Reduction". *Philosophy and Phenomenological Research* 52: 1–26.
Kim, Jaegwon. 1993. "The Nonreductivist's Troubles with Mental Causation". In J. Heil and A. Mele (eds), *Mental Causation*. Oxford: Oxford University Press, 189–210.
LeDoux, Joseph. 1996. *The Emotional Brain.* New York: Simon and Schuster.
Luria, Alexander R. 1980 *Higher Cortical Functions in Man,* 2nd ed. New York: Basic Books.
Mac Cormack, Earl. 1996. "Fractal Thinking: Self-organizing Brain Processing". In Mac Cormack and Stamenov 1996.
Mac Cormack, Earl and Maxim Stamenov (eds). 1996. *Fractals of Brain, Fractals of Mind.* Amsterdam: John Benjamins.
Macrides, Foteos, H. B. Eichenbaum, and W. B. Forbes. 1982. "Temporal Relationship between Sniffing and the Limbic Theta Rhythm during Odor Discrimination Reversal Learning". *Journal of Neuroscience* 2: 1705.
Maslow, Abraham. 1954. *Motivation and Personality*. New York: Harper & Row.
Merleau-Ponty, Maurice. 1942/1963. *The Structure of Behavior.* A. Fischer (trans.). Boston: Beacon.
Miller, Lawrence. 1990. *Inner Natures: Brain, Self and Personality.* New York: Ballantine.
Monod, Jacques. 1971. *Chance and Necessity.* New York: Random House.
Muller, Herbert. 1998. "Primacy of Experience, and Working Metaphysics". In Owner-jcs-online@lists.zynet.co.uk. See also mdmu@MUSICA.MCGILL.CA.
Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83: 435–450.
Natsoulas, Thomas. 1993. "What is Wrong with Appendage Theory of Consciousness?" *Philosophical Psychology* 6: 137–154.
Nauta, Walle J. 1971. "The Problem of the Frontal Lobe: A Reinterpretation". *Journal of Psychiatric Research* 8: 167–187.

Neisser, Ulric. 1994. "Ecological Psychology". Lecture at Southern Society for Philosophy and Psychology, Atlanta, Georgia.

Newton, Natika. 1996. *Foundations of Understanding*. Amsterdam: John Benjamins.

Petersen, S. E., P. T. Fox, A. Z. Snyder, and M. E. Raichle. 1990. "Activation of Extrastriate and Frontal Cortical Areas by Visual Words and Word like Stimuli". *Science* 249: 1041- 1044.

Posner, Michael I. and Mary K. Rothbart. 1992. "Attentional Mechanisms and Conscious Experience". In A. D. Milner and M. D. Rugg (eds), *The Neuropsychology of Consciousness*. London: Academic Press.

Pribram, Karl. 1980. "Mind, Brain, and Consciousness: the Organization of Competence and Conduct". In Julian Davidson and Richard Davidson (eds), *The Psychobiology of Consciousness*. New York: Plenum Press, 47–64.

Quine, W. V. O. 1960 *Word and Object*. Cambridge, Mass.: MIT Press, 33ff.

Sartre, Jean Paul. 1971. *Sketch for a Theory of the Emotions*. London: Methuen.

Schües, Christina. 1994. "The Anonymous Powers of the Habitus". *Study Project in the Phenomenology of the Body Newsletter* 7: 12–25.

Searle, John. 1984. *Minds, Brains and Science*. Cambridge: Harvard University Press.

Skinner, B. F. 1953. *Science and Human Behavior*. New York: Macmillan.

Spence, K. W., and J. T. Spence. 1964. "Relation of Eyelid Conditioning to Manifest Anxiety, Extraversion and Rigidity". *Journal of Abnormal and Social Psychology* 68: 144–149.

Stamenov, Maxim. 1995. "The Fractal-like Roots of Mind: A Tutorial in Direct Access". In Earl Mac Cormack and Maxim Stamenov (eds), *Fractals of Brain, Fractals of Mind*. Amsterdam: John Benjamins, 273–322.

Thelen, Esther, and Linda Smith. 1994. *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge: MIT Bradford.

Tienson, John. 1987. "An Introduction to Connectionism". *Southern Journal of Philosophy* 26: 1–16.

Varela, Francisco, Evan Thompson, and Eleanor Rosch. 1993. *The Embodied Mind*. Cambridge: The MIT Press.

Watson, John. 1900/1930. *Behaviorism*. Chicago: University of Chicago Press.

Watson, John. 1913. "Psychology as the Behaviorist Views It". *Psychological Review* 20: 157-158.

Watt, Douglas. 1998. "Affect and the 'Hard Problem' Neurodevelopmental and Cortico-limbic Network Issues". *Consciousness Research Abstracts: Toward a Science of Consciousness, Tucson 1998:* 91–92.

Watt, Douglas. In press. "The Centrecephalon and Thalamocortical Integration: Neglected Contributions of Periaqueductal Gray". *Consciousness & Emotion*.

White, Robert. 1965. "Motivation Reconsidered". *Psychological Review* 65: 297–333.

CHAPTER 2

Affective Consciousness and the Instinctual Motor System

The Neural Sources of Sadness and Joy

Jaak Panksepp
Bowling Green State University

1. On the varieties of consciousness

How many fundamental forms of consciousness exist in the human brain? No one knows, but along with some others, I prefer the answer "more than one." Considering that Freud (1923) once argued for three distinct forms of unconsciousness, it should not seem far-fetched that several distinct types of consciousness may emerge from brain activities. Indeed, Block (1995) has compelling argued for a distinction between *phenomenal* and *access* consciousness. Obviously, consciousness has to be comprised of a multiplicity of brain processes working in harmony, but split-brain studies indicate how effectively a seemingly unitary form of consciousness can be divided in two (Springer and Deutsch 1998). Indeed, new psychotherapeutic insights have emerged from the recognition that the division of consciousness in split-brain studies may reflect a fundamental duality of consciousness in waking life (Schiffer 1998). However, since those examples may simply constitute surgical and epigenetic divisions of a singular type of brain process, a more interesting issue is: Are there really several categorically distinct *types* of neural transactions within each human brain, indeed perhaps each mammalian brain, from which subjective awareness can arise?

Besides dividing consciousness into right and left brain varieties, we may also be able to divide it reasonably, albeit coarsely, into dorsal (cortical) and ventral (subcortical) varieties that roughly correspond to cognitively focused forms (perceptions and thoughts) and affectively focussed ones (emotions and feelings). Of course, this division would not be independent of right-left divisions, for it is evident that the consciousness of the left hemisphere is more

logical-linear-propositional, while that of the right hemisphere is more affective-holistic-impressionistic. But the dorsal-ventral dimension goes much deeper into the classical distinction between reason and the passions. The dorsal-ventral forms of awareness, even more than the right-left ones, affirm the qualitative differences that we recognize as highly valenced affective states aroused by deeply meaningful life events as opposed to our fairly neutral everyday perceptual experiences and thoughts as we navigate the world. Cognitive science has largely been based upon an analysis of the latter (how we perceive objects and actions and their relationships in space and time); because of many obvious methodological and conceptual difficulties, it has generally avoided the raw emotional side of life. Advances are presently being made on the more gentrified forms of emotions, but few are working on the ancestral underbelly that will be my main concern here. In Freudian terms, I will be concerned with the nature of the id — the instinctual drives and the psychological surface that faces toward the inner world.

It is especially difficult to agree how ancient evolutionary representations such as our basic feeling states should be discussed and studied. The sad solution, for too many scholars of the mind, has been to ignore them altogether. That is an intellectual tragedy, for it fails to acknowledge some of the most important, most valuable and painful birthrights of our human existence. Thus emotions have remained a scientific and philosophical puzzle of enormous magnitude — a challenge that we eventually must confront in evolutionary and psychobiological terms as we increasingly recognize that all our evolutionarily close brethren also have feelings. Affective consciousness is not simply constructed from our more recently evolved cognitive capacities but from our ancestral heritage of body and brain working together as a cohesive unit capable of anticipating future events in valenced ways. Only our master skill — the liberating capacity of symbolic thought (or from another vantage, the prisonhouse of language) — coaxes us to believe otherwise, through our uniquely anthropoid capacity to be aware of awareness, yielding abstractions which no other species appears to appreciate. My position will be that in order to understand the evolution of consciousness, we must come to understand emotions in more purely neuro-affective and neuro-behavior terms.

In any event the current range of educated opinion in the area extends from the majority of behavioral neuroscientists who would prefer not to entertain the possibility that other mammals have feelings (e.g., LeDoux 1996, 1999) to those, like myself (Panksepp 1982 1998a), who think that viewpoint is misleading and hence, in the long run, counterproductive. Damasio (1999a, p. 39) has recently commented on this sad aspect of the field, especially "neuroscience's reluctance to accept that complex nonhuman creatures have feelings an attitude that goes beyond the necessary prudence over the fact that such creatures may or may not

know they have such feelings." To put it mildly! We should never forget that poking other animals with a sharp pin will make them go "ouch" (in their own species-typical way) as readily as do humans. When we share our basic pleasures of life with them, they also seem to take delight.

In my estimation, there are many reasons to believe that the internally experienced affective values of the nervous system are created from much older subcortical circuits than those that mediate cognitive awareness. By this I do not mean to imply that the various affective systems of the brain are not closely linked to immediate life experiences of the organism, but to assert that evolution left powerful neural residues of highly generalized ancestral memories that remain genetically encoded as affective potentials within our brains and the brains of many other animals. Although the traditional distinction between affective and cognitive processes is presently not a popular academic stance, there is a mountain of evidence to suggest the dichotomy is a meaningful way to parse neuro-mental space. This view in no way seeks to deny their remarkable blending in our first-person subjective experiences, nor the fact that cognitive abilities have co-evolved with affective processes in many higher regions of the brain (e.g., anterior cingulate, orbitomedial frontal and periamygdaloid temporal cortices). Those are self-evident phenomenological truths and established neurological facts. However, in other examples of co-evolution, for instance predator-prey relations, it is always essential to distinguish species rather than conflate them. I think the same rule should apply to the distinct, albeit interactive, functions of the brain.

The acceptance of such dichotomous dimensions of mind can provide penetrating new perspectives on various long-standing problems in consciousness studies — e.g. the nature of 'the self,' mind-brain dualism, and the explanatory gap. Accordingly, my goal here is to refresh the classic idea that brain tissues which generate the basic affective feeling states can be credibly distinguished from those that generate exteroceptive/cognitive consciousness. I envision these global types of consciousness to relate to each other in figure-ground relationships, with the ancient affective processes, linked more closely to central nervous system (CNS) somatomotor and visceral processes, providing the essential ground states from which more recent cognitive skills, linked more closely to exteroceptive events, have emerged. Thus, the premise I would like to develop is that cognitive abilities are built upon affective foundations. Some might claim that there is some type of cognitive primacy to consciousness since emotions are typically transient, but there are some emotional systems like the SEEKING system of the brain that sustain tonic levels of interest, investigatory activities and an urge to make sense of the world (for overview, see Chapter 8; Ikemoto and Panksepp, 1999).

Of course, one great difficulty in dividing consciousness into distinct forms

arises from the fact that all forms rely on some common brain mechanisms, such as the ascending reticular arousal and attentional systems (Baars 1997; Newman 1997). Another arises from the fact that concepts can be blended, transformed and democratized by our plastic linguistic abilities, making all strict criteria and conceptual dividing lines intrinsically debatable. Notwithstanding these difficulties and in line with a great deal of traditional thinking, as well as compelling lines of brain research, the distinction is not only heuristic but meaningful. Substantial parts of the neural apparatus for our fundamental emotional feelings are demonstrably different from those that mediate our exteroceptive perceptions and thoughts. We can see these differences in the firing rates of neurons (slow vs. fast, respectively), the neurochemistries (long-acting peptides vs. short-acting amino acids), as well as the locations of key neuroanatomies (subcortical vs. cortical).

The two are also phenomenologically different, emotions being powerfully felt in the body and being psychologically all encompassing (they seem to be part of the very 'self'), while cognitions are more spatio-temporally resolved and are not as experientially all-consuming. In a sense, this distinction is a classic one that was embedded in the classical notion that all sensations have primitive *protopathic* components, characterized by intense, poorly localized feelings, and more recently evolved *epicritic* components that provide detailed cognitive resolution to sensation (Jackson 1884). It was assumed that the latter were built, in some fundamental way, on the former. The proposition that the ancient emotional mechanisms have powerful effects on the cognitive apparatus is being repeatedly confirmed by modern brain imaging techniques. Using global measure of neuronal arousal in the brain via *c-fos* immunohistochemistry, investigators find that emotional arousal leads to a massive activation of the cortico-cognitive apparatus (Beck and Fibiger 1995; Campeau *et al.* 1997; Kollack-Walker 1997).

We must now come to terms with such fundamental brain issues if we are ever to understand the nature of primary-process consciousness. Regrettably, the neglect of emotions and the ascent of cognitive issues in consciousness studies has created a rather misleading picture of the way awareness may be created within the CNS. Our capacity for feelings may well have been the fountainhead of consciousness in brain evolution. This should be self-evident if we view consciousness as based ultimately on the nature of our basic biological values.

If this premise is correct, consciousness must have sprouted from some primordial adaptive functions of the brain that we still share with other creatures. That should come as no surprise to those who subscribe to evolutionary principles. After all, evolution works best with body parts and processes that already exist. There is every reason to believe that biological values were encoded in brain systems and simple response tendencies before complex capacities evolved which allowed organisms to cognitively navigate their material and social terrain. It may be a bit more surprising to contemplate the possibility that what we

presently experience as routine waking consciousness, with its rich sensorial awareness, may arise from neural processes that governed emotional motor actions in ancestral species rather than those that simply sensed the world in more neutral ways. We will consider this possibility here, for our basic emotions are still intimately intertwined with the generation of distinct and powerful forms of action readiness. It is possible that the raw primal feeling states we experience as a result of such emotional tendencies are strongly linked to, perhaps even caused by, those readiness potentials. This type of powerfully self-referenced, neural organization may have served as a pre-adaptation for the emergence of ever more complex exteroceptive and cognitive forms of consciousness.

Most prevailing theories of emotions — the constructivist, appraisal, propositional attitude, and componentialist views — are typically based on fairly high-level cognitive analyses of emotions. This does not make them incorrect, but as a result they have little to say about the raw affective domain where our deepest emotional existence is lived. Perhaps not surprising, all too often theorists ignore and minimize those forces in human life. Most people go out of their way to avoid powerful negative feelings, the fundamental despairs of our lives — feelings of coldness, fear, hunger, rage, thirst, loneliness, and the other varieties of pain.

So far only the categorical approaches that accept the validity of distinct primitive passions — the various affect programs of the brain — have provided a glimpse into those primal entities of mind. At present, behavioral brain research on the underlying circuits offers the most robust empirical gateway to understanding what emotions and feelings really are (Panksepp 1982 1991). The rest, except for a depth-psychology that has rarely been scientifically cultivated in the past half century (for a striking exception, see Solms 1997), is largely a description of surface complexities, which in humans can be extraordinarily interesting — remarkably subtle but all too often deceptive (Friedlund 1994). Indeed, surface behavioral appearances may more accurately reflect the dynamics of the underlying emotional brain processes in animals than in humans. The body and the emotional mind work as a unit, as long as the cortico-cognitive mind does not intervene with its own interesting agendas that emerge from added layers of awareness and escalating cognitive complexities that permit interesting emotional repressions and other regulations. Although the notion that emotional behaviors, such as facial expressions, are veridical readouts of emotional states has been criticized (Friedlund 1991), it is generally agreed that there is some useful information in such external signs, even in adult humans who can strategically utilize socio-cultural display rules (Buck 1994; Scherer, Wallbott and Summerfield 1986). I think most reasonable people would agree that the amount of veridical readout about emotional states within external surface appearances diminishes as one goes from animals and young children to older

humans, and from people who care about and know each other well to those who do not.

The dichotomy between cognitive and affective consciousness developed here does not seek to exclude other possible configurations. Such basic distinctions as those between perceptions and thoughts are obviously useful in consciousness studies. As already mentioned, the dichotomy also does not imply that affective and cognitive forms of consciousness do not normally operate cooperatively in the well integrated brain-mind. The distinction is neuroscientifically, phenomenologically and psychobiologically meaningful and has some clear advantages in promoting a cogent science of consciousness. To really understand emotions, we must recognize the importance of evolutionary global state variables, which I have called 'emotional qualia' — 'equalia' arising from emotional evolution[1] — in the construction of mind. I.e., consciousness is not simply created from the exteroceptively triggered varieties of individual experiences; it is based on ancient neural matrices that symbolized the organism as an active, feeling creature of the world. Evidence supporting these assertions has been extensively documented (Panksepp 1998a). If true, focussing on sensory systems like vision may not take us very far into fathoming the mystery of consciousness (Crick 1994).

In sum, from the present perspective, cognitive consciousness, with its many thoughts and perceptions, is directed toward the external world, largely in a sensory manner. Its highest propositional forms are integrally intertwined with neocortical, especially left hemisphere, functions. Affective consciousness arises fundamentally from the inward-directed CNS functions that evolution provided in the form of value-action compasses, largely constructed from distinct types of interoreceptive visceral and instinctual motor readiness systems; it is also more powerfully constrained by limbic and medial brainstem activities which may have especially strong influences on the semantically impoverished right hemisphere.

A basic premise is that in brain evolution the affective distance between humans and related animals is smaller than the cognitive distance. In traditional Freudian terms, affective consciousness is related more to id functions, while cognitive consciousness is related to ego and superego functions (Panksepp 1999b, c). Interactions between the two occur at many levels within the brain, but it will be worth considering that the prior emergence of the equalia of affective consciousness in brain evolution served as the foundation for the emergence of the more resolved qualia of cognitive consciousness. Conversely, the emergence of cognitive consciousness allowed organisms to better regulate emotional/affective processes and to utilize them as informational tokens within its linguistic-type deliberations. However, emotions are much more than informational tokens or signals; they are intrinsically valenced. They have many genetically coded attributes of goodness and badness that attract and repel us. In

those capacities they control our basic intentions that only gradually mature into propositional attitudes as they interact with real world events.

While cognitive consciousness now helps us navigate through the jungle of highly resolved information provided by the externally directed senses, affective consciousness elaborates the raw feels and consequent value structures that still inform us of primal survival issues, ingrained in brain systems constructed by our genetic heritage — from our experience of various anxieties and hungers, to our peak experiences, whether orgasmic or more prolonged satisfactions. While cognitive consciousness is informationally well resolved and capable of being simulated through the traditional modeling approaches of artificial intelligence, affective consciousness is more global and nebulous, more urgent and organic, and less capable of being instantiated computationally (Picard 1997). To some extent, affective processes are probably active every waking moment of our lives. Although there are bound to be many sub-aspects of consciousness within each of these two global categories (the many distinct varieties of emotions, and the various modes of perception, attention and thought), it is important to restore the broad affective-visceral and cognitive-somatic categories to our modes of thinking if we are going to make headway on a variety of important empirical issues in neuroscience and psychology, and parallel conceptual issues in philosophy and consciousness studies. Without trying to minimize the many complexities that this distinction hides, I will try to highlight what affective consciousness may look like within the nervous system. But first, let me pause to share a bit of data on the issues just covered.

2. Interlude: A small survey of opinion concerning the varieties of consciousness

To my knowledge, no one has ever simply asked people about their viewpoints on the existence of multiple forms of consciousness. Accordingly, I asked 32 people (about two-thirds female and one third male), in one of my upper level undergraduate/graduate classes, the following two simple binomial questions, yielding a 2×2 contingency table of attitudes on this issue: (1) "No one knows how many distinct forms of consciousness exist in the human brain/mind. Some believe there is only one fundamental form of consciousness, while others believe there are two or more, such as those that mediate emotions and those that mediate cognitive activities. If you were going to take sides on this issue, which side do you believe is more correct? One, or more than one?" (2) "We are all individuals with various strengths, both cognitive and emotional. If you were asked to decide whether you deem yourself to be more fundamentally a cognitive individual or an emotional one, to which camp would you tend to subscribe?

Emotional or cognitive?" This was done at the beginning of a semester before I had a chance to mold their minds with my own views. The results of this 2 × 2 contingency relationship were as shown in Table 1:

Table 1.

Type (Question 1 — Question 2)	N	% of total	%male
Two Forms — Emotional	17	53.1	14%
One Form — Emotional	4	12.5	25%
Two Forms — Cognitive	4	12.5	50%
One Form — Cognitive	7	21.9	57%

These results indicated that more than half of these students viewed themselves primarily as emotional creatures who believe they have at least two fundamental forms of consciousness, a viewpoint with which I am prone to agree. Of course, this trend may have been partially due to the prevalence of females in the sample. Among those who took the mono-conceptual cognitive view, males prevailed. No big surprise, perhaps. Although this was a modest sample, it suggests why academic writing on the topic tends to be exceedingly 'mono-theistic' with respect to the essential role of cognitive processes in the creation of consciousness — a predominantly male viewpoint. It is probably the case that males as a group do fit a conventional stereotype — they pride themselves on their logic and are less in tune with their emotional lives than females. I suspect this response bias might be even more skewed in the cognitive direction among philosophers and perhaps in the affective direction among individuals more devoted to the arts.

I also asked one additional simple question: Since emotions are complex processes of the brain, with many attributes, how would you prioritize the following five aspects from least to most important, if we are ever going to scientifically understand emotions: namely, the study of (i) facial, (ii) vocal, (iii) feeling, (iv) cognitive or (v) autonomic aspects of emotions? The results were essentially identical to the previous time that I asked this same question (see Figure 1, Panksepp 1999b, p. 17), with 'feelings' being the most important dimension to be studied, followed closely by 'cognitive' and 'autonomic' processes, while 'facial' and 'vocal' dimensions of emotions were generally deemed to be least important. Thus, among our students, affective processes are deemed of foremost importance for understanding the nature of emotions. I suspect this folk-psychological point of view is closer to the heart of the matter than are many currently popular academic views. Indeed, "knowledge by acquaintance" is a more powerful and insistent source of knowledge that "knowledge by description" (Buck 1999), but it is also more fuzzy and fleeting

since it is so intimately linked to our first-person experiences. Science can more easily probe descriptions of the world based on our more 'objective' exteroceptive senses, especially vision and hearing, than those based on our inner affective experiences. However, to make sense of the brain, we must begin to develop new methodologies that can bring subjective and objective issues together in empirically fruitful ways.

3. A short history of why progress in understanding affective processes has been so slow

From such results, one might expect that there would be a widespread hunger to scientifically understand the neural nature of affective experiences, but that has not been the case in mainstream psychology or neuroscience. It has been no different in modern philosophy.[2] The reasons for the neglect of affect are not difficult to surmise: We are all captivated by the vast variety and richness of our sensory experiences and our resulting thoughts, which we can often convey with exceptionally cool-headed linguistic grace and clarity. Most of our deepest feelings are notoriously global and diffuse in our consciousness, and hence difficult to put into words. To really capture the power of affect, we often need to use poetic language and metaphoric images, neither of which is a respected mode of communication among the rigorously disposed. Thus, for most scholars, emotions have remained as mysterious and inexplicable as Freud left them three score and several years ago.

Our failure to try to deal with the fundamental nature of affect is a pity, for it has great potential to provide a credible solution to the mind-body problem: Affective consciousness may be a primitive, subcortically 'embodied' mass-action form of consciousness, emerging fundamentally from our instinctual body representational apparatus, that we share homologously with many other animals (Panksepp 1998b). In other words, at an affective level, mind and body are a unified process. Unfortunately we are accustomed to thinking that motor outputs are just that, mere outputs, and that our central motor representations in the brain are not very important for how we internally feel. I think this is a shortsighted view promoted by our cognitive tendency to focus on somatic sensations as being more psychologically interesting than motor processes. However, in the case of primary emotions, central motor processes may be more essential for creating feeling states than sensory processes.[3] This is not to minimize the importance of sensory aspects in sophisticated creatures, but to simply recognize the neglected importance of a self-referential motor system in establishing the foundation upon which sensorial systems could be built. I would go so far as to suggest that consciousness cannot be simulated without taking those features into account.

Emotional affects may largely be neural representations of various bodily states, not only of the corporeal body but also the fundamental neurosymbolic representations of the body within the brain — perhaps taking the form of quite primitive visceral and somatic self-representations in ancient regions of the brain (Panksepp 1998b). At a foundational level, such representations of body and the establishment of a neural potential for affective consciousness may be one. The dynamics of the later supervenes on the former.

Only with the emergence of cognitive consciousness, and the eventual development of higher order symbolic abilities, did the massive divide between mind, body and brain emerge in the minds of thinkers. However, if we destroy the affective foundations of the brain, the mind collapses. On the other hand, if we destroy the cognitive components, then organisms simply become inept and stupid, but they are still affectively *in* the world. In fact, decorticate animals appear to be considerably more emotional than normals, but some would claim that their hyperemotionality simply reflects affectively meaningless motor displays. That, however, is only a supposition. No one has adequately probed affective capacities of decorticate animals in needed ways, but we do know that their instinctual motor apparatus is intact (Kolb and Tees 1990; Panksepp, Normansell, Cox and Siviy 1994).

Some would insist that all we can really know about in this domain are animal movements, and nothing about the potential for feelings in their brains. That is wrong. We can know a great deal about their affective responses, through the study of various instinctual behaviors as well as conditioning and preference tests, and in so doing, we can learn much about our own basic feelings. We can reach credible cross-species conclusions, as long as we are willing to triangulate among all relevant sources of information — the neuroscientific and the behavioral data that animal models can provide, and the affective verbal reports that humans can share (Panksepp 1998a). Unfortunately, that project remains in its infancy. For brain scientists not to accept other animals, especially the other mammals, into our shared affective circle is rapidly becoming an untenable position. It is inconsistent with our modern understanding of the genetic underpinning of mind and the epigenetic psychic landscapes that emerge from the interaction of those potentials with world events. At present, even the agnostic view — that we can never know about these matters with certainty — is only sustaining the hegemony of an outdated intellectual tradition, namely radical behaviorism, that continues to impede a scientific understanding and appreciation of human and animal emotions. However, because of the neuroscience revolution, we can now focus on powerful emotion-related predictions that can be made across species, and in that way, come to truly understand the biological nature of our own feelings for the first time. However, this does require committment to new conceptual approaches.

In my estimation cognitive science, even of the neuroscience variety, has not yet developed optimal frameworks for decoding the nature of global central processes such as affect, even though it is excellent in specifying how cognitive inputs get access to emotional systems (LeDoux 1996). A different conceptual approach, which I have called "affective neuroscience" may help us properly conceptualize the full richness of those global affective processes of the mammalian brain which, rather spontaneously (i.e., through the auspices of past evolutionary selection) help establish each animal's psychobehavioral priorities (Panksepp 1991 1998a). This requires us to work with animal and human models of emotionality concurrently. The animal work can provide testable hypotheses of how emotions are organized in the brain, and human phenomenological reports can verify whether these hypotheses are correct. The underlying assumption here is that basic emotional value systems are homologously organized in all mammalian species.

Of course, as soon as these value systems become operative, they begin to construct and weave an infinitely individualistic approach to the world through learning (Rolls 1999; Panksepp 2000a). Hence, in the mature organism, we rarely see their operations independently from the cognitive structures that they have helped construct. However, in the animal laboratory we can see 'pure' emotional phenomena repeatedly if we electrically or chemically stimulate specific parts of the brain. The aim of the rest of this paper is to discuss how feelings might actually be constructed in the brain, and to specifically explore two of the most important and least studied emotions of which humans and animals are capable — sadness and joy.

4. Theories of how affective feelings are created in the brain

Darwin (1872/1998) was among the first to enumerate the fundamental properties of emotional systems. He suggested three principles: (i) that each basic emotional system of the brain arises from brain activity (i.e., his *principle of action*, due to the constitution of the nervous system), (ii) interacts in specifiable ways with other systems (his *principle of antithesis*) and (iii) that the arousal of the various emotional states guide learning (his *principle of serviceable associated habits*). He also explicitly recognized that key features of all emotions are distinct internally experienced feelings, but he chose to neglect those aspects since they were impossible to systematically analyze with the tools available in his era. The task is much easier today because of the precise behavioral and neuroscientific tools that we have at our disposal.

At present only a few prominent investigators believe that a key issue of emotion research as well as consciousness studies should be a clarification of

how brains construct emotional feelings (most prominently, MacLean 1990). It is surprising how few hypotheses along those lines have been entertained in the modern literature. There appear to be only three major lines of thought: (1) that affective feelings are created by "somatic markers," reflecting bodily changes that accompany emotions (Damasio 1994), which is a modern variant of the traditional James-Lange perspective that emotions arise ultimately from bodily commotions as well as the central representations of those processes; (2) that feelings are constituted of information arising from subcortical emotional systems interacting with higher working memory (LeDoux 1996 1999) or linguistic abilities of the forebrain (Rolls 1999), and (3) that affective states emerge from the intrinsic neurodynamics of basic emotional circuits interacting with primordial SELF[4] structures (neurosymbolic "virtual bodies") of the brain (Panksepp 1998a, b). This last view also aspires to provide a concrete infrastructure for theories of affect such as Freud's (see Solms and Nersessian 1999, and Panksepp 1999b, with the accompanying commentaries). More recently, Damasio (1999b) has also highlighted the importance of these subcortical systems more than he did in his previous synthesis, a view that is congenial to my way of thinking.

Despite all the controversy, there appears to be growing agreement that various subcortical systems are absolutely essential for orchestrating emotional action readiness as well as the creation of emotional feelings, whether it be directly through the capacities of the lower neural substrates (Panksepp 1998) or only through their interactions with higher cortico-cognitive functions (LeDoux 1996; Rolls 1999). These differences may eventually turn out to be more matters of emphasis than substance, so that a synthetic, consensual view will emerge as we all take the full complexity of the brain into account. As with other polarized controversies concerning the nature of mind, such as biological (nature) vs. social constructivist (nurture) approaches, the truth probably lies somewhere in between, and a fully integrated viewpoint will eventually be constructed through mutually respectful critiques and discussions (LeDoux 1999; Nielsen and Kazniak 1999; Panksepp 1999b, Watt 1998). Usually the investigators that do not believe in the causal reality of feelings focus on the very rapid unconditioned eruptive emotional response patterns, while those who believe internal affective experiences modify behavior tend to focus on the longer-term behavioral changes emotions produce.

Of course, only the eventual empirical comparison of predictions from the various viewpoints will be able to further resolve key issues, such as the necessity and contributions of lower vs. higher systems in the creation of affective experience. At least, we finally have credible conceptual schemes to dissect affective influences experimentally, and to evaluate how they fit with more traditional cognitive conceptions of consciousness. An attractive (i.e., testable) aspect of the affective neuroscience perspective is that the essential

substrates for emotional feelings arise from specific subcortical circuits that can be directly studied (MacLean 1990; Panksepp 1998a). This view also readily permits emotional values to interact intimately with the Extended Reticular-Thalamic Activating System (ERTAS), which helps construct conscious awareness of external events (Baars 1996; Newman 1997; Watt 1998).

One of the main areas of the brain where these interactions may occur is the periaqueductal gray matter (PAG) of the midbrain, and a case has been made that this is the epicenter for the generation of affect (Panksepp 1998a,b). This ancient brain area, along with surrounding tectal and tegmental zones may constitute a primal form of affective self-representation within the brain. I will not repeat the arguments and evidence, but merely indicate that this hypothesis is in agreement with some fairly stringent psychobiological criteria: (1) Namely, where in the brain is the most massive concentration of emotional information? Within the PAG. (2) Where in the brain can one evoke the largest number of coherent emotional responses with the lowest intensities of brain stimulation? The PAG again. (3) Where in the brain does the smallest amount of neural damage have the most severe effects on emotions? In the PAG. I could go on. It is important to re-emphasize that the PAG has easy access to the ascending reticular activating system to arouse and regulate the activity of the cortico-cognitive parts of the brain, and it has direct and massive reciprocal influences with all other brain areas that have been implicated in emotions from the hypothalamus to basal forebrain, the extended amygdaloid nuclei and the anterior cingulate and frontal cortices (Holstege, Bandler, and Saper 1996). In other words, the "primes" at the base of emotional systems (Buck 1985; Panksepp 1982), epigenetically establish broad fields of competence in many brain areas that are hierarchically arranged and richly connected with each other. In this view, it is important to emphasize that at a very young age, children show intense arousal of the lower substrates, but as they mature, the higher systems become more metabolically active (Chugani 1998). I would assume that the lower mappings are reiterated in higher mappings as organisms mature.

The intimate interaction of affective and attentional systems at primitive levels of the neuroaxis provide ways in which externally-directed cognitive searchlights can be optimally directed within the perceptual fields of higher brain areas. It should not escape our attention that such interactions provide yet other ways for emotional feelings to be enhanced and blended with cognitive processes. The interaction of ancient emotional command systems of the brain with ERTAS architectures can help broadcast emotional influences widely throughout a neurodynamically unified brain (for a thorough airing of such issues, see Watt 1998). In some yet unfathomed way, consciousness probably depends on very broad neuroelectric oscillatory coherences among large ensembles of neurons — with qualitatively different coherences emerging with different emotional states.

Although an enormous number of specifics remain to be detailed, it does seem that emotional systems interact directly with various workspaces of consciousness, thereby generating internally experienced affective states that permit organisms to operate in the world with a limited set of prototypical psychological and behavioral attitudes. Through the ontogenetic and epigenetic emergence of more complex informational structures (Oyama 1985), the basic emotions allow organisms to sustain types of intentional attitudes toward the dynamics of their social and non-social environments. There are bound to be 'affect-logic' types of cognitive structures that emerge around each and every emotional system of the brain (Wimmer and Ciompi 1996). In short, there is abundant room for all of the views discussed above to contribute to the integrated neurodynamic psychic states that we recognize as the various emotional feelings. However, my own conviction is that without a clear incorporation of basic emotional circuits as foundational structures in any scheme, we will be missing the essential heart of affective consciousness in the brain. I suspect the main reason this view is not more widely accepted, or even discussed, is simply because most investigators interested in understanding emotions and feelings do not work on the brain and are wary of any form of reductionism. Hence they are prone to disregard this arena of intellectual activity. However, now that more investigators are beginning to understand the importance of establishing superve-nience relationships between basic psycholoigcal and neuroscientific priniciples, there are signs of change in the air (Buck 1999; Izard 1993; Solms 1997).

It presently seems likely that neuropeptides are essential for creating emotional and motivational specificity within the emotional command systems described above (Buck 1999; Panksepp 1993a): Just consider a few lines of evidence: the ability of CRF to promote anxiety and separation distress, of αMSH to promote fearful hiding, of opioids to reduce many negative emotions, of oxytocin and prolactin to reduce separation distress and to promote nurturance, of Leutinizing Hormone–Releasing Hormone (LH–RH) to promote sexuality, of angiotensin to instigate thirst and the many neuropeptides that seem to instigate hunger (for review of evidence, see Panksepp 1998a). The remarkable fact is that the peripheral physiological and central effects of these hormones are remark-ably congruous. Presumably these systems help create resonances in the extended trajectory of the emotion integrating systems, from PAG to frontal regions of the brain, that directly mediate affective states. I would imagine that certain dynam-ics allow these systems to captivate a great deal of the brain (Freeman 1995; Panksepp 1999c), providing distinct, affect-specific neuro-"gravitational" forces or attractors by which perceptions and cognitions can be brought in line with various emotional demands. The fact that many of these neuropeptides can act not only synaptically, modulating global brain control systems such as acetylcho-line, norepinephrine, and serotonin systems, but also in more diffuse paracrine ways

that also allow wide broadcasting of their influences in the brain (see Panksepp 1998a, Chap. 6).

In sum, a great deal has been learned about certain basic emotional systems of the brain. We now have a working understanding of FEAR systems (Panksepp 1990; Rosen and Schulkin 1998) and its connections to world events (LeDoux 1996). We have insight into the nature of RAGE systems because of the groundbreaking work of John Flynn (1967) and his students Richard Bandler (Bandler and Keay 1996) and Alan Siegel (Siegel, Roeling, Gregg, and Kruk 1999). We have a profound understanding of the SEEKING or "I want" system (Robinson and Berridge 1993; Ikemoto and Panksepp, 1999; Panksepp 1981; 1986), especially through the exquisite electrophysiological work of Wolfram Schulz (1998). There are, of course, brain systems for sexual LUST and maternal CARE. We have a provisional understanding of separation-distress or sorrow/PANIC and joy/PLAY systems (Panksepp 1998a, 2000b).

From my point of view, the neurodynamic effects of these systems on some type of primordial neurosymbolic representation of the SELF, probably established in motor coordinates, constitute the essence of emotional feelings (Panksepp 1998a,b). These neurodynamics may have functional isomorphisms with the psychological dynamics of each of the emotions, helping us envision how to explain the gap between neural states and psychological states within the brain. Such isomorphisms are not as evident in the generation of cognitive qualia, even though there is still hope that, if we look at brain states in the right way, they may become evident. The fact that one can initiate many different emotional feelings by repeating the idealized motor dynamic of emotional movements (Clynes 1988), affirms how intimately emotional actions and internal feeling states are coupled.

Of course, we have a long way to go before anything like a comprehensive understanding of these systems, especially sorrow and joy, is achieved. A focus on these basic social emotions may be especially apt in any discussion of mammalian consciousness, since it was a major transition in brain evolution when creatures who truly cared for each other evolved from those that did not. The resulting social attachments provided an opportunity for widening the psychology of social space, which probably had a major impact on the emergence of special subtleties in affective consciousness and in the way organisms experienced their inner selves (Insel 1997; Nelson and Panksepp 1998; Panksepp 1981).

5. The brain substrates of sorrow/PANIC

Our work on the nature of separation-distress or sorrow/PANIC was initiated in 1972 when brain opioid receptor systems were first discovered. We began by

entertaining the idea that this system might mediate social bonding. Until then, social bonding was thought to emerge from young organisms' receipt of other primary reinforcers, such as food, water and warmth from caretakers. The possibility that there was a distinct brain system for social affect was not in the mainstream of scientific thought. Our hypothesis arose from a recognition that both narcotic addiction and social dependence shared three remarkable similarities — i) a powerful attachment or active pleasure phase, ii) a sustained habituation or tolerance phase where the explicitly felt pleasure emanating apparently from the object of desire diminished, and iii) a traumatic withdrawal phase if that source of now mild satisfaction was taken away. The theory was fundamentally neuropsychological — the nature of brain substrates for animal emotions and human affective experiences were deemed of equal importance.

After some false starts, many empirical tests confirmed the theory (Panksepp *et al.* 1980): Opioids, both of the exogenous and endogenous varieties, turned out to be remarkably potent attenuators of social separation-induced vocalizations (crying) in the young of several species, including dogs, guinea-pigs, and domestic chicks (Herman and Panksepp 1978; Panksepp *et al.* 1978a, b). This was followed by the discovery that these opioid effects were due to quite specific influences on subcortical brain circuits (Herman and Panksepp 1981, Panksepp *et al.* 1988). Of an enormous number of psychoactive agents evaluated, only a few others, such as oxytocin, prolactin and glutamate receptor antagonists produced comparable calming effects, and only a few, such as Corticotrophin Releasing Factor (CRF), curare and glutamate receptor agonists, could specifically promote such vocalizations (Panksepp 1991, 1998a). Many other social processes were modified in ways that were highly compatible with the opioid theory of social affect (Panksepp *et al.* 1980), and the findings were extended into the human clinical realm, with the opioid-excess theory of autism (Panksepp 1979).

Although there was much resistance to emotion research in the behavioral neuroscience community when we started, especially to issues related to affective processes in other animals, within a decade others were confirming and extending the above findings in other species, especially laboratory rodents (Kehoe and Blass 1986) and primates (Kalin, Shelton and Barksdale 1988; Keverne, Martensz and Tuite 1989), including several subtle primate social behaviors (Kalin, Shelton and Lynn 1995; Keverne, Nevison and Martel 1999). At present it is generally accepted that social-emotional regulation is strongly controlled by opioid and oxytocinergic dynamics within mammalian and avian brains (Carter 1998; Nelson and Panksepp 1998). Our ability to turn on separation calls in adult animals by stimulating specific areas of the brain established the approximate locations of these emotional systems — which are strongly concentrated in subcortical structures such as ventral septal and dorsal preoptic

areas, the bed nucleus of the stria terminalis, various dorsomedial thalamic zones, and the dorsal regions of the PAG (Herman and Panksepp 1981; Jürgens 1998; Panksepp *et al.* 1988). When this emotional system, which surely has various subcomponents, is quiescent, presumably because of the neurochemical influences of supportive pro-social frameworks, individuals tend to be emotionally confident for they have a 'secure base.' With the perception of social loss, the dynamics shift toward panicky feelings, yielding fertile ground for the emergence of depression, despair and other variants of psychic pain (Panksepp *et al.* 1991).

The consequences of this work for understanding affective consciousness are straightforward. In humans, we can say that opioids, along with oxytocin and some other brain chemistries, are brain features that produce the social comfort that one feels in the presence of friendly companionship. However, with time, the conscious perception of good feelings tends to fade, rapidly becoming a subconscious penumbra for daily activities. Only when the secure base is compromised by the loss of a loved one does one promptly recognize (through rapidly aroused emotional feelings) how dependent they were on the social support that is no longer consciously felt. Thus it is not surprising that low opioid activity in the brain, especially after opioids have been high for some time, generates feelings of isolation and distress, providing motivation for animals and humans to seek social contacts and companionship. Now social contact again produces an active sense of satisfaction and even joy. Of course the level of joy that one feels is highly dependent on the quality of social engagement that one achieves. If the interactions are highly playful, consisting of positive physical and psychological give-and-take, one experiences remarkably strong positive feelings that are not just the absence of loneliness. They reflect active engagement of playful-joy systems of the brain — systems just beginning to be probed by behavioral neuroscientists.

6. The brain substrates of playful joy and laughter

A few years after initiating our research program on the neurobiology of separation-distress, we started to experimentally analyze social rough-and-tumble (R and T) play in juvenile rats. At the time there was considerable resistance in my field to conceptualizing behavioral actions in such terms. Our first papers on the topic were soundly rejected because we used the term "play" to characterize the behavior we were observing. Gradually, as the empirical results were replicated by colleagues, the work found outlets (Panksepp and Beatty 1980), and became a modest growth area in behavioral neuroscience (Panksepp *et al.* 1984; Panksepp 1993b; Pellis and Pellis 1998; Thor and Holloway 1984; Vanderschuren *et al.* 1997). Parenthetically, it is remarkable that *human* R and T

play has yet to be adequately characterized in a standardized observational setting, even though there are some excellent studies of playground activities (Blurton Jones 1972; Humphreys and Smith 1987; Pellegrini 1995; Smith and Lewis 1985). In any event, it is becoming generally accepted that there are fundamental brain systems for R and T play that generate a positive affective state (Calcagnetti and Schechter 1992; Normansell and Panksepp 1990). Although there is bound to be controversy whether this type of CNS arousal should be called "joy" I would wager that the concept is on the right track.

When kids inform us that the thing they enjoy most in life is "to play," I believe we should seriously consider that the arousal and expression of play impulses is one of the primal sources of joy within the brain. Of course, one problem is that we also tend to use this term for other satisfactions — ones that arise from various consummatory rewards, and other achievements, as well as other brain systems such as the lateral hypothalamic self-stimulation or SEEKING system. This can yield much confusion for the coherent scientific utilization of emotional words. This is one reason I have chosen to designate emotional primes that exist in the brain with capitalized labels (Panksepp 1998a). Although we can be sure that all our semantic usages are socially constructed, we can also be confident that social joy is tightly linked to PLAY systems in the brain. When this system is intensely aroused, we tend to exhibit one simple fixed-action pattern, indicating perhaps a specific JOY process — laughter.

For a long time, it was thought that this behavior was unique to humans and perhaps some other great apes (Provine 1997), but we have recently found evidence that has brought that anthropocentric supposition into question. Laboratory rats also exhibit a powerful vocal chirping-type response during play and in response to tickling (Knutson *et al.* 1998; Panksepp and Burgdorf 1999). Their laughter-like response occurs in a very high frequency range (~50-kHz) that humans cannot perceive without technological assistance (i.e., from so-called 'bat detector'). Our attempts to share this discovery have received as much resistance as our earlier work on play (for a discussion, see Panksepp and Burgdorf 1999), but it is a robust and reliable phenomenon that anyone with the necessary technology and initiative can observe. Of course, the interpretation of the phenomenon, as always, is open to alternative views.

The reason we believe it is an evolutionary homolog of primitive human laughter is because if fulfills the following empirical criteria: (1) The repetitive sound characteristics are reminiscent of primate laughter; (2) the vocalization is brought out powerfully by positive social interchange (e.g., play) and even more so by tickling; (3) individuals that chirp the most in response to tickling, also play the most among each other; (4) young animals chirp more than older animals; (5) animals approach hands that have tickled them much more than those that have petted them, and they are attracted to other stimuli that have been

associated with tickling; (6) the response classically conditions rapidly to cues that predict tickling; (7) animals will learn instrumental responses to receive tickling; (8) young animals like to spend more time with older animals that chirp a lot rather than with those that chirp infrequently; and (9) just like humans who are more ticklish in certain areas of the body (e.g., ribs), young rats have 'tickle skin' concentrated at the nape of the neck where they typically direct their own play activities (Panksepp 1998a; Panksepp and Burgdorf 1999). Finally, (10) all fearful and negative affective stimuli we have tested (cat-smell, foot shock, new places, being held by the scruff of the neck) reduce this response. The results could not be more clear. Hence, in line with the evidence, we must provisionally conclude that the animals are in fact exhibiting a joy-laughter response. Indeed, this response is a temperamental characteristic of animals, for it can be successfully selected for and against within four generations of selective breeding (Panksepp, Burgdorf and Gordon, 2000).

If we could identify the brain areas that generate this response, as well as the urge to play, we may have the beginning of a credible neuroanatomy for social joy in mammals, including, we believe, humans. We may also discover some new ways to treat childhood problems such as attention deficit hyperactivity "disorders" (Panksepp 1998a, c). If we identify the fundamental neurochemistries for this type of joy, might it not also yield new ideas for treating depression? There is much important empirical work left to do in this field, and so far, few are doing it. I look forward to the next century of behavioral brain research, when more investigators will begin to work on the fundamental nature of the many emotions of the mammalian brain from which so much of our human nature appears to arise.

7. Consciousness and emotion studies in the 21st century

In closing, let me share a few personal perspectives and hopes for the future. In the coming century we will need to better appreciate the likelihood that consciousness and emotions arise from a diversity of brain substrates, and that the study of these substrates will provide the most productive avenue to understand the sources of consciousness and the affective platform upon which other forms of consciousness may have been built. John Searle (1998), in the tradition of Aristotle, has been at the forefront of calls for naturalistic approaches to the study of consciousness and the self. I agree with his vision, but would add that it is long past time for philosophy to become an increasingly experimental field of inquiry. There are many critical empirical questions concerning the subjective nature of mind that other disciplines are not yet courageous enough to tackle. Because of the neuroscience revolution, I believe philosophers presently have an

opportunity to make major empirical contributions to our understanding of these matters outside their traditional realms of logical discourse (e.g., see Solms 1997 and accompanying commentaries).

We can only get so far with rational argumentation, and we might be wise to consider that the "dream of reason," namely logic not guided by empirical observations, can create intellectual monstrosities as readily as it can inform. The evolved brain-mind is not simply a propositional logic machine. The mammalian brain is an organ that has been constructed in the long course of evolution to survive the challenges of varying environments. Some of these challenges were so consistently encountered by our ancestors that inbuilt mechanisms for certain sensory-motor responses and internal psycho-motor tendencies became the instinctual birthright of 'the mammalian brain,' yielding the affective equalia that we still share with all our furry relatives. Although we humans now have a stupendous neocortical thinking cap, which resembles a random-access, *tabula rasa* of remarkable proportions, the capacities of those higher brain regions are completely supported from below by instinctual emotional processes. The lower brain capacities automatically inform us of our most pressing concerns, and still help energize our behaviors in emotionally life-challenging situations. Of course, the guidance of behavior has to be done in conjunction with here and now cognitive abilities. These emotionally driven higher brain activities do not necessarily generate profound thoughts, but they do inspire important ones, surrounded by feelings that keeps us on track, informing us of various behavioral alternatives and, sometimes, of our deeper nature.

I do believe that primary-process consciousness, a primitive form of affective (raw-feel) consciousness, first emerged on the face of the earth at about the time that our basic emotional systems were being constructed in the caldron of evolution. We now need to devote more empirically guided conceptual effort to understanding exactly what was constructed, and what consequences it had for the evolution of the rest of the brain as well as the epigenetic development of each individual. Many investigators have recently been trying to crack the code of fear learning within the amygdala, hippocampus and bed-nuclei of the stria terminalis. Similarly, we need cadres of dedicated individuals who are willing to pursue equally fine work on the learning and neuronal changes that transpire in the many other emotional systems of the mammalian brain. Even more, we need investigators who are willing to take the affective nature of mammalian emotions more seriously than ever before, to create a lasting understanding of the many details of the unconditioned emotional systems that provide an enormous integrity to the fully embodied character of animal behavior and to the uninhibited joys and sorrows of human beings as well. We also must develop better approaches to dealing with internally triggered subjective experiences as they truly present themselves within the human brain-mind (e.g., Pennebaker, Mayne,

and Francis 1997). In doing that, we should not forget how important the body is in all this, especially the various representations of the body within the brain.

Indeed, for deep progress, we may need to develop new paradigms of psychoanalytic research, ones that views human nature more openly as being built on pre-linguistic, pre-cognitive foundations (Rapaport 1953). Psychoanalytically oriented research will be needed that tries to analyze emotions and other subjectively experienced mental changes in terms of the first-person processes that they are (Solms 1997). Our 'subjects' need to be collaborators in our intellectual journeys if we truly wish to probe the nature of affect — co-conspirators, so to speak, in the search for honest descriptions of their affective lives. We need to break through the semantic, social-desirability, and confabulatory skills of the left hemisphere (Schiffer 1998), to get to what the right hemisphere is feeling. The first- and third-person perspectives need to be integrated into a coherent whole in the study of emotions (Nielsen and Kazniak 1999), more so than in any of the other human sciences. When that is done well, we shall truly have some new approaches for the understanding of human emotions and consciousness.

This type of knowledge should have enormous implications for our philosophical conceptions of the many primitive psychic forces, on the edges of our refined cognitive consciousness, which undergird human nature. Recognition of such issues will improve our ability to conceptualize what has gone wrong when this nature becomes deviant in psychiatric disorders and depraved in the moral development of some individuals. This work, if done well, will open up new vistas in biological psychiatry and solidify our conceptions of ourselves as creatures of the world. Indeed our newfound neurochemical knowledge now permits us to determine how affective dynamics are related to specific brain systems, and how far back these systems go in evolution. For instance, the vasopressin/oxytocin family of neuropeptides which have figured so heavily in our recent underdstanding of vertebrate social behaviors (Carter 1998; Insel 1997; Nelson and Panksepp 1998; Panksepp 1993a), go back to primitive invertebrates. Earthworms and molluscs still regulate their socio-sexual affairs with homologous molecules (Kesteren *et al.* 1995; Satake *et al.* 1999).

Obviously, to maintain an image of human beings as creatures to be cherished, we have to create an intellectual framework that enobles us along with the other animals. Although the material world may have no intrinsic values — nothing that is intrinsically good or bad — our evolved brains do have such values. We must recognize these values and respect them for the powerful evolutionary solutions that they are. They are not just fantasies, but the fundamental core of our being that is ultimately reflected in the social structures that we do build with our imaginations. Our educational systems need to recognize these inborn values as the fundamental sources of our morals, and be willing to

educate our children as to their deeper nature. As Nietzsche (1885/1996, p. 100) asserted, and the life work of Freud affirmed, the "moralities are also merely a sign language of the affects." And now, for the first time in our intellectual history, we can finally come to terms with the psychobiological nature of those 'forces' in the body and the brain.

Notes

1. The term "equalia" is used to denote those evolutionary qualia, such as basic affective feelings, that were constructed as neurosymbolic birthrights of our brain through evolutionary selection operating on our ancestors as opposed to the "qualia" that come to us more directly by experiencing the world with our external senses. (See Panksepp 1999b.)
2. To estimate the magnitude of the neglect of affective processes in philosophy, I turned to The *Nature of Consciousness* (Block, Flanagan and Guzeldere 1997): Of the 50 contributions, only one — Michael Tye's "A Representational Theory of Pains and Their Phenomenal Character" — aspires toward any substantive coverage of any one of the many powerful affective experiences of which we, and many other animals are capable. The one major exception was Frank (1988), and more recently there has been some noteworthy activity in the area (Griffiths 1997). As in psychology, we presently appear to be at the beginning of an "Affective Revolution" in philosophy.
3. There are obviously various types of affects. Here, the emotional affects are deemed to be those related to what I have called Grade A, Blue-Ribbon Emotional Systems (Panksepp 1982, 1998a), built around extensive command-circuits in subcortical brain regions. Then other affects, like the pleasure of taste and touch, are closely related to primary sensory inputs, and homeostatic affects such as hunger and thirst, related to internal homeostatic receptors. Only affective processes of the basic emotional systems may be closely related to the brain's motor programs, although we should not forget that the pleasure of taste is enhanced by chewing and swallowing, mirth through laughing, and sexual pleasure through copulatory movements.
4. The capitalization here, as well as for emotional systems, is used as a convention to designate that one is referring to specific neural systems that are an essential infrastructure for the concept (Panksepp 1998a).

References

Baars, B. J. 1998. *In the Theatre of Consciousness: The Workspace of the Mind.* Cambridge, MA: MIT Press.

Bandler, R. and Keay, K. A. 1996. "Columnar Organization in the Midbrain Periaqueductal Gray and the Integration of Emotional Expression". In Holstege, R. Bandler and C. B. Saper (eds), *The Emotional Motor System, Progress in Brain Research,* Vol. 107: Amsterdam: Elsevier, 287–300.

Beck, C. H. M. and Fibiger, H. 1995. "Conditioned Fear-induced Changes in Behavior and the Expression of the Immediate Early Gene *C-fos*: With and Without Diazepam Treatment". *Journal of Neuroscience* 15: 709–720.

Block, N. 1995. "On a Confusion about a Function of Consciousness". *Behavioral and Brain Sciences* 18: 227–287.

Block, N., Flanagan, O. and Guzeldere, G. (eds). 1997. *The Nature of Consciousness: Philosophical Debates.* Cambridge, MA: MIT Press.

Blurton Jones, N. 1972. "Categories of Child-child Interaction". In *Ethological Studies of Child Behavior*, N. Blurton Jones (ed), New York: Cambridge University Press.

Buck, R. 1985. "Prime Theory: An Integrated View of Motivation and Emotions". *Psychological Review,* 92: 389–413.

Buck, R. 1994. "Social and Emotional Functions in Facial Expression and Communication: the Readout Hypothesis". *Biological Psychology* 38: 95–115.

Buck, R. 1999. "The Biological Affects: A Typology". *Psychological Review* 106: 301–336.

Campeau, S., Falls, W. A., Cullinan, W. E., Helmreich, D. L., Davis, M., and Watson, S. J. 1997. "Elicitation and Reduction of Fear: Behavioural and Neuroendocrine Indices and Brain Induction of the Immediate-early Gene C-fos". *Neuroscience* 78: 1087–1104.

Calcagnetti, D. J. and Schechter, M. D. 1992. "Place Conditioning Reveals the Rewarding Aspect of Social Interaction in Juvenile Rats". *Physiology and Behavior* 51: 667–672.

Carter, C. S. 1998. "Neuroendocrine Perspectives on Soical Attachment and Love". *Psychoneuroendocrinology* 23: 779–818.

Carter, S., Kirkpatrick, B. and Lederhendler, I. I. (eds). 1997. *Neurobiology of Affiliation,* Vol. 807, *Annals of the New York Academy of Sciences.* New York: New York Academy of Sciences.

Clynes, M. 1988. "Generalized Emotions: How It May Be Produced, and Sentic Cycle Therapy". In M. Clynes and J. Panksepp, *Emotions and Psychopathology.* New York: Plenum Press, 107–170.

Chugani, H. T. 1998. "A Critical Period of Brain Developoment: Studies of Cerebral Glucose Utilization with PET". *Preventive Medicine* 27: 184–188.

Crick, F. 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul.* New York: Simon and Schuster.

Damasio, A. R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain.* New York : G. P. Putnam.

Damasio, A. R. 1999a. "Commentary". *Neuro-Psychoanalyis* 1: 38–39.

Damasio, A. R. 1999b. *The Feeling of What Happens.* New York: Harcourt Brace.

Darwin, C. 1872/1998. *The Expression of the Emotions in Man and Animals, Third Edition,* New York: Oxford University Press.

Flynn, J. P. 1967. "The Neural Basis of Aggression in Cats". In *Neurophysiology and Emotion,* D. C. Glass (ed), New York: Rockefeller University Press, 40–60.

Frank, R. H. 1988. *Passions Within Reason: The Strategic Role of the Emotions.* New York: Norton.

Freeman, W. 1995. *Societies of Brains: A Study in the Neuroscience of Love and Hate.* Hillsdale, NJ: Lawrence Erlbaum.

Freud, S. 1823/1981. *The Ego and the Id.* In J. Strachey (ed), *The Standard Edition of the Complete Psychological works of Sigmund Freud,* London: The Hogarth Press.

Holstege, G., Bandler, R. and Saper, C.B (eds). 1996. *The Emotional Motor System: Progress in Brain Research,* Vol. 107. Amsterdam: Elsevier.

Humphreys, A. and Smith, P. 1987. "Rough-and-tumble, Friendship, and Dominance in Schoolchildren: Evidence for Continuity and Change with Age". *Child Development* 58: 201–212.

Kollack-Walker, S., Watson, S. J. and Akil H. 1997. "Social Stress in Hamsters: Defeat Activates Specific Neurocircuits within the Brain". *Journal of Neuroscience* 17: 8842- 8855.

Friedlund, A. 1991. "Sociality of Solitary Smiling: Potentiation by an Implicit Audience". *Journal of Personality and Social Psychology* 60: 229–240.

Friedlund, A. 1994. *Human Facial Expression: An Evolutionary View.* San Diego, CA: Academic Press.

Griffiths, P. E. 1997. *What Emotions Really Are.* Chicago: Univ. of Chicago Press.

Harris, J. C. and Newman, J. D. 1988. "Combined Opiate/adrenergic Receptor Blockade Enhances Squirrel Monkey Vocalization". *Pharmacology, Biochemistry and Behavior* 31: 223–226.

Herman, B. H. and Panksepp, J. 1980. "Effects of Morphine and Naloxone on Social Attachment in Infant Guinea Pigs". *Pharmacology Biochemistry and Behavior* 9: 213–220.

Herman, B. H. and Panksepp, J. 1981. "Ascending Endorphin Inhibition of Distress Vocalization". *Science* 211: 1060–1062.

Ikemoto, S. and Panksepp, J. 1999. "The Role of Nucleus Accumbens Dopamine in Behavior: A Unifying Interpretation with Special Reference to Reward Seeking". *Brain Research Reviews,* 31: 6-41.

Insel, T. R. 1997. "The Neurobiological Basis of Social Attachment". *American Journal of Psychiatry* 154: 726–735.

Izard, C. E. 1993. "Four Systems for Emotion Activation: Cognitive and Noncognitive Processes". *Psychological Review* 100: 68–90.

Jackson, J. H. 1931/1984. Evolution and Dissolution of the Nervous Systems". In J. Taylor (ed), *Seclected Writings of John Hughlings Jackson.* London: Hodder and Stoughton, 2: 45–75.

Jürgens, U. 1998. "Neuronal Control of Mammalian Vocalization, with Special Reference to Squirrel Monkey". *Naturwissenchaftern* 5: 376–388.

Kalin, N. H., Shelton, S. E. and Barksdale, C. M. 1988. "Opiate Modualtion of Separation-induced Distress in Non-human Primates". *Brain Research* 440: 285–292.

Kalin, N. H., Shelton, S. E., and Lynn, D. E. 1995. "Opiate Systems in Mother and Infant Primates Coordinate Intimate Contact During Reunion". *Psychoneuroendocrinology* 7: 735–742.

Kehoe, P. and Blass, E. M. 1986. "Opioid Mediation of Separation Distress in 10 Day Old Rats: Reversal of Stress with Maternal Stimuli". *Developmental Psychobiology* 19: 385–398.

Kesteren, R. E., Smit, A. B., De Lange, R. P. J., Kits, K. S., Van golen, F. A., R. C. Van Der Schors, N. De With, Burke, J. F. and Geraerts, W. P. M. 1995. "Structural and Functional Evolution of the Vasopressin/oxytocin Superfamily: Vasopressin-related Concopressin is the Only Member Present in *Lymnaea*, and Is Involved in Control of Sexual Behavior". *Journal of Neuroscience* 15: 5989–5998.

Keverne, E. B., Martensz, N. and Tuite, B. 1989. "ß-Endorphin Concentrations in CSF of Monkeys Are Influenced by Grooming Relationships". *Psychoneuroendocrinology* 14: 155- 161.

Keverne, E. B., Nevison, C. M. and Martel, F. L. 1999. "Early Learning and the Social Bond". In *The Integrative Neurobiology of Affiliation,* C. S. Carter, I. Lederhendler, B. Kirkpatrick, Cambridge, MA: MIT Press, 263–73.

Kolb, B. and Tees, C. (eds). 1990. *The Cerebral Cortex of the Rat.* Cambridge, Mass.: MIT Press.

Knutson, B., Panksepp, J. and Burgdorf, J. 1998. "Anticipation of Play Elicits High-Frequency Ultrasonic Vocalizations in Young Rats". *Journal of Comparative Psychology* 112: 65–73.

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

LeDoux, J. 1999. "Psychoanalytic Theory: Clues from the Brain". *Neuro-Psychoanalyis* 1: 44–49.

MacLean, P. D. 1990. *The Triune Brain in Evolution: Role in Paleocerebral Functions.* New York: Plenum Press.

Nelson, E. and Panksepp, J. 1998. "Brain Substrates of Infant-mother Attachment: Contributions of Opioids, Oxytocin, and Norepinephrine". *Neuroscience and Biobehavioral Reviews* 22: 437–452.

Newman, J. 1997. "Putting the Puzzle Together, Part I: Towards a General Theory of the Neural Correlates of Consciousness". *Journal of Consciousness Studies* 4: 47–66.

Nietzsche, F. 1885/1996. *Beyond Good and Evil.* New York: Vintage.

Nielsen L. and Kaszniak, A. 1999. The The University of Arizona, Consciousness Studies electronic seminar on "The Investigation of Conscious Emotion: Combining First Person and Third Person Methodologies" Feb. 22–March 5 1998 http://www.consciousness.arizona.edu/emotion/library.html

Normansell, L. and Panksepp, J. 1990. "Effects of Morphine and Naloxone on Play-Rewarded Spatial Discrimination in Juvenile Rats". *Developmental Psychobiology* 23: 75–83.

Oyama, S. 1985. *The Ontongeny of Information: Developmental Systems and Evolution.* Cambridge, UK: Cambridge University Press.

Panksepp, J. 1979. "A Neurochemical Theory of Autism". *Trends in Neuroscience* 2: 174–177.

Panksepp, J. 1981. "Brain Opioids: A Neurochemical Substrate for Narcotic and Social Dependence". In S. Cooper (ed) *Progress in Theory in Psychopharmacology.* London: Academic Press, 149–175.

Panksepp, J. 1981. "Hypothalamic Integration of Behavior: Rewards, Punishments, and Related Psychobiological Process". In P. J. Morgane and J. Panksepp (eds), *Hand-*

book of the Hypothalamus, Vol. 3, Part A. Behavioral Studies of the Hypothalamus. New York: Marcel Dekker, 289–487.

Panksepp, J. 1982. "Toward a General Psychobiological Theory of Emotions". *The Behavioral and Brain Sciences 5,* 407–467.

Panksepp, J. 1986. "The Anatomy of Emotions". In R. Plutchik (ed), *Emotion: Theory, Research and Experience Vol. III. Biological Foundations of Emotions.* Orlando: Academic Press, 91–124.

Panksepp, J. 1990. "The Psychoneurology of Fear: Evolutionary Perspectives and the Role of Animal Models in Understanding Human Anxiety". In *Handbook of Anxiety.* Amsterdam: Elsevier/North-Holland Biomedical Press, 3–58.

Panksepp, J. 1990. "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 K. Strongman (ed), *International Reviews of Emotion Research.* Chichester, UK: Wiley, 59–99.

Panksepp, J. 1993a. "Neurochemical Control of Moods and Emotions: Amino Acids to Neuropeptides". In *The Handbook of Emotions,* M. Lewis and J. Haviland (eds). New York: Guilford, 87–107.

Panksepp, J. 1993b. "Rough-and-tumble Play: A Fundamental Brain Process". In *Parents and Children Playing,* K. B. MacDonald (ed). Albany, NY: SUNY Press, 147–184.

Panksepp, J. 1998a. *Affective Neuroscience: The Foundations of Human and Animal Emotions,* New York: Oxford University Press.

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. 1998c. "Attention Deficit Disorders, Psychostimulants, and Intolerance of Childhood Playfulness: A Tragedy in the Making?" *Current Directions in Psychological Sciences* 7: 91–98.

Panksepp, J. 1999a. "The Riddle of Laughter: Neural and Psychoevolutionary Underpinnings of Joy". *Current Directions in Psychological Sciences* 9: In Press.

Panksepp, J. 1999b. "Emotions as Viewed by Psychoanalysis and Neuroscience: An Exercise in Consilience, and Accompanying Commetaries". *NeuroPsychoanalysis* 1: 15–89.

Panksepp, J. 2000a, In Press. "The Neurodynamics of Emotions: An Evolutionary-Neurodevelopmental View". In M. D. Lewis and I. Granic (eds), *Emotion, Self-Organization, and Development.* New York: Cambridge Univ. Press.

Panksepp, J. 2000b, In Press. "Emotions as Natural Kinds within the Mammalian Brain". In *The Handbook of Emotions , 2nd ed.* M. Lewis and J. Haviland (eds), New York: Guilford.

Panksepp, J., Herman, B., Conner, R., Bishop, P. and Scott, J. P. 1978. "The Biology of Social Attachments: Opiates Alleviate Separation Distress". *Biological Psychiatry* 9: 213–220.

Panksepp, J. and Beatty, W. W. 1980. "Social Deprivation and Play in Rats". *Behavior and Neural Biology* 30: 197–206.

Panksepp, J., Vilberg, T. , Bean, N.J. , Coy, D.H. and Kastin, A.J. 1978. "Reduction of Distress Vocalization in Chicks by Opiate-like Peptides". *Brain Research Bulletin* 3: 663- 667.

Panksepp, J., Herman, B.H., Villberg, T., Bishop, P. and DeEskinazi, F.G. 1980. "Endogenous Opioids and Social Behavior". *Neuroscience and Biobehavioral Reviews* 4: 473–487.

Panksepp, J. , Siviy, S. , and Normansell, L. 1984. "The Psychobiology of Play: Theoretical and Methodological Perspectives". *Neuroscience and Biobehavioral Reviews 8:* 465–492.

Panksepp, J., Normansell, L.A., Herman, B. Bishop, P. and Crepeau, L. 1988. "Neural and Neurochemical Control of the Separation Distress Call". In J.D. Newman (ed), *The Physiological Control of Mammalian Vocalization.* New York: Plenum Press.

Panksepp, J., Normansell, L., Cox, J.F. and Siviy, S.M. 1994. "Effects of Neonatal Decortication on the Social Play of Juvenile Rats". *Physiology and Behavior* 56: 429–443.

Panksepp, J., Yates, G., Ikemoto, and Nelson, E. 1991. "Simple Ethological Models of Depression: Social-isolation Induced "Despair" in Chicks and Mice". In *Animal Models in Psychopharmacology,* B. Olivier and J. Moss (eds), Holland: Duphar, 61–181.

Panksepp, J. and Burgdorf, J. 1999. "Laughing Rats? Playful Tickling Arouses High Frequency Ultrasonic Chirping in Young Rodents". In S. Hameroff, D. Chalmers and A. Kazniak, *Toward a Science of Consciousness III,* Cambridge, Mass.: MIT Press, 231-242.

Panksepp, J., Burgdorf, J., and Gordon, N. 2000. "Toward a Genetics of Joy: Breeding Rats for 'Laughter'". In A. Kazniak (ed), Proceedings of the 1998 Ischia Conference on Emotions and Consciousness, University of Arizona Consciousness Studies Program.

Pellegrini, A.D. 1995. "A Longitudinal Study of Boys' Rough-and-tumble Play and Dominance During Early Adolescence". *Journal of Applied Developmental Psychology* 16, 77–93.

Pellis S.M. and Pellis, V.C. 1998. "Play Fighting of Rats in Comparative Perspective: A Schema for Neurobehavioral Analyses". *Neuroscience and Biobehavioral Reviews* 23: 87–101.

Pennebaker, J.W., Mayne, T.J. and Francis, M.E. 1997. "Linguistic Predictors of Adaptive Bereavement". *Journal of Personality and Social Psychology* 72: 863–871.

Picard, R.W. 1997. *Affective Computing,* Cambridge, Mass. : MIT Press.

Provine, R.R. 1997. "Contagious Yawning and Laughter: Significance for Sensory Feature Detection, Motor Pattern Generation, Imitation, and the Evolution of Social Behavior". In *Social Learning in Animals: The Roots of Culture,* C.M. Heyes and B.G. Galef, (eds). New York: Academic Press.

Rapaport, D. 1953. "On the Psycho-analytic Theory of Affects". *The International Journal of Psycho-analysis.* 34: 177–198.

Robinson, T. and Berridge, K. 1993. "The Neural Basis of Drug Craving: An Incentive-Sensitization Theory of Addiction". *Brain Research Reviews* 18: 247–291.

Rolls, E.T. 1999. *The Brain and Emotion.* Oxford, UK: Oxford University Press.

Rosen, J. B. and Schulkin, J. 1998. "From Normal Fear to Pathological Anxiety". *Psychological Review* 105: 325–350.

Satake, H., Takuwa, K., Minakata, H. and Matsushima, O. 1999. "Evidence for Conservation of the Vasopressin/oxytocin Superfamily in Annelida". *Journal of Biological Chemistry* 274: 5605–5611.

Scherer, K. R., Wallbott, H. G. and Summerfield, A. B. (eds). 1986. *Eperiencing Emotion: A Cross-cultural Study.* Cambridge, UK: Cambridge University Press.

Schultz, W. 1998. "Predictive Reward Signal of Dopamine Neurons". *Journal of Neurophysiology,* 80: 1–27.

Searle, J. R. 1998. "How to Study Consciousness Scientifically". *Brain Research Reviews* 26: 379–387.

Schiffer, F. 1998. *Of Two Minds: The Revolutionary Science of Dual-brain Psychology.* New York: Free Press.

Siegel, A., Roeling, T. A., Gregg, T. R. and Kruk, M. R. 1999. "Neuropharmacology of Brain-stimulation-evoked Aggression". *Neuroscience & Biobehavioral Reviews* 23: 359–89.

Smith, P. K. and Lewis, K. 1985. "Rough-and-tumble Play, Fighting, and Chasing in Nursery School Children". *Ethology and Sociobiology* 6: 175–181.

Springer, S. P. and Deutsch, G. 1998. *Left Brain, Right Brain: Perspective from Cognitive Neuroscience.* (5th ed.) New York: W. H. Freeman & Co.

Solms, M. 1997. "What Is Consciousness?" *Journal of the American Psycholanalytic Association* 45: 681–703; discussion 704–778.

Solms, M. and Nersessian, E. 1999. "Freud's Theory of Affect: Questions for Neuroscience". *NeuroPsychoanalysis* 1: 5-14 .

Thor, D. H. and Holloway, W. R.1984. "Social Play in Juvenile Rats: A Decade of Methodo- logical and Experimental Research". *Neuroscience & Biobehavioral Reviews* 8: 455–464.

Vanderschuren, L. J., Niesink, R. J. and Van Ree, J. M. 1997. "The Neurobiology of Social Play Behavior in Rats". *Neuroscience & Biobehavioral Reviews* 21: 309–326.

Watt, D. F. 1998. The Association for the Scientific Study of Consciousness electronic seminar on "Emotion and Consciousness," Sept. 21-Oct 9, 1998, http://www.phil.vt.edu/assc/esem.html

Wimmer, M. and 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.

Chapter 3

Consciousness, Motivation, and Emotion

Biopsychological Reflections

Bill Faw
Brewton-Parker College

This paper examines some nervous system mechanisms involved in consciousness, motivation, and emotion. We begin with preliminary thoughts on the relationships between various meanings of consciousness: having phenomenal experiences, being aware of objects and situations, and wakefulness. Then we look at a key issue in 'emotional consciousness,' specifically whether it is part of basic consciousness mechanisms or merely one 'channel' through which we have conscious content. At the end of the chapter are diagrams which present many of the brain-stem, sub-cortical and cortical areas referred to in the text. The diagrams of 'your brain on numbers' can be especially helpful in differentiating cortical areas.

1. 'Having phenomenal experiences' and 'being aware-of'

We begin with the assumption that the most fundamental element of any conscious experience is the *having of a phenomenal experience* (Ned Block 1995). An alternative starting point is to posit *consciousness-of* as the core (Natsoulas 1978; Lycan 1996). But this confounds the having of phenomenal experiences with the question as to whether all phenomenal experiences *represent something* to the experiencer (Bernard Baars, personal communication).

The task of *perceptual* systems is, precisely, to 'represent' objects and events in the world and within our body to us. The task of *cognitive* systems is to represent our present experiences and endeavors in more abstract form. In mammals, such propositional-representational information processing utilizes the cerebral neocortex (Panksepp, in Watt 1998). (References in Watt 1998 refer to various contributions to the 1998 Association for the Scientific Study of Consciousness

e-seminar on Emotion and Consciousness, focused upon Douglas Watt's contributions.)

In distinction to perceptual and cognitive systems, which have the basic goal of 'representing' things to us, basic bodily-feelings and *emotional* states might be 'pre-propositional' in origin (Panksepp, in Watt 1998), existing in animals without propositional-representational cortex and being generated from sub-cortical sources even in humans. If this is the case, then emotional consciousness should be especially dissociable into phenomenal experience and what the emotion might *represent* to the experiencer (Newman, in Watt 1998).

1.1 *'Having phenomenal experiences' and 'being awake'*

We have phenomenal experiences in many *states* of consciousness such as in alert to drowsy wakefulness; shallow to deep slow-wave-sleep; rapid eye movement (REM) sleep; various degrees of coma; and hypnotic, trance, delirious, and drug-induced states. States of consciousness differ from each other in both their quantitative degree of 'alertness' and in their qualitative characteristics of phenomenal experience. Thus, consciousness as the 'having of phenomenal experiences' is a much broader term than consciousness as 'being awake' — the latter representing only one of many mental states in which the former is manifest. In turn, within wakefulness there is a range of *sub-states* wherein we are more or less mentally alert, more oriented to internal mind or to external world, and more focused or diffused in attention. It is extremely important to find the experiential qualities and brain mechanisms that all conscious states have in common and those that distinguish different states of consciousness.

1.2 *The roles of motivation and emotion in core consciousness*

'Motivation' is a much wider concept than 'emotion.' Motivation is inferred whenever an animal acts in a goal-directed way (Baars, in Watt 1998). Motivation is a process that involves some coding for 'salience' but may have little or no emotion — and little or no consciousness — to it. Much of motivated behavior comes from completely non-conscious homeostatic (= status quo) control, which is enacted through motor neurons, constriction and relaxation of blood vessel walls, and increasing or decreasing of the sugar, insulin, adrenalin, testosterone, or estrogen levels in the blood.

I make the moderately strong claim that at least normal waking states of consciousness and most 'altered states' *have motivational and emotional* as well as perceptual or cognitive components. After all, the 'schizo' (split) of the 'phrenia' (mind) in schizophrenia is defined clinically as an abnormal split between the normal interdependence of emotion, perception, and cognition in

consciousness. My claim is a tentative version of Watt's and Panksepp's (Watt 1998; Watt, in Vogeley 1999): that emotion is a central organizing global component of consciousness — part of the experiential qualities and brain mechanisms that all conscious states have in common. An alternative is the view, variously expressed by Joseph LeDoux, John Smythies, and John Taylor in the Watt e-seminar, that 'emotional consciousness' is just one 'channel' though which we derive conscious content, in the way visual cortex gives us visual content.

The analogy suggested is to see consciousness as the sights and sounds in a television set. Some circuits in a traditional television set, such as the high-voltage-cathode-ray-tube, are necessary in order to get *any* television show — any sights and sounds. Then there are *specific* circuits that bring in specific channels. In the first position losing 'emotional consciousness' would be to lose all consciousness — all sights and sound; while in the second position losing our emotional system would be like losing 'channel 5' or even channels 5 through 10, which may be a big loss, but still allows other channels to come in. We will revisit these various issues in our concluding section.

1.3 *Level one emotional consciousness mechanisms*

It is artificial and tentative but perhaps helpful to divide the complex emotional system into levels or zones of emotional processing. The following levels are my own development of Panksepp's hints (in Watt 1998). These levels are hierarchical in four concurrent ways. The lower the level, the lower in the nervous system, the phylogenetically more primitive, the more apt to be controlled by higher levels, and the more non-consciously activated. The body of this chapter will explicate these five levels.

At the most basic level are the *servo-mechanism systems* that produce bodily-condition changes and actions that are crucial parts of emotional phenomenal experiences and behavioral responses. This level includes pain reflexes, the autonomic motor system, and the hormonal system. The second level is a brainstem zone of interacting *arousal/affective circuits*, containing five brainstem arousal systems which might be the lowest level for the activation of consciousness and for organizing fight, flight, and approach emotional responses. The third level incorporates the *hypothalamic-thalamic-basal-forebrain-septal areas,* to which Level 2 systems project. The fourth level incorporates the *amygdaloid and hippocampal complexes* and related sub-cortical parts of the 'limbic system.' The fifth level incorporates *anterior motivational/autonomic control areas*: temporal pole, anterior insula, posterior-ventral-medial pre-frontal cortex and anterior cingulate gyrus.

1.4 *Spinal-cord/brain-stem pain reflexes of the somatic motor system*

We will encounter three pain systems in this chapter: pain-reflex, slow-pain-emotional, and fast-pain-perceptual systems. The most primitive is the pain-reflex system to bring about direct and immediate responses to minimize tissue damage. When your finger comes in contact with a flame, the pain sensory neuron activates a spinal cord inter-neuron which, in turn, activates a spinal cord motor neuron, which pulls your finger away from the flame. All of this happens quickly and independently of pathways that carry pain information to the cortex to make you conscious of the pain. The spinal reflexes often exist even in spinal-cord-damaged persons whose cortex is never notified of the pain. Those with intact nervous systems can defy this system, but only by imposing a very high level 5 conscious system to inhibit it.

Visual, auditory, olfactory, gustatory and head-touch senses have comparable non-conscious sensorimotor reflex mechanisms in the brainstem with the same purpose: to pull us instantly out of harm's way before there is further tissue damage. These are processed by the midbrain's colliculi. We pull away from a bear that we see, hear, touch, and/or smell, hopefully before we activate the bear's *taste* receptors, even before we are consciously aware of the bear's existence. We do this because the retina, cochlea, and other receptors send information to the colliculi for reflex responses, independently of their projections to the cortex via the thalamus for conscious responses. Thus, much of one's motor responses are controlled at a totally non-conscious level. If one becomes aware that one made such a response, that awareness is due to other systems that reach consciousness.

1.5 *Visceral activation of the autonomic motor system*

The autonomic (= self ruling) nervous system has two branches. Its 'sympathetic' (= with pathos) branch prepares blood vessels, organs, cortex, sensory receptors, and muscle tone for fight/anger or flight/fear motor responses. Its 'parasympathetic' branch prepares the same vessels and organs for resting, grooming, mating, and digesting. Reciprocal inhibition between the two systems maintains the body's homeostasis and then moves things toward fight/flight or rest response-states as needed when commanded from Levels 2, 3, 4, or 5.

The *autonomic* motor system enacts motor responses over which we have little conscious control, like heart rate, blood pressure, or routine breathing, while the *somatic* motor system activates motor responses over which we can have conscious control, like walking, throwing, and talking. We can control some aspects of our breathing through somatic musculature — to hold our breath or hyperventilate — but need to do that as a conscious-control constraint on the on-

going non-conscious autonomic control. We have less conscious input into heart-rate and blood pressure control; but can influence even those through using somatic muscles, imagination, or biofeedback to activate sympathetic or parasympathetic autonomic systems.

We obviously can 'become conscious' of our adrenaline/noradrenaline (= epinephrine/ norepinephrine) rush, quick breathing, and so on, through *interoceptive* (= perception of bodily states) feedback, just as we can become conscious of our somatic muscle movements through *proprioceptive* (= perception of bodily positions) feedback, as well as through our *exteroception* (= perception of the outside world), such as seeing our blushing face in a mirror, hearing our breathing, and the like.

Interoceptive feedback from our autonomic nervous system is what William James (1884) and Carl Lange (1885), independently, felt was at the core of the 'conscious experience' of emotion. Lange limited it to autonomic feedback, while James added somatic feedback from our facial expression and other emotional responses. James' son-in-law, Walter Cannon (1927), demonstrated several problems with their theories, including the observation that sympathetic nervous system increases in heart beat, etc., might be found in any strong emotion — so that one cannot determine *which* emotion one is experiencing just from autonomic feedback. Thus, autonomic nervous system arousal and the feedback we gain from it probably play a more major role in the *quantitative/intensity* dimension of experienced emotion — how strong the emotion is, from full sympathetic arousal to full parasympathetic peacefulness — than in the *qualitative* dimension — which emotion I am feeling.

1.6 *Electrolyte-hormone/endocrine systems*

Level 1 Electrolyte-hormonal systems are regulated by the level-3 hypothalamus directly or through the pituitary body which dangles from it. One of the many ways in which hormonal systems control *emotional* responses is through the 'Hypothalamic-Pituitary-Adrenal Cortex (HPA) Axis.' This is one of two classic 'stress' circuits that are sort of twins. In the HPA circuit, the hypothalamus causes the pituitary to release ACTH, which stimulates the adrenal *cortex* to release various stress-fighting steroids such as cortisol. Depressed and anxious patients tend to show hyper-responsiveness in the HPA axis (Pivac *et al.* 1997; McAllister-Williams *et al.* 1998; Sonino and Fava 1996).

The other circuit is also controlled by the hypothalamus, which activates the sympathetic nervous system to activate the adrenal *medulla* — an inner part of the adrenal gland, surrounded by the adrenal cortex — to release epinephrine and norepinephrine. This circuit is generally activated first, and is a fast-acting response. The adrenal-cortex circuit is slower to be activated but lasts longer.

2. Level two brainstem zone of interacting arousal/affective circuits

Level 1 servo-mechanisms enter consciousness only through extensive upper-level feedback and control. Level 2 areas, in mutual-interaction with their respective levels 3, 4, and 5 targets, seem to be *responsible for consciousness itself* and the beginnings of emotional experiences and behavioral responses. Thus, these brainstem-to-cortex pathways underlie phenomenal experience.

Instead of there being one linear continuum of conscious states (say from alert wakefulness through sleep to deep coma: a beta–alpha–theta–delta wave EEG continuum), there seem to be four or five distinct continua, each of which involves a distinct brain-stem-to-cortex major neurotransmitter system (Faw 1997). In 1949 Moruzzi and Magoun discovered and named the reticular formation (RF), a loose net-like array of interconnected neurons, running in the core of the brainstem through the medulla, pons, and midbrain. While the nervous system's general excitatory (glutamate) and inhibitory (GABA) neuro-transmitters are amino acids, there are a number of specialized neurotransmitter systems that are much more circumscribed in location and function, that have their cell bodies in a few specific patches of the brainstem reticular formation. Newman (1997) has labeled the acetylcholine projection to the thalamus the *Extended Reticular Thalamic Activating System* (ERTAS). In the spirit of Schiff (1999), but quite independently, I would extend Newman's 'extended' system to include all five of these brainstem arousal systems and their level 3 connections.

Each ERTAS system seems to have its own unique role in motivational, emotional, perceptual, and cognitive systems, while their interplay seems to be what makes 'phenomenal experience' possible. Consciousness in the sense of 'being awake' arises from specific combinations of this interplay.

The RF has discrete and interacting nuclei that project the neurotransmitter *acetylcholine* from a string of nuclei, *norepinephrine* from a parallel string but especially the locus ceruleus of the Pons, *dopamine* from the ventral tegmental area (VTA) and substantia nigra of the midbrain, *serotonin* from the string of raphe nuclei, and *endorphins* from the periacquaductal gray (PAG) area. There are other places where each of these specialized neurotransmitters are found, but their roles in arousing the conscious mind seem to come from these locations in the RF. Each specialized neurotransmitter system projects from the brainstem to the cortex either through the hypothalamus or the thalamus (Moruzzi and Magoun 1949).

2.1 *Acetylcholine brainstem consciousness-oscillator system*

The leading nervous-system candidate to be the sine qua non for *having phenom-enal experiences* is the acetylcholine brainstem-reticular-formation-to-thalamus-to-

cortex arousal system (Crick 1984; Llinas *et al.* 1994; Coenen 1995; Newman 1997). These acetylcholine nuclei have long been implicated in the PGO spikes that activate and sustain REM dreams and REM muscle paralysis (Pinel 1997). Brainstem norepinephrine, dopamine, and serotonin centers all inhibit the acetylcholine centers that produce REM sleep, such that their activation can wake you out of REM sleep by an intense or meaningful sensory stimulation such as a fire alarm, a bright light, the smell of smoke, someone shaking you, or your baby's quiet sounds. REM begins with an inhibition of norepinephrine, dopamine, and serotonin centers, presumably by the neurotransmitter GABA. This disinhibits the former's inhibition of acetylcholine systems, which then produce the dreaming, muscle paralysis and so on of REM (Hobson 1997).

More recently, the brainstem acetylcholine nuclei have been found to constitute a *"brainstem oscillator"* (Coenen 1995; Steriade and McCarley 1990; Newman 1997), which activates the intralaminar nuclei (ILN) of the thalamus, which arouse cortex diffusely to better receive and process specific information coming through the thalamus. This may be the basic arousal system (Llinas *et al.* 1994; Newman 1997) — the high voltage cathode-ray tube — or merely part of the team of arousal systems (Faw 1997; Baars, in Watt 1998; Schiff 1999). This system is also involved in 'gating,' as we will see when we look at Level 3.

Acetylcholine pathways show their highest activation during both the normal waking state and REM states (Hobson 1997) and their lowest activation during slow wave sleep. The destruction or blocking of this pathway leads to permanent and very deep coma (Newman 1997). This pathway does not seem to be totally shut down during slow wave sleep, leading to speculation that it might generate even the perseverative mentation typically found during slow-wave sleep (Flanagan 1997). This is consistent with the hypothesis that it is involved in creating phenomenal experiences (Everitt and Robbins 1997). We will explore a second set of acetylcholine nuclei in the level 3 basal forebrain area.

2.2 *Norepinephrine vigilance system*

Parallel to the string of acetylcholine nuclei is a string of norepinephrine nuclei, running from the medulla through the upper pons. The bilateral pontine locus ceruleus (= blue place) norepinephrine cells also project to the cortex but through the lateral hypothalamus (Moruzzi and Magoun 1949; Mountcastle 1978) rather than through the thalamus. It projects to all areas of the cortex, but most diffusely to perceptual areas of posterior cortex and the right dorsal-lateral prefrontal lobe (Posner and Rothbart 1992).

This norepinephrine system is most excited during alert wakefulness; progressively less excited through drowsiness and slow-wave sleep; and then at its lowest activity during REM sleep (Aston-Jones and Bloom 1981; Hobson

1997). It is involved in directing the cortex to the outside world, creating a global mode of vigilant alertness to meaningful stimuli in the outside world (Aston-Jones and Bloom 1981a and b; Aston-Jones *et al.* 1986; Sara 1985; and Segal 1985; Fernandez-Duque and Posner 1997). Its projection to the right dorsal-lateral pre-frontal lobe seems to be for 'vigilance' (Posner and Rothbart 1992), to allow us to have selective attention (Robbins 1997), seemingly by interrupting the mental processing controlled by left dorsal-lateral prefrontal, to gear us to the outside world (Posner and Rothbart 1992).

Excessive firing of this circuit may be found in manic states, fear, hypervigilance, and negative emotional arousal (Chapman and Nakamura, in Watt 1998); while *hypo*-activity may be found in depression.

2.3 *Dopamine 'motivational' system*

In the midbrain are two large clusters of dopamine cells, the ventral tegmental area (VTA) and the substantia nigra (= black substance). The VTA cluster has reciprocal projections through the level 3 lateral hypothalamus to the nucleus accumbens and projects from there to the level 5 anterior cingulate gyrus and pre-frontal lobe. It also activates level 4 amygdala and hippocampus. This is the basic pathway involved in Olds' and Milner's (1954) 'reward' or 'pleasure' circuits (Persico *et al.* 1998). This dopamine system does not seem to mediate 'primary reinforcers,' such as food, but is somehow involved in 'tagging' new objects and experiences with positive valence, linking primary reinforcement rewards with novel stimuli, so that newly conditioned stimuli lead to operant responses in the future (Salamone *et al.* 1997; Grace, in Watt 1998). This seems to be even more specifically a *motivational* circuit than the norepinephrine circuit — so that any thought, perception, experience, or action motivating voluntary action is linked to this pathway.

This VTA-accumbens dopamine circuit is involved in cravings and psychological dependence on psychoactive drugs, such as cocaine, opiates, nicotine, PCP and marijuana (Wise 1996), which are likely habit forming because they act on brain circuits that subserve more natural and biologically significant rewards. The suppression of DA receptors has been implicated in cocaine-related depression (Malison *et al.* 1998; Lerman *et al.* 1998).

The other major midbrain dopamine projection system, containing three quarters of all the dopamine in the brain (Thompson 1993) is just dorsal to the first. It is the dopamine cluster that is depleted in Parkinson's Disorder (PD). It projects from the substantia nigra of the midbrain to the basal ganglia. This DA projection is crucial for the starting, shifting and stopping of all voluntary movement and the conscious orchestration of novel, unpracticed movements.

2.4 *Serotonin emotional-tone system*

Like the norepinephrine and dopamine systems, but unlike the acetylcholine system, the brainstem-to-cortex serotonin pathway also projects through the hypothalamus and is active during wakefulness, but relatively inactive in both slow-wave and REM sleep (Hobson 1997). This pathway begins in the 'raphe (=seam) nuclei,' which run the length of the brainstem, and projects to wide areas of the cortex, but most profusely to the frontal lobe. These projections also go to the basal forebrain, hippocampus, basal ganglia, amygdala, and to areas involved in the conversion of serotonin into melatonin to create drowsiness and sleep.

Serotonin has a complex relationship with many negative emotions and behavior, such as depression (McAllister-Williams *et al.* 1998; Blier *et al.* 1997; Sonino and Fava 1996); suicide (Bakish *et al.* 1997; Pivac *et al.* 1997); anxiety (Hedgess, *et al.* 1996); premature ejaculation (Waldinegar *et al.* 1998); Obsessive Compulsive Disorder (Laird 1996); Pre-Menstrual Syndrome (Kouri and Halbreich 1997); negative symptoms of schizophrenia (Silver and Shmugliakov 1998); pathological crying (Low and Chong 1998); and aggression, conduct disorder, and alcoholism (Sander *et al.* 1998). In some of the above disorders, the serotonin effects on the disorder are distinct from its effect on depression. Serotonin systems also cooperate with the periacquaductal gray (PAG) endorphin system in lessening or even shutting down the transmission of the slow pain projections to the cortex.

Finally, much of the psychedelic effect of LSD is mediated by serotonin receptors (Leonard 1997, for a review). Much of this relates to perceptual-system hallucinations. The marked distortions of time and motivation might be related to the distorted toning of the pre-frontal areas by the serotonin systems. Because of serotonin's implication in depression and violence, this pathway seems to play the role of setting cortical tone at a mellow-euphoric level, through its own efforts and through its regulation of pain.

2.5 *Peri-acquaductal gray (PAG) — Endorphin emotional valence system*

Extending for some length in the very dorsal brainstem, surrounding the brain-stem 'aqueduct,' is the periacquaductal gray (PAG) area, an area rich in neurons releasing endorphins — endogenous morphine neurotransmitters. We have already hinted at the best known function of the PAG, its cooperation with serotonin systems in lessening or even shutting down the transmission of the slow/throbbing/emotional pain projections to the cortex.

We have already introduced the most primitive pain-system, the pain reflex system that brings about direct, immediate, and non-conscious responses to minimize tissue damage. Second in the sequence of nervous system evolution is

the slow-pain emotional system (Thompson 1993). C-fiber pain receptors bring pain information regarding skin or organs to the spinal cord or brainstem, synapsing in the Substantia Gelatinosa (SG) zone of the spinal cord, releasing the neurotransmitters Substance-P and glutamate into the synapses of neurons that will then transmit the pain information toward the brain (Thompson 1993). In the SG are many tiny inter-neurons that, when stimulated from above, release endorphin neurotransmitters onto the pre-synaptic pain neurons, inhibiting their release of Substance P and glutamate — thus attenuating or even stopping the transmission of pain messages.

Spinal cord pain transmission can be inhibited from above only after it notifies higher areas of the first pain stimuli. It does this by projecting to various parts of the brain-stem reticular formation, including the endorphin PAG area. PAG endorphin neurons activate lower-brainstem serotonin raphe neurons (Leonard 1997), which project down into the SG columns in the spinal cord, activating in turn the endorphin inter-neurons in the SG that inhibit the release of Substance P, thus keeping the bodily system from being overwhelmed with pain perception.

The slow-pain messages that do get through continue to project to the PAG and other brain stem areas. The PAG and the brainstem reticular areas are all in mutual contact. Slow-pain pathways continue up to the level 3 ILN-thalamus and hypothalamus and to level 4 amygdala and level 5 posterior-ventral-medial pre-frontal and cingulate gyrus. These areas project back to the reticular core and PAG (Agggleton, Burton, and Passingham 1980; Newman 1997).

This extensive circuit allows several things. It allows bottom-up messages of intense pain to trigger an automatic spinal-cord/brainstem/PAG-Raphe/spinal-cord loop to attenuate pain messages. It also allows various top-down messages signaling 'mental pain' or setting a broader context that can either increase or decrease the pain thresholds of the lower circuit. The level 2 PAG also serves as a way-station for level 3 hypothalamic control of the level 1 servo-mechanism autonomic nervous system (Chapman and Nakamura, in Watt 1998).

Finally, Panksepp (in Watt 1998) has long championed the PAG area as the most basic core of emotion and as part of a brain-stem implicit-self system, and thus as a bottom-up emotional valence tagger — tagging experiences as subjectively negative (painful) or positive (pleasurable). In fact, Panksepp reports that total destruction of the PAG in some species eliminates all detectable consciousness.

2.6 *Cortical arousal systems, emotion, and motivation*

I propose that these five and other systems within the central nervous system determine the very *existence* and *direction* (Hobson 1997) of consciousness (being conscious of external objects while awake versus dream objects while in

REM versus no objects while in deep coma), and some of the *qualitative* dimensions, specifically whether objects are discerned in a realistic or dreamy 'trip' way, and some of the quality and quantity of emotional states.

Several of these brainstem systems are implicated in depression, most clearly serotonin and norepinephrine, but likely also dopamine, endorphins, and even acetylcholine (Neumeister *et al.* 1998; Leo 1996; Miyai and Reding 1998; Robbins 1997; Leonard 1997 for a review). It is very likely that the hypo-functioning of most of these systems — and the compensatory hyper-sensitiza-tion of their receptors — plays a specific role in depression. I speculate in my Abnormal Psychology classes that norepinephrine's lack leads to lethargy, a reluctance to wake up, and an inability to move; while its super-abundance in mania may be what causes racing thought, speech, action, and the inability to sleep. Dopamine's lack may relate to the anhedonic nature of some depression, where nothing brings pleasure anymore. Serotonin's lack is the most correlated with suicide. Serotonin might be essential for establishing a normal emotional mood. Since endorphin projections play a role in reinforcement, their lack or dysregulation likely plays a role in depression. Acetylcholine's lack may play some role in the slowing of thought processes in depression.

The norepinephrine, dopamine, serotonin, and endorphin cortical arousal systems seem to have overlapping as well as distinct emotional/motivational components. All four of them seem to play a role in motivating us to perceive and respond to the *external world*. In contrast, it seems likely that the basic phenomenal-consciousness engine, the acetylcholine arousal system, is, in itself, much less involved in emotional systems. This system seems to underlie states of consciousness that range from the most emotional to the least emotional and from the most externally to the most internally oriented.

However, the acetylcholine system has an interesting and paradoxical role in motivated motor behavior. In REM sleep (where the acetylcholine system is the predominant player), afferent acetylcholine projections cause the motor cortex to be very active, with the latter presumably sending commands to motor neurons to activate muscles to act out the dream; while the efferent acetylcholine areas drastically lower muscle tone so that muscles do not make meaningful movements. This might be considered a self-aborted motivational system. While the autonomic system supplies much of the quantitative intensity of emotions, the dopamine reward circuits and PAG may supply the basic qualitative dimension of emotions.

3. Level three diencephalon areas to which level 2 projects

The third level incorporates the hypothalamic/thalamic/basal-forebrain/septal areas, to which level 2 systems project, both to the projections and to other

features of the area. We will deal with this area in the following order: (1) acetylcholine intralaminar nuclei (ILN) and reticular nuclei of the thalamus for attentional gating, (2) dopamine nucleus accumbens staging area for higher attentional control, (3) acetylcholine basal forebrain cognitive enabler, (4) hypothalamus as a level 3 motivation controller, and (5) perceptual-thalamus as an information processor.

3.1 *Acetylcholine intralaminar nuclei thalamus (ILN): Attentional gating*

Level 2 brainstem acetylcholine nuclei activate the level 3 intralaminar nuclei (ILN) of the thalamus, which in turn arouse the level 5 cortex diffusely to better receive and process the specific information coming from the thalamus. This may be the basic arousal system or merely part of the team of arousal systems.

The brain-stem acetylcholine system is also involved in 'gating.' ILN activation interacts with the 'reticular' nuclei of the thalamus (nRT, a mesh-like covering over much of the thalamus) which seems to operate as an attentional filter to activate specific thalamic nuclei, which can then transmit close to 100% of their incoming sensory information to the sensory areas of the cortex (Steriade and McCarley 1990; Crick 1984; Newman 1997). The nRt gates perceptual input into consciousness through a cortex-wide array of *thalamocortical loops:* perceptual- thalamus to cortex to nRt-thalamus to perceptual-thalamus, with collateral projections to the 'shell' of the nRt (Crick 1984; Newman 1997; Newman, in Watt 1998). The nRt uses the neurotransmitter GABA for its inhibitory contacts (Crick 1984). These loops are activated with '40 Hertz' oscillations, which bind their contents into units of conscious experience (Crick and Koch 1990; Llinas *et al.* 1994, 1996), an effect that is enhanced by selective attention and attenuated by habituation (Tiitinen 1995).

3.2 *Nucleus accumbens staging area for higher attentional control*

The VTA-dopamine-fed nucleus accumbens has a major relationship with the reticular nuclei of the thalamus (nRt), which we have just described as being central to attentional 'gating.' Accumbens projections act on nRt neuron dendrites to globally control the flow of sensory inputs to cortex (Newman, in Watt 1998). Level 4 amygdala and hippocampus and level 5 dorsal-lateral prefrontal exert their attentional control downward through this level 3 ILN by way of accumbens.

Disorders in this dopamine system are implicated in schizophrenia (Thompson 1993; Posner and Rothbart 1992; Csernansky and Bardgett 1998; Heresco-Levy *et al.* 1996), especially with *excess* release of dopamine in the nucleus accumbens (Gray 1998) and a *deficit* in its dopamine projection to pre-frontal cortex (O'Donnell and Grace 1998; Grace, in Watt 1998), presumably allowing

subcortical areas such as the amygdala to dominate attention and behavior, with less pre-frontal control.

3.3 *Acetylcholine basal forebrain cognitive enabler*

We have mentioned a second set of acetylcholine nuclei. This is in the level 3 basal forebrain area between ventral-medial frontal lobe and the hypothalamus and not far from the nucleus accumbens. Destruction of this second set of acetylcholine nuclei has long been implicated in Alzheimer's Disease, through the deterioration of acetylcholine projections to the amygdala, hippocampus, and areas of the cortex.

These basal forebrain acetylcholine projections are also involved in the modulation of arousal and attention (Nobili and Sannita 1997) by enhancing the cortical processing of signals in each area to which it projects (Everitt and Robbins 1997; Robbins 1997), so that it enables the *hippocampus* to encode new memories (Woolf 1996); the *amygdala* to form fear and anger 'conditioned emotional responses'; and the *cingulate gyrus* to use response rules through operant conditioning discrimination (Everitt and Robbins. 1997), and so on.

3.4 *Hypothalamus as a level 3 motivation controller*

Just anterior and dorsal to the midbrain is the hypothalamus, which we have identified in its direct control over electrolytes in the blood and bodily fluids, its control over the hormonal system in general and the HPA axis in particular through the pituitary, and its control over the level 1 autonomic nervous system through the level 2 PAG. Specific nuclei of the hypothalamus control the drive states of various specific hungers, water and mineral-water thirsts, temperature regulation, sleep cycles, sex, stress control, and autonomic arousal. The hypothalamus is the primary homeostatic motivational center of the individual. Most of this homeostatic control is done completely non-consciously.

3.5 *Perceptual-thalamus as an information processor*

In each perceptual modality there is processing on many levels, some of which does not contribute directly to perceptual consciousness; some of which remains 'subliminal' or non-conscious but directly contributes to perceptual consciousness; and yet other of which is crucial for the very existence of perceptual consciousness. We have already noted that sensory projections to the midbrain colliculi lead to immediate non-consciously-activated withdrawal from the threatening sight, sound, or what have you.

Sensory projections through the thalamus to the cortex — which are crucial

for the very existence of perceptual consciousness — are quite distinct from the colliculi paths. Perceptual-thalamus is a complex sensory way-station, where direct fibers from the optic tract and multi-synaptic projections from the auditory, gustatory, and tactile tracts project on their way to the primary visual, auditory, gustatory, and somatosensory areas of the cortex. The thalamus projects to cortical sensory areas for conscious perception. These cortical sensory areas, in turn, activate hippocampus circuits for the encoding of memories; amygdala, hypothalamus, and autonomic nervous system circuits for the expression of emotions; and various motor systems for motor responses.

The perceptual-thalamus also incorporates a fast-pain touch-perceptual system. This system goes directly from the spinal cord to an area of the thalamus that is contiguous to touch areas; and from the thalamus to the parietal lobe somatosensory system, again contiguous to the other aspects of the touch system. This seems to be a 'non-emotional' pain system that has developed to give us precise information as to where tissue damage is occurring. This seems to be the most conscious and cognitive part of the pain system.

But perceptual-thalamus can, on its own, operate many of the same memory, emotion, and action mechanisms that the cortex operates. The thalamus does this by running parallel loops. In parallel with the loops from the thalamus to cortex to these various systems, the thalamus projects directly to the hippocampus, amygdala, and motor systems, controlling most of the same memory, emotion, and motor mechanisms (Wilson *et al.* 1983; reviewed in Faw 1987). Given the limited size of each thalamic sensory area and the limited sensory field feeding into each thalamic neuron, it is assumed that the thalamus activates these memory, emotion, and motor systems in a much less selective and precise way than does the cortex.

We have now seen a very sophisticated four-way action of the thalamus in relationship to the sensory systems below and the cortex above. To return to our television analogy, *perceptual-thalamus* sends specific sensory information from various 'channels' (1) to the cortex for conscious perception and (2) to sub-cortical areas for immediate responses; (3) *the reticular-nuclei-thalamus* operates as a channel selector for perceptual-thalamus, allowing many areas of the nervous system to compete in choosing which information the cortex will attend to; and (4) the *intralaminar-nuclei-thalamus* operates as on-off-volume switch, determining the level of activation of the cortex so that it can receive the channels.

4. Level four subcortical limbic emotional mechanisms

The fourth level incorporates the amygdaloid and hippocampal formations and related sub-cortical parts of the 'limbic system.' Virtually all levels 2 and 3 areas

reciprocally connect with level 4 structures. We will focus on the amygdala, with a brief discussion of the hippocampal area.

4.1 *Amygdala as a fourth-level emotional evaluator and conditioner*

Viewed from the side, the human brain looks something like a boxing glove. The frontal lobe is the front of the glove and the temporal lobe the thumb. The hippocampus, amygdala, and parts of the basal ganglia are like the human thumb — with the amygdala near the thumb nail — within the glove's thumb.

Basic to understanding the amygdala is to note its major reciprocal connections (Aggleton, Burton, and Passingham 1980). It triangulates: (1) external-world-perceptual information from each sensory system and from multi-sensory areas from perceptual areas of the thalamus and temporal lobe (Herzog and van Hoesen 1976); (2) internal-world-motivational-visceral information from the PAG, thalamic ILN, septal-area substantia innominata, reward circuits through the lateral hypothalamus, and each of our levels 2–3 neurotransmitter systems (Carpenter and Sutin 1983; Everitt and Robbins 1997); and (3) anterior motivational/autonomic control areas: temporal pole, anterior insula, posterior-ventral-medial pre-frontal, and anterior cingulate gyrus.

This convergence suggests to Aggleton and colleagues that the amygdala plays a role in "integrating information about the sensory aspects of stimuli and their motivational and emotional significance" (365). In their view, the cortical perceptual areas that feed to the amygdala play the role of object recognition through sight, touch, sound, smell, and taste. Then the amygdala associates objects with rewards — so that "visual stimuli … gain motivational and emotional significance" (Jones and Mishkin 1972; Herzog and van Hoesen 1976; van Hoesen and Pandya 1975; Turner *et al.* 1980; Leonard *et al.* 1985). Related to this is the claim that the amygdaloid complex is a key modulator of memory consolidation for emotional events, although not the site of storage (Cahill and McGaugh 1998).

This view of amygdala functioning fits with primate studies. Monkeys with lesioned amygdalae or posterior-ventral-medial pre-frontal cortex (or with severed connections in or out of the amygdala) are markedly impaired in simple one-trial learning of *object-reward* or *place-reward* associations with trial-unique junk objects and places, especially if the rewarded object or place was reversed from the previous trial (Jones and Mishkin 1972; Spiegler and Mishkin 1981; Faw 1987 for a review). Such monkeys also have retarded acquisition of conditioned avoidance, conditioned suppression, matching to sample, and learning sets (Jones and Mishkin 1972). All of this suggests to them that amygdaloid lesions cause stimuli to lose their aversive or attractive properties — what they call "psychic blindness" or a poly-sensory emotional 'agnosia.'

This relates to what we noted with the dopamine VTA — nucleus accumbens reward circuit, that it is somehow involved in 'tagging' new objects and experiences with positive valence, linking primary reinforcement rewards with novel stimuli, so that the newly conditioned stimuli lead to operant responses in the future (Salamone *et al.* 1997; Grace, in Watt 1998). There is conflicting evidence regarding the brainstem endorphin PAG, as to whether it merely relates to innate emotional responses or participates in the learning of conditioned (primarily defensive) emotional responses (Panksepp, in Watt 1998). It may be that the level 2 and 3 PAG and VTA circuits handle both innate and learned emotional stimuli, but that it takes the level 4 amygdala to actually learn to associate new (conditioned) stimuli with wired-in (unconditioned) stimuli that then lead to conditioned emotional responses.

The amygdala has been best known for its roles in *fear and anger responses*. Animals with experimental lesions in the amygdala may show no fear or anger toward objects and animals which previously elicited such emotional responses — the Kluver-Bucy Syndrome (Kluver and Bucy 1939; Carpenter and Sutin 1983; Rolls *et al.* 1977; Correll and Scoville 1965; Jones and Mishkin 1972; Mishkin and Delacour 1975; Mora *et al.* 1976; Weiskrantz and Warrington 1979). In contrast, variable intensities of electrical stimulation of the amygdala can either block or elicit environmentally inappropriate fear or anger (Carpenter 1983). While the electrically-elicited anger has been called "sham rage", there is evidence that it has emotional quality as well as motor responses in it (Watt 1998).

The most common response to electrical stimulation of the amygdala in alert animals is an 'arrest' reaction, in which spontaneous behaviors cease and the animal snaps to aroused attention. This is identical in behavior and in EEG to what happens with brainstem reticular formation stimulation (Carpenter and Sutin 1983). This 'arrest' reaction is assumed to be the first phase of either fight or flight reactions — presumably the phase during which the animal determines which response is the more prudent.

The amygdala has been found to have *distinct centers* triggering fear and anger responses to perceptual stimuli (Ursin and Kaada 1960; Carpenter and Sutin 1983; Spiegler and Mishkin 1981). Each sensory system projects, in ordered form, to both the fear and the anger areas of the amygdala — much of it through the cortex of the temporal pole (Aggleton, Burton and Passingham 1980). Thus, the amygdala is key to both stimulus evaluation of threat and the production of defensive responses (Kasniak, in Watt 1998; Faw In press).

As part of its task of determining which objects warrant which emotional responses, the amygdala is essential for *recognizing fear* (Broks *et al.* 1998; Morris *et al.* 1996) and perhaps *anger* (Adolphs *et al.* 1998a) in the *faces* of others, required for accurate social judgments based on others' facial appearance. This recognition of fear and anger in the faces of others is found even in species

such as monkeys that are incapable of recognizing their own faces in a mirror (Gallup 1968 1999; Faw 1987 for review).

This face-expression recognition is mediated by a string of *face-receptive cells* at various stages in the visual system (Kanwisher 1999; Perrett *et al.* 1982; Baylis *et al.* 1985 1987; Desmione *et al.* 1984; Rolls and Baylis 1986; Leonard *et al.* 1985; Perrett 1999), and in an area of amygdala that is a major convergence area for levels 2, 3, and 5 information (Leonard *et al.* 1985; Faw 1987). Failure of monkeys or persons with amygdaloid lesions to react appropriately to faces is "related to damage to or disconnection of a system in the amygdala which is concerned with emotional responses to faces and which receives face-specific inputs from the superior temporal sulcus region in which face processing occurs" (Perrett *et al.* 1982: 341). Damage here may disrupt social behavior in the dominance hierarchy and other social functioning (Leonard *et al.* 1985).

While most of the face cells in the amygdala have long latency, clinical evidence suggests a likely manifestation of the direct *thalamic* triggering of perceptual processing in terms of face recognition, seen in a paradox sometimes found with *prosopagnosia*. Prosopagnosics cannot recognize even previously familiar faces, but some will show sympathetic nervous system arousal directly proportional to the previous familiarity of these faces which they no longer consciously recognize. A less pathological experience likely associated with this thalamic projection would be to meet someone whose very looks make you fearful and who looks 'familiar'; but you don't yet realize that this person looks a lot like a bully who used to extort your lunch quarters in fourth grade. The thalamus sends enough visual information to the emotion systems to trigger your fearful response — even while your cortex is trying to figure out why you are having such an emotional response. In short, this direct thalamic control over memory, emotion, and motor systems appears to explain why so much of the processing of these action systems is non-conscious.

The amygdala presumably uses thalamic (or thalamic-hippocampal) perceptual information to trigger quick emotional responses: to form immediate 'first impressions' and survival-preserving 'gut feelings.' Temporal-lobe information is used by the amygdala for activating conscious emotional responses and the formation of new stimulus-reward associations. Recent evidence suggests that the *left* amygdala might specialize in responses to consciously perceived faces and the *right* in responses to subliminally-presented faces (Morris *et al.* 1998).

The amygdala is well equipped to initiate emotional responses. It activates various motor-cortex-basal-ganglia loops to produce facial emotional expressions (Adolphs *et al.* 1998b; Cherrier *et al.* 1997). The amygdala projects to the hypothalamus to alter hormonal responses through the pituitary and autonomic responses through the dorsal PAG (Panksepp and Chapman and Nakamura, in Watt 1998) to activate bodily responses appropriate to anger (fight) and fear

(flight). Adrianov (1996) sees the close interrelationships between the hypothalamic nuclei and the amygdala as the basis for the *mutual exchange of motivational and emotional messages*.

4.2 *Hippocampus as a level 4 novelty evaluator episode recorder*

The hippocampal complex has many tasks, central of which seem to be: (1) to compare new perceptions with memory traces of old experiences (Nadel and Moscovitch 1998) in order to help shift attention to unexpected and meaningful stimuli and (2) to encode new memories, especially of the '*episodes*' of one's life. Many '*procedural*' memories and Pavlovian-conditioned autonomic responses are processed and stored in the cerebellum (Thompson 1993). In addition to these, many '*semantic*' memories that do not have episodic context seem to be processed and stored directly in perceptual and language posterior regions. It appears to be the episodes of one's life, the memories with context, that most involve the hippocampus. The episodes or at least the 'library cards' for them are likely stored in the Temporal Pole — the end of the boxing glove thumb.

The sudden appearance of intense stimuli, such as a bright light, loud sound, smell, or smoke, seem to directly activate '*involuntary' attention* cells within the midbrain colliculi and within cortical perceptual areas. Intense-stimulus types of attention-grabbing events do not need to rely on memory — but operate from wired-in responses needed for survival. But, something grabbing one's attention because it is 'unexpected' or 'personally meaningful' does depend on memory: the former on very short-term memory; the latter on either short-term or long-term memory. This may be why the hippocampus is involved.

Level 3 perceptual-thalamus supplies perceptual information both directly to level 4 hippocampus and amygdala (resulting in 'short latency' perceptually-responsive cells that fire a mere 25 ms or so after stimulus presentation) and indirectly through the higher perceptual processing of the cortex (resulting in 'higher order' cells firing some 200 ms after presentation — Wilson *et al.* 1983). The hippocampus and amygdala are also in reciprocal interaction. My own speculation is that the hippocampus begins using short-latency *thalamic* information for comparing new perceptions to test for unexpected or meaningful new events and then incorporates long-latency temporal-lobe information for revisions in stimulus testing and for encoding into new memories.

5. Level five anterior cortical control areas

5.1 *Pre-frontal lobe as an 'executive committee'*

Baddeley (1992) originated what has become doctrine in cognitive psychology,

the concept of Working Memory (WM) as more than just a 'consciousness of what's happening now,' but as having three major components: *two streams* that bring information into consciousness and serve as *loops* in processing the information that is there and a '*central executive*' which controls the loops — both to maintain information in WM for further processing and to encode the information into long term memories. The two streams are called the '*visuo-spatial sketch-pad*' and the '*auditory articulatory loop*.' These bring in sensory information and allow one to use mental imagery and word-thinking to process experiences. These are controlled by a '*central executive*,' which is linked to the frontal lobe — generally to the dorsal-lateral prefrontal cortex.

In a recent journal interaction, Parkin (1998) debated Baddeley (1998) with Parkin stating that the "central executive does not exist". Instead, there is a diverse pattern with different executive tasks related to different cortical areas. Baddeley responded that the central executive should not be seen as an 'organ' that 'exists,' but as a concept that explains. He applauded dividing the concept into subsystems and seeking the brain interactions involved.

Their debate leads me to coin the phrase that the prefrontal portion of the frontal lobe is not an 'executive,' but an '*executive committee*.' In a development of the Baddeley model, I have suggested (Faw 1999) that there are *five streams* coming from posterior cortex and subcortical areas to *five pre-frontal executive regions*. Each pre-frontal member of the executive committee represents one set of 'constituents' and chairs at least one 'sub-committee' that can do a lot of work on its own (without committee approval), but can also make its case (in certain circumstances) for taking over central control of *conscious attention* and *willed action*. The members of a sub-committee include the posterior *input* areas and two types of *output* mechanisms under the control of each pre-frontal area: *motor sequence processing loops* from prefrontal → basal ganglia → thalamus → back to the prefrontal area and a *motor control path* from prefrontal → supplementary/premotor → motor strip. Many of the anatomical distinctives and tasks that I will suggest derive from Passingham (1995).

My "executive committee model" was written in outline form for the 1999 conference of the Association for the Scientific Study of Consciousness and is now being developed in article form. An outline of this model will suffice here, briefly defining each stream in terms of its pre-frontal area, its constituents, and its basic 'executive committee' tasks. I will then focus on two streams that have a special role in the control of emotions. (All references to cortical "area xx" refer to the Brodmann numbering of lateral and medial cortex areas presented in the diagram.) The names I give these pre-frontal areas are a bit eccentric, in an attempt to be more precise than people often are in designating pre-frontal areas.

STREAM ONE: *Area*: Anterior-Ventral-Orbital Prefrontal (frontal pole and ventral convexity — areas 11 and lateral 12 and possibly more). *Constituents*:

visual, auditory, and touch information about objects from sensory processing areas on or bordering the temporal lobe. *Task*: represent the external world in 'object' coordinates — 'What Is It?' ventral perceptual stream.

STREAM TWO: *Area*: Posterior-Ventral-Medial Prefrontal (areas 13 and 14 and medial 12). *Constituents*: taste, smell, bodily feelings, mood, and emotion input from the amygdala, entorhinal cortex, temporal pole and insula. *Task*: represent one's internal world in 'emotional/self' coordinates — 'What's It to Me?' stream.

STREAM THREE: *Area*: Broca Area-plus (dominant hemisphere areas 44 and 45 and a supplementary dorsomedial speech area anterior to SMA Area 6, Passingham 1995). *Constituents*: auditory and visual language and word-thinking processing areas 22, 39, and 40. *Task*: represent one's mental world in language coordinates — 'What Do I Think About This?' stream.

STREAM FOUR: *Area*: Anterior Cingulate Gyrus: areas 32 and 24. *Constituents*: object-evaluating and memory-encoding circuits from the hippocampal formation, posterior cingulate gyrus, and amygdala (to some extent). *Tasks*: (1) represent external and internal worlds in historical coordinates — 'What Response Worked Last Time?' stream — and (2) help focus attention onto what in current experience is most relevant — 'What is Worth My Attention?' stream.

STREAM FIVE: *Area*: Dorsal-Lateral prefrontal: areas 46 and 9. *Constituents*: somatic and visuo-spatial (and visuo-object?) information from areas in or bordering the parietal lobe — the 'where is it?' dorsal perceptual stream. *Tasks*: (1) keep Working Memory loops active (= 'run the meeting') until action is 'willed' — 'When Have we Made the Decision?'; and (2) represent the external world in (body, limb and eye) action coordinates in order to be the principal initiator of willed action — 'Where Is It, So I Can Carry Out the Decision?'

The dorsal-lateral prefrontal area, especially on the left side, clearly chairs the executive committee, but as a 'prime minister' not as a dictator. The prime minister can be outvoted. The two streams that most represent one's *emotional/motivational self* in committee decisions are stream two (posterior- ventral-medial pre-frontal) and stream four (anterior cingulate gyrus). Let us examine their contributions to committee life and see how they can 'gang up' on the committee's prime minister.

5.2 *Representing the emotional/motivational self in committee decisions*

Stream Two posterior-ventral-medial prefrontal (PVM: areas 13, 14 and medial 12) represents one's internal world in 'emotional/self' coordinates. It does this by receiving *taste* (from the parietal-temporal juncture) and *smell* (directly from olfactory bulb) input (which 'represent' the external world, but as to how it 'tastes and smells to me') and by receiving *interoceptive* input — reports of

stomach aches, tiredness, heartburn, mood, and emotion — from the amygdala and insula (Passingham 1995; Luu and Pribram 1997). Luu and Pribram consider this the 'ventral amygdaloid — ventral pre-frontal limbic trend'. "Early emotional experience with a caregiver may help develop the synapses and axons linking brainstem VTA, hypothalamus and right orbital systems" (Shore 1994, cited in Watt 1998).

We have noted how the brainstem serotonin raphe and the endorphin PAG cooperate to stop the *emotional-slow-pain system* 'dead in its tracks' in the spinal cord dorsal horn and that slow-pain pathways continue up to the level 3 ILN-thalamus and hypothalamus, to the level 4 amygdala, and from there to level 5 anterior insula, PVM, and anterior cingulate. It is this slow-wave system that is inhibited by the release of endorphins or the taking of morphine-related pain killers, by the use of hypnosis for pain control and medical operations, and that is severed in pre-frontal lobotomy.

This 'what is it to me?' stream can be said to chair the '*emotional working memory*' (LeDoux, in Watt 1998) s*ub-committee*. It allows the pre-frontal committee to be notified of the 'threat' or 'reward' aspect of incoming stimuli (Passingham 1995; Luu and Pribram) or of one's current stream of thought; allows bodily feelings and emotions to compete for attention and willed action; helps interpret social situations and make social judgments (Fletcher in Vogeley 1999; Cicerone *et al.* 1997; Vogeley 1999, Adolphs *et al.* 1998); and helps process all emotions (Reiman 1997; Damasio, in Watt 1998) and the emotional aspects of pain (Peyron *et al.* 1998). This area is activated when people make 'gut decisions' as in the 'Prisoners' Game' (Bechara *et al.* 1998). And it may be the most important area in differentiating people according to basic personality traits (Fletcher, in Vogeley 1999; LeDoux, in Watt 1998; Vogeley 1999; Taylor, in Watt 1998; Soloff 1998; Viinamaki *et al.* 1998).

Dysfunction in this PVM area and its basal-ganglia-thalamic loops have been implicated in a variety of disorders such as Obsessive Compulsive Disorder (Wilson 1998); Attention Deficit Hyperactivity Disorder; moriatic aphasia (Ghika-Schmied *et al.* 1998); depression (Elliott *et al.* 1998; Busch and Alpern 1998); personality disorders (Viinamaki *et al.* 1998); anxiety disorders such as generalized anxiety, phobia, and PTSD; alcoholism; autism; schizophrenia; and immature impulsive behavior — the 'Phineas Gage' condition. In many of these disorders, *several* pre-frontal areas (and their streams and motor loops) are involved in hyper- or hypo-functioning to produce some parts of the symptoms of the disorder. Untangling that is a busy task.

Underlying much of the above is that the PVM area is involved in associating *new responses* to rewards and punishments (Passingham 1995). We have seen that monkeys with lesioned amygdalae or our PVM area are markedly impaired in a simple test of one-trial learning of object-reward or place-reward associations,

especially if the rewarded object or place was reversed from the previous trial. The basic role of the amygdala in the processing of emotion seems to be *object-reward* association learning.

Spiegler and Mishkin (1981) found that the *amygdaloid* impairment was distinct from the *PVM* impairment. While amygdalectomy impaired object-reward association learning, these frontal lesions led to a *difficulty in suppressing the previously established habit*. This leads to perseveration (continuation) of already established stimulus-reinforcement associations. This is equivalent to Passingham's (1995) conclusion that the amygdala associates *objects* with rewards/punishments, while this prefrontal area associates *responses* with rewards/punishments. That is, the amygdala determines that this new situation will get the 'same old' anger response given in similar old situations; but the PVM allows one to break that pattern of response and give a new response to the current situation.

Resuming our executive committee analogy: motivational/emotional information from taste, smell, and internal body sensors feeds into the amygdala to be associated with perceptual information about objects and persons in the outside world. The amygdala takes innate or previously-learned immediate action on some innate or previously learned object-reward/punishment associations, forms new associations, and projects profusely to our PVM pre-frontal area to incorporate into voluntary focused attention and deliberate action these object-reward/punishment associations and previously learned responses to these objects and situations. The PVM pre-frontal area takes these previously-learned object- and response-associations to the executive committee. But the PVM is responsible both to its constituents and to the full committee. If the full committee determines that the previously-learned motivational/emotional responses to a certain situation are 'too childish,' 'immoral,' 'too time consuming,' or just 'stupid,' the PVM might have the task of inhibiting the old and learning a new response to the situation. A lesion in the PVM or disconnecting the PVM from its constituents or from the rest of the executive committee areas can lead to different types of disorders — resulting in spontaneous childish impulsive behavior, in repetitive compulsive behaviors, or in social cluelessness.

5.3 *Representing the 'days of our lives' in committee decisions*

Stream Four to the Anterior Cingulate Gyrus (areas 32 and 24) represents the classic 'limbic lobe' medial arc, from the parahippocampal gyrus (areas 26–30 with hippocampal complex within), the posterior cingulate gyrus (areas 31 and 23), to the anterior cingulate gyrus. Luu and Pribram (1997) consider this the "dorsal hippocampal — septal — dorsal cingulate limbic trend".

The input this stream brings to the 'committee' must relate to the two tasks

we noted for the hippocampal complex: situation-evaluation (comparing new experiences with memory traces of old experiences) and the recording of these new episodes. Hippocampal *situation-evaluation* is used both for the more automatic posterior 'involuntary attention' systems — external-stimulus-cued attention (Posner and Rothbart 1992) — and for the more deliberate anterior 'voluntary attention' system — internally-cued attention (Yamaguchi and Kobayashi 1998; Posner and Rothbart 1992) — the best known role of anterior cingulate.

The hippocampal role of *encoding new episodic memories* seems to be widely distributed among executive committee members. The left dorsal-lateral prefrontal (which is interconnected with the hippocampus and the anterior cingulate — Passingham 1995) has been implicated strongly in the encoding of new episodic memories (Vogeley 1999; Gur *et al.* 1997; Ferreira *et al.* 1998), presumably by holding information on line while the bilateral anterior cingulate focuses attention and the hippocampal complex forms the new memory. Anterior cingulate focus of attention enhances the pathways subserving the attended objects and attenuates pathways of non-attended objects (Posner and Rothbart 1992).

'Connecting the dots' from much of this chapter, the *mechanisms of attention* are something like the following (Newman 1997; Watt 1998). In *external-cue-activated attention*, perceptual cortex and comparing mechanisms in the hippocampus note sudden intense, unexpected, or meaningful stimuli and activate brainstem ERTAS systems to stimulate the intralaminar nuclei (ILN) of the thalamus to activate the cortex and open specific gates in the posterior reticular nuclei of the thalamus (nRt), allowing perceptual information to flow readily through attended gates while inhibiting flow of information through other gates.

The anterior cingulate gyrus projects to vast portions of the anterior nRt gates — about equal to the entire perceptual projection to the thinner posterior nRt gates — presumably to activate *internally-directed attention* (Watt 1998). The nucleus accumbens in basal forebrain also projects to the nRt gates. The accumbens receives projections from the dorsolateral prefrontal lobe, the hippocampus, and the amygdala. The accumbens may, thus, be a higher 'gate' for levels 4 and 5 input to the nRt (except for the anterior cingulate that contacts the nRt directly). This, presumably, allows the accumbens to tip the balance between 'present planning/working-memory' dorsal-lateral, 'past experience' hippocampal, and immediate threat amygdala activation of attention (Watt 1998).

In addition to the anterior cingulate activating voluntary attentional mechanisms, it likely brings to the committee some hippocampal recollection of the episodic historical context in which current experiences can be evaluated, so that deliberate responses can be enacted within a much broader perspective of one's history of responses than the amygdala-PVM immediate-crisis responses. The PVM and the anterior cingulate assert control over the autonomic nervous system through output to the amygdala and hypothalamus.

5.4 *Insurrection in the executive committee*

We have claimed that the prime minister, Lord Dorsal-Lateral Prefrontal Cortex, can be ganged up on and outvoted. (Total overthrow of the prime minister must be what happens in schizophrenia!) We have also seen that activation of the amygdala can lead to an 'arrest' response, where all spontaneous activity is stopped and the animal becomes hyper-vigilant. In addition, stimulation of the anterior cingulate and amygdala leads to the inhibition of spontaneous movement (Passingham 1995). But how does all of this lead to insurrection?

Marshal and colleagues (1997) have run a marvelous PET scan study that shows this well. They ran PET scans on a patient with 'hysterical paralysis' (called "conversion disorder" in DSM-IV), who had not been able to move her left leg since a traumatic event, but had no physiological reason to be paralyzed. When moving her good right leg, they found PET activation of the normal pathway from dorsal-lateral prefrontal to premotor to motor areas and no anterior cingulate nor orbital (PVM) activation. When she attempted unsuccessfully to move her 'paralyzed' left leg, the PET showed prefrontal to premotor activation, but not right primary motor cortex. Instead, the right orbital frontal and right anterior cingulate cortex were significantly activated. Marshall and colleagues suggested that "these 2 areas inhibit prefrontal (willed) effects on the right primary motor cortex when the S tries to move her left leg" (p. B1). In the terms we have been using, the amygdala — PVM stream sensed that there was great threat in moving that leg and the hippocampus-anterior-cingulate stream recalled the traumatic incident that led to the paralysis. Together, they outvoted the dorsal-lateral prime minister in its attempted 'willed action' to move the left leg.

6. Revisiting our questions and assumptions

In this section, we come back to the big questions about the relations between motivation, emotion, and consciousness; between 'having phenomenal experiences' and being 'aware-of' such and such; and the question as to whether emotional consciousness is part of the core mechanisms of consciousness or a mere 'channel.' Let us begin by revisiting pain.

6.1 *Pain as a complex perceptual/emotional system*

The anatomy of pain represents a combination of perceptual, reflex, motivational, and emotional systems, showing a remarkable interweaving of non-conscious and conscious processing that can be decomposed into at least three systems, each of which we have encountered in this chapter: *pain-reflex systems* for each sensory

modality, the *slow-pain emotional system*, and a *fast-pain system for pain-perception*. In a sense, the three pain systems seem to 'represent' tissue damage in three different ways and have three distinct functions. The pain-reflex system does not bother to 'represent' the damage in any *conscious* way, but in an automatic, take-charge way. Its function is to minimize further damage. The pain-emotional system 'represents' the pain in very vague and global ways and has the function to make you feel so bad that you do something about your tissue damage. The pain-perceptual system 'represents' the pain in very precise analytical ways, so that if you are motivated to do something about the tissue damage (through the pain-emotional system), you can figure out where the damage is.

The pain-reflex system runs strictly non-consciously; the pain-emotion system runs in both conscious and non-conscious paths; while the pain-perception system is strongly conscious. The pain-reflex system uses primitive reflexes; the pain-emotion system uses middle-evolutionary limbic, emotional systems; and the pain-perception system uses high-tech cognitive systems. When things are functioning well, these three basic types of neural systems blend into holistic conscious experiences and well-tooled responses. When these things are split apart, reflexes and emotional and cognitive systems dissociate.

6.2 *Conscious and non-conscious aspects of motivation and emotion*

We have noted that much of *motivated* behavior comes from completely non-conscious homeostatic control enacted through motor neurons, constriction and relaxation of blood vessel walls, and increasing or decreasing various substance levels in the blood. But there are at least two ways in which some of this non-conscious motivated behavior reaches consciousness.

Some of the motivational systems *utilize the production of conscious feelings* — such as hunger, thirst, feeling cold or hot, and sexual arousal — to aid in the *production of more precise motivated responses*. Therapists assume that the conscious aspect of such feelings has a causative role, so that, e.g., some obese patients are trained to become better monitors of their 'feelings of hunger' in order to better control their food intake. These conscious motivational feelings tend to trigger very strong emotional responses from humans, so that hunger or thirst might produce panic and sexual desire might produce all kinds of emotions, including panic. The *fulfillment* of such motivational drives is also often bathed in emotions.

In addition, *feedback* from much of the results of automatic motivational control is capable of entering consciousness. For instance, we are probably not consciously aware of how we speed up our heartbeat or breathing due to sympathetic arousal, even when we use biofeedback, but a sudden increase in

heartbeat or breathing rate might intrude upon our consciousness. Consciousness of such an automatic change in our functioning seems to be very valuable in our making a conscious decision to relax, take a deep breath, or whatever.

Just as we have looked at the division of conscious and non-conscious processing in the motivational systems in general and the pain system in particular, let us look at the issue of conscious and non-conscious processing in the *emotions*. The only way that one could claim that *emotions are always conscious* would be to limit the definition of 'emotions' to '*emotional experience*.' But even if we equate 'emotion' with 'emotional experience,' it seems artificial to say that I am only 'angry' during the moments when I am focusing my attention on my angry feelings, but not during the times when I 'forget myself' and stare with awe at the rainbow or the times when I am working on something. It is also artificial to limit 'emotion' to 'emotional experience' when emotion has so many different hormonal, autonomic, motoric, facial expression, and cognitive dimensions. Why is the 'experience' emotional when the foot stamping or heart thumping is not?

6.3 *Motivation/emotion as part of core-consciousness?*

If it turns out that destruction of the PAG in humans leads to permanent coma (as Panksepp, in Watt 1998, reports for some other species), and if the PAG is seen as a basic motivational/emotional area, then motivation/emotion will turn out to be part of core consciousness. If the midbrain Ach → thalamic ILN is the essential pathway for the most fundamental type of conscious awareness (Newman 1997 and, in Watt 1998) and if the brainstem acetylcholine system is the least motivational/emotional of the arousal systems (as suggested early in this paper), then basic consciousness might be emotionally neutral. If the *entire* brainstem arousal quintet is seen as essential for consciousness — or even the quartet excluding the PAG — there seem to be very strong motivational/emotional components in basic consciousness.

Looking at the higher end of the emotional control system, the PVM pre-frontal lobe is the one pre-frontal area most clearly linked with emotion, with the anterior cingulate close behind. Eliminate the former and you probably still have a fair amount of conscious experiences — as ventral-prefrontal lobectomy patients can report. The loss of a functioning anterior cingulate may be more questionable regarding consciousness, as with persons with akinetic mutism, and so on. Still, the loss may be a loss in terms of the ability to deliberately focus, rather than total loss of consciousness. What is clear is that the loss of emotional consciousness would greatly alter the quality of consciousness. It might be more akin to losing the television's color and focus, rather than just losing a channel.

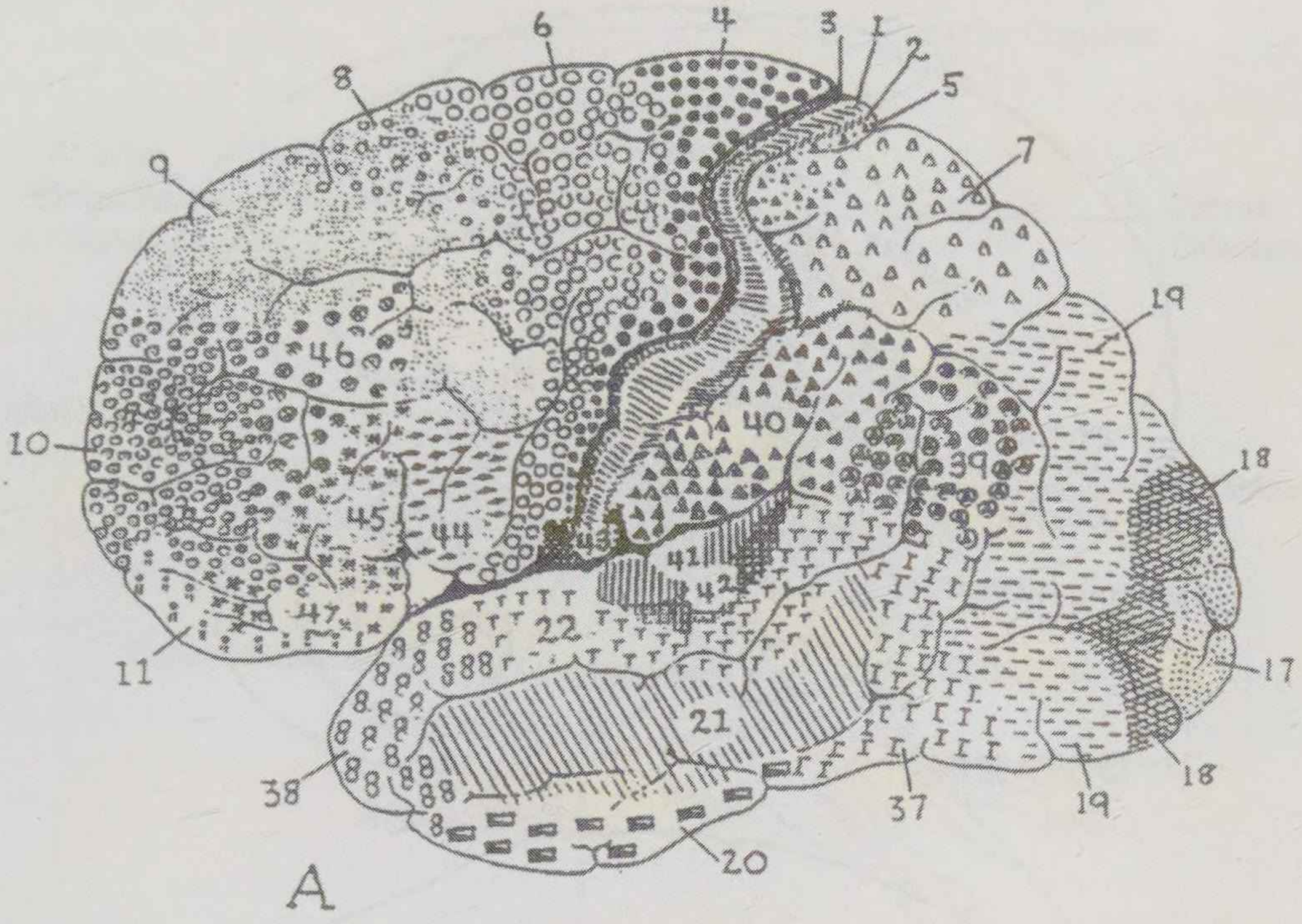
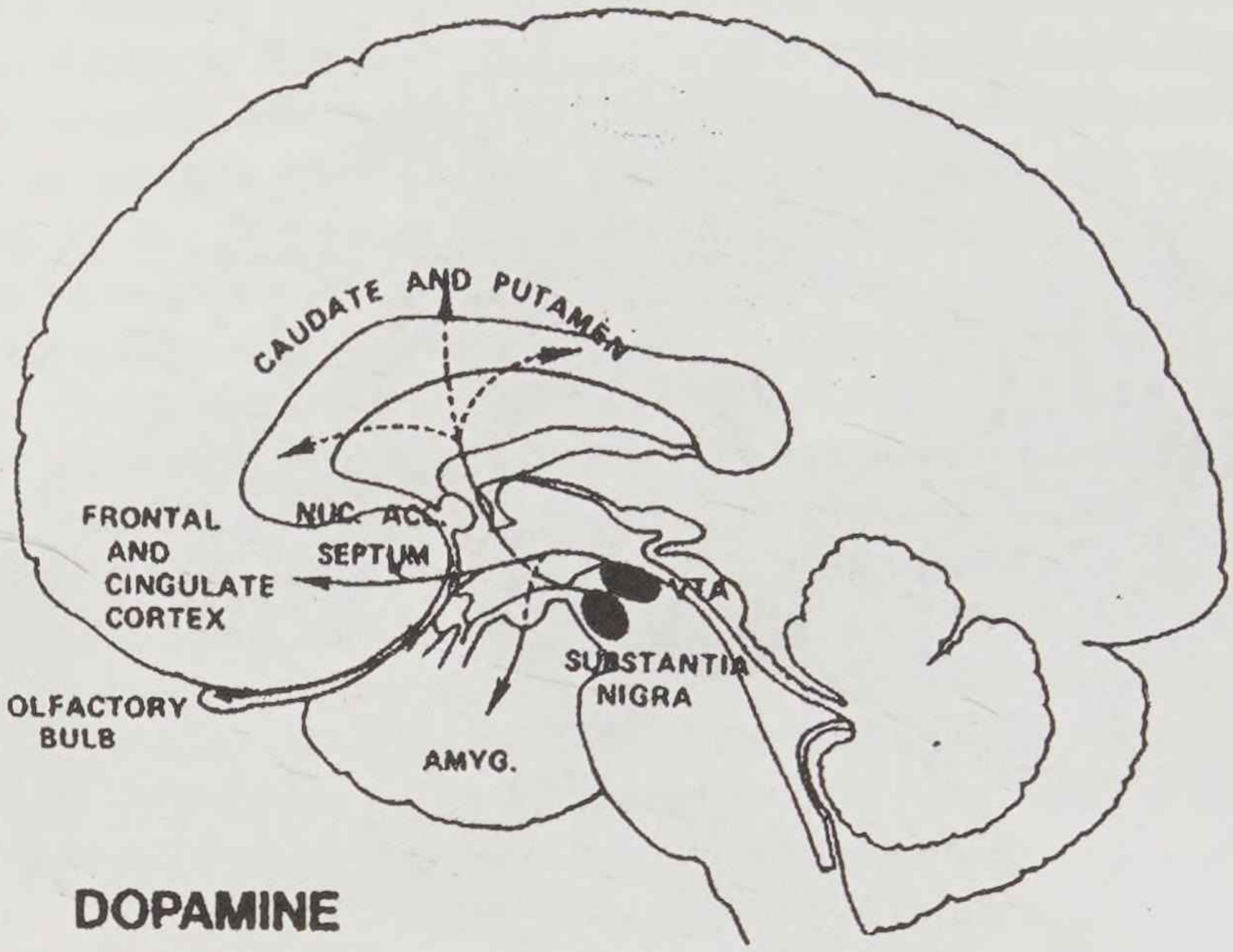
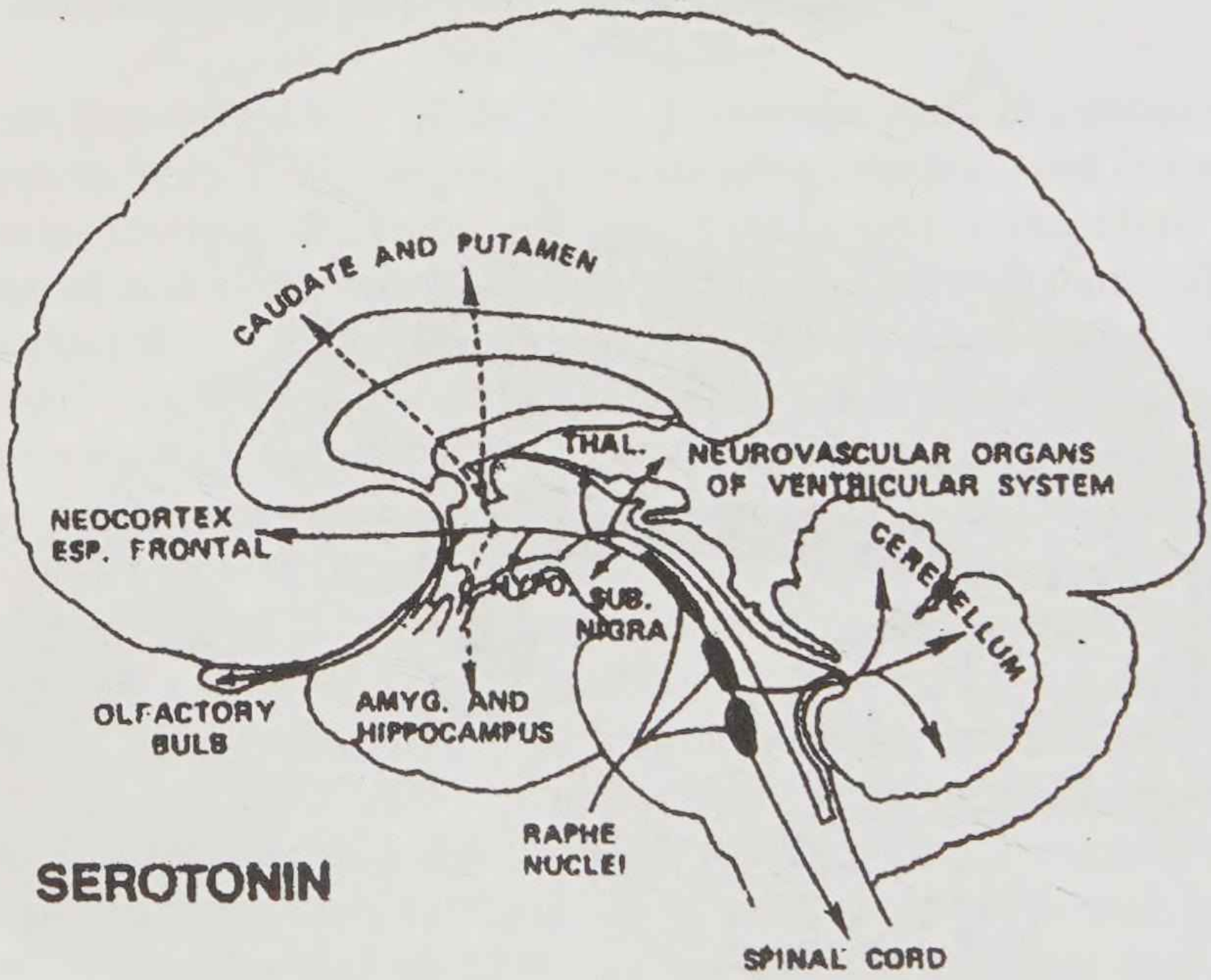


Figure 1. **Human cortex:** Cytoarchitectural map of: (A) Convex surface of cortex; (B) Medial surface of cortex (after Brodmann).



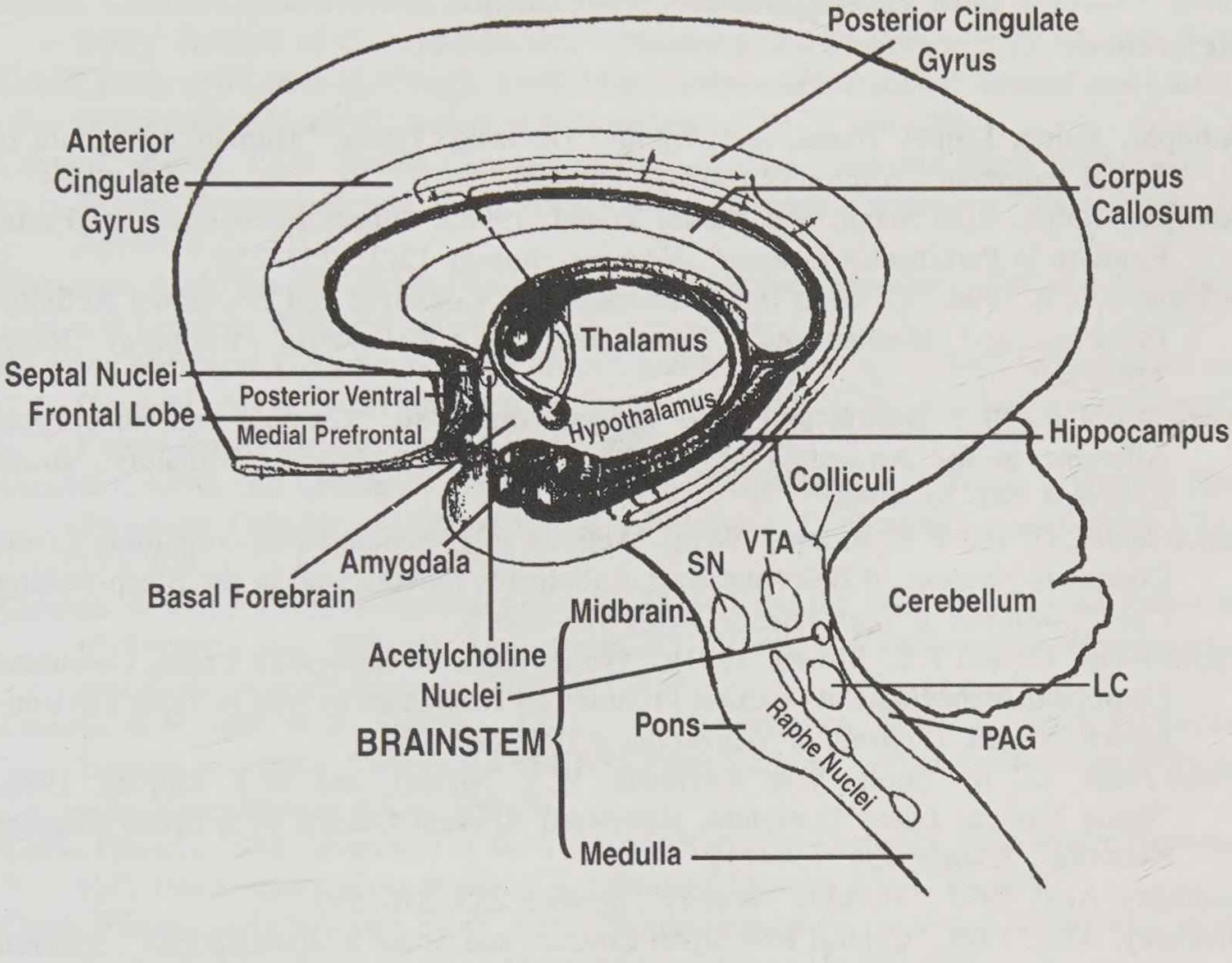
DOPAMINE



SEROTONIN

Figure 2. **Dopamine**: Locations and projections of two of the major dopaminergic systems are illustrated in the upper diagram. **Serotonin:** Serotonin fibers arising from the raphe nuclei of the brain stem are depicted in the botom diagram. AMYG = amygdala; NUC ACC = nucleus accumbens; VTA = ventral tegmental area; HYPO = hypothalamus; SUB NIGRA = substantia nigra; THAL = thalamus.

Facing page: Subcortical brain regions.


Posterior Cingulate Gyrus
Anterior Cingulate Gyrus
Corpus Callosum
Thalamus
Septal Nuclei
Posterior Ventral
Frontal Lobe
Medial Prefrontal
Hypothalamus
Hippocampus
Colliculi
VTA
SN
Amygdala
Basal Forebrain
Midbrain
Cerebellum
Acetylcholine Nuclei
LC
Pons
Raphe Nuclei
PAG
BRAINSTEM
Medulla

References

Adolphs, Ralph, Daniel Tranel, and Antonio Damasio. 1998a. "Human Amygdala in Social Judgment". *Nature 393(6684):* 470–474.

Adolphs, Ralph, Rina Schul, and Daniel Tranel. 1998b. "Intact Recognition of Facial Emotion in Parkinson's Disease". *Neuropsychology* 12(2): 253–258.

Adrianov, O. S. 1996. "Cerebral Interrelationships of Cognitive and Emotional Activity: Pathways and Mechanisms". *Neuroscience and Behavioral Physiology* 26(4): 329–339.

Aggleton, J. P., M. J. Burton and R. E. Passingham. 1980. "Cortical and Subcortical Afferents to the Amygdala of the Rhesus Monkey (Macaca Mulatta)". *Brain Research* 190: 347- 368.

Aston-Jones, G. and F. E. Bloom. 1981a. "Activity of Norepinephrine-containing Locus Coeruleus Neurons in Behaving Rats Anticipates Fluctuations in the Sleep-waking Cycle". *Journal of Neuroscience* 1 (8): 876–886.

Aston-Jones, G. and F. E. Bloom. 1981b. "Norepinephrine-containing Locus Coeruleus Neurons in Behaving Rats Exhibit Pronounced Responses to Non-noxious Environmental Stimuli". *Journal of Neuroscience* 1(8): 887–900.

Aston-Jones, G., M. Ennis, V. A. Pieribone, W. T. Nickell, and M. T. Shipley. 1986. "Brain Nucleus Locus Coeruleus, Restricted Afferent Control of a Broad Efferent Network". *Science* 234: 734–737.

Baddeley, Alan. 1992. "Working Memory". *Science* 255: 556–559.

Baddeley, Alan. 1998. "Central Executive: Concept and Some Misconceptions". *Journal of International Neuropsychological Society* 4(5): 523–526.

Baylis, G. C., E. T. Rolls, and C. M. Leonard. 1985. "Selectivity Between Faces in the Responses of a Population of Neurons in the Cortex in the Superior Temporal Sulcus of the Monkey". *Brain Research* 342: 91–102.

Baylis, G. C., E. T. Rolls, and C. M. Leonard. 1987. "Functional Subdivisions of the Temporal Lobe Neocortex". *Journal of Neuroscience* 7(2): 330–342.

Bakish, David, Patrizia Cavazzoni, Jerzy Chudzik, and Arun Ravindran. 1997. "Effects of Selective Serotonin Reuptake Inhibitors on Platelet Serotonin Parameters in Major Depressive Disorder. *Biological Psychiatry* 41(2): 184–190.

Bechara, Antoine, Hanna Damasio, Daniel Tranel, and Steven Anderson. 1998. "Dissociation of Working Memory from Decision Making within the Human Prefrontal Cortex". *Journal of Neuroscience* 18(1): 428–437.

Blier, Pierre, Richard Bergeron, and Claude de Montigny. 1997. "Selective Activation of Postsynaptic 5-HT-1a Receptors Induces Rapid Antidepressant Response". *Neuropsychopharmacology* 16(5): 333–338.

Block, Ned. 1995. "On a Confusion about a Function of Consciousness". *Behavioral and Brain Sciences* 18: 2.

Broks, Paul, Andrew Young, Elizabeth Maratos, and Peter Coffey. "Face Processing Impairments after Encephalitis: Amygdala Damage and Recognition of Fear". *Neuropsychologia* 36(1): 59–70.

Busch, Chardelle and Herbert Alpern. 1998. "Depression after Mild Traumatic Brain Injury; Review of Current Research". *Neuropsychology Review* 8(2): 95–108.

Cahill, Larry and James McGaugh. 1998. "Mechanisms of Emotional Arousal and Lasting Declarative Memory". *Trends in Neurosciences* 21(7): 294–299.

Cannon, Walter. 1929. *Bodily Changes in Pail, Hunger, Fear and Rage. An Account of Recent Researches into the Function of Emotional Excitement.* New York: D. Appleton.

Carpenter, Malcolm and Jerome Sutin. 1983. *Human Neuroanatomy, (8th ed.).* Baltimore and London: Williams and Wilkins.

Cherrier, Monique, Gary Small, Scott Komo, and Asenath LaRue. 1997. "Mood State and Cerebral Metabolism in Persons with Age-associated Memory Impairment". *Psychiatry Research: Neuroimaging* 76(2,3): 67–74.

Cicerone, Keith and Lawrence Tanenbaum. 1997. "Disturbance of Social Cognition after Traumatic Orbitofrontal Brain Injury". *Archives of Clinical Neuropsychology* 12(2): 173–188.

Coenen, A. M. L. 1995. "Neuronal Activities Underlying the EEG and Evoked Potentials of Sleeping and Waking: Implications for Information Processing". *Neuroscience and Biobehavioral Reviews* 19(3): 447–463.

Correll, R. E. and W. B. Scoville. 1965. "Performance on Delayed Match Following Lesions of Medial Temporal Lobe Structures". *Journal of Comparative and Physiological psychology* 60(3): 360–367.

Crick, Francis. 1984. "Function of the Thalamic Reticular Complex: Searchlight Hypothesis". *Proceedings of the National Academy of Sciences* 81: 4586–4590.

Crick, Francis and Christoph Koch. 1990. "Some Reflections on Visual Awareness". *Cold Spring Harbor Symposium on Quantitative Biology* 15: 953–962.

Csernansky, John and Mark Bardgett. 1998. "Limbic-cortical Neuronal Damage and the Pathophysiology of Schizophrenia". *Schizophrenia Bulletin* 24(2): 231–248.

Desimone, Robert, Thomas Albright, Charles Gross, and Charles Bruce. 1984. "Stimulus-selective Properties of Inferior Temporal Neurons in the Macaque". *Journal of Neuroscience*, 4(8): 2051–2062.

Elliott, Rebecca, B. J. Sahakian, A. Michael, and E. S. Paykel. 1998. "Abnormal Neural Response to Feedback on Planning and Guessing Tasks in Patients with Unipolar Depression". *Psychological Medicine* 28(3): 559–571.

Everitt, Barry and Trevor Robbins. 1997. "Central Cholinergic Systems and Cognition". *Annual Review of Psychology* 48: 649–684.

Faw, Bill. 1987. *Knowing One's Own Face: Neuro-Cognitive Analysis.* Unpublished Master's Thesis. Richmond, VA: Virginia Commonwealth University.

Faw, Bill. 1997. "Outlining a Brain Model of Mental Imaging Abilities". *Neuroscience and Biobehavioral Reviews* 21(3): 283–288.

Faw, Bill. 1999. "Executive Control of Consciousness by the Frontal Lobe". Presented at *Association for the Scientific Study of Consciousness.* London, Ontario.

Faw, Bill. In press. "My Amygdala-Orbital-Frontal Circuit Made Me Do It". *Consciousness & Emotion.*

Fernandez-Duque, Diego and Michael Posner. 1997. "Relating the Mechanisms of Orienting and Alerting". *Neuropsychologia* 35(4): 477–487.

Ferreira, C., M. Verin, B. Pillon, and R. Levy. 1998. "Spatio-temporal Working Memory and Frontal Lesions in Man". *Cortex* 34(1): 83–98.

Flanagan, Owen. 1997. "Prospects for a Unified Theory of Consciousness or, What Dreams Are Made for". In J. D. Cohen and J. W. Schooler (eds), *Scientific Approaches to Consciousness*. Mahwah, NJ: LEA.

Gallup, Gordon. 1968. "Mirror-image Stimulation". *Psychological Bulletin* 70(6): 782–793.

Gallup, Gordon. 1999. "The Mirror Test". Paper at *3rd Annual Meeting Association for the Scientific Study of Consciousness.* London, Ontario.

Ghika-Schmied, F., G. Assal, J. Ghika, and J. Bogousslavsky. 1998. "Moriatic Aphasia: Syndrome of Acute Carotid Occlusion on the Dominant Side". *Neurology* 50(5): 1511- 1512.

Gray, Jeffrey. 1998. "Integrating Schizophrenia". *Schizophrenia Bulletin* 24(2): 249–266.

Gur, Ruben, J. Daniel Ragland, Lyn Mozley, and P. David Mozley. 1997. "Lateralized Changes in Regional Cerebral Blood Flow During Performance of Verbal and Facial Recognition Tasks: Correlations with Performance and 'effort'". *Brain and Cognition* 33(3): 388- 414.

Hedges, Dawson, Frederick Reimherr, Robert Strong, and Huston Corinne. 1996. "Open Trial of Nefazodone in Adult Patients with Generalized Anxiety Disorder". *Psychopharmacology Bulletin* 32(4): 671–676.

Heresco-Levy, Uriel, Gail Silipo, and Daniel C. Javitt. 1996. "Glycinergic Augmentation of NMDA Receptor-mediated Neurotransmission in the Treatment of Schizophrenia". *Psychopharmacology Bulletin* 32(4): 731–740.

Herzog, Andrew and Gary van Hoesen. 1976. "Temporal Neocortical Afferent Connections to the Amygdala in the Rhesus Monkey". *Brain Research* 115: 57–69.

Hobson, J. A. 1997. "Consciousness as a State-dependent Phenomenon". In J. D. Cohen and J. W. Schooler (eds), *Scientific Approaches to Consciousness*. Mahwah, NJ: LEA.

James, William. 1884. "What Is an Emotion?" *Mind* 9: 188–205.

Jones, B. and Mortimer Mishkin. 1972. "Limbic Lesions and the Problem of Stimulus-Reinforcement Associations. *Experimental Neurology* 38: 362–377.

Kanwisher, Nancy. 1999. "Neural Correlates of Changes in Perceptual Awareness in the Absence of Changes in the Stimulus". Paper at *3rd Annual Meeting Association for the Scientific Study of Consciousness.* London, Ontario.

Kluver, H. and P. Bucy. 1939. "Preliminary Analysis of Functions of the Temporal Lobes in Monkeys". *Archives of Neurological Psychiatry* 42: 979–1000.

Kouri, Elena and Uriel Halbreich. 1997. "State and Train Serotonergic Abnormalities in Women with Dysphoric Premenstrual Syndromes". *Psychopharmacology Bulletin* 33(4): 767- 770.

Laird, Lyle. 1996. "Issues in Monopharmacotherapy and Polypharmacotherapy of Obsessive-Compulsive Disorder". *Psychopharmacology Bulletin* 32(4): 569–578.

Lange, Carl. 1885/1922. In James and Lange: *The Emotions.* Baltimore.

Leo, Raphael. 1996. "Movement Disorders Associated with the Serotonin Selective Reuptake Inhibitors". *Journal of Clinical Psychiatry* 57(10): 449–454.

Leonard, C.M., E.T. Rolls, F.A.W. Wilson, and G.C. Baylis. 1985. "Neurons in the Amygdala of the Monkey with Responses Selective for Faces". *Experimental Brain Research* 15: 159–176.

Leonard, Brian. 1997. *Fundamentals of Psychopharmacology (2nd ed.)*. Chichester: John Wiley and Sons.

Lerman, Caryn, Neil Caporaso, David Main, and Janet Audrain. 1998. "Depression and Self-medication with Nicotine: Modifying Influence of the Dopamine D4 Receptor Gene". *Health Psychology* 17(1): 56–62.

Llinas, R., U. Ribary, M. Joliot, and X-J. Wang. 1994. "Content and Context in Temporal Thalamocortical Binding". In *Temporal Coding of the Brain*. G. Busaki (ed), Heidelberg: Springer-Verlag.

Llinas, R. and D. 1996). "The Brain as a Closed System Modulated by the Senses". In *The Brain-Mind Continuum,* R. Llinas and P.S. Churchland (eds). Cambridge, MA: MIT Press.

Low, Bee Lee and Siow Ann Chong. 1998. "Treating Poststroke Pathologic Crying with Fluoxetine". *Journal of Clinical Psychopharmacology* 18(3): 252.

Lycan, William. 1996. *Consciousness and Experience*. Cambridge, MA: MIT.

Malison, Robert, Susan Best, Christopher van Dyck, and Elinore McCance. 1998. "Elevated Striatal Dopamine Transporters during Acute Cocaine Abstinence as Measured by Beta-CIT SPECT". *American Journal of Psychiatry* 155(6): 832–834.

Marshall, John, Peter Halligan, Gereon Fink, and Derick Wade. 1997. "Functional Anatomy of a Hysterical Oaralysis". *Cognition* 64(1): B1-B8.

McAllister-Williams, R.H., I.N. Ferrier, and A.H. Young. 1998. "Mood and Neuropsychological Function in Depression: Role of Corticosteroids and Serotonin". *Psychological Medicine* 28(3): 573–584.

Mishkin, Mortimer and Jean Delacour. 1975. "Analysis of Short-term Visual Memory in the Monkey". *Journal of Experimental Psychology: Animal Behavior Processes* 1(4): 326- 334.

Miyai, Ichiro and Michael Reding. 1998. "Effects of Antidepressants on Functional Recovery Following Stroke: Double-blind Study". *Journal of Neurologic Rehabilitation* 12(1): 5- 13.

Mora, F., E.T. Rolls, and M.J. Burton. 1976. "Modulation during Learning of the Responses of Neurons in the Lateral Hypothalamus to the Sight of Food". *Experimental Neurology* 53: 508–519.

Morris, J.S., C.D. Frith, D.I. Perrett, and D. Rowland. 1996. "Differential Neural Response in the Human Amygdala to Fearful and Happy Facial Expressions". *Nature* 383(6603): 812–815.

Morris, J.S., A. Ohman, and R.J. Dolan. 1998. "Conscious and Unconscious Emotional Learning in the Human Amygdala". *Nature* 393(6684): 467–470.

Moruzzi, G. and H.W. Magoun. 1949. "Brain Stem Reticular Formation and Activation of the EEG". *Electroencephalographic Clinical Neurophysiology* 1: 455–473.

Mountcastle, Vernon. 1978. "Organizing Principle for Cerebral Function: Unit Module and the Distributed System:. In G.M. Edelman and V.B. Mountcastle (eds), *The Mindful Brain (7–50).* Cambridge, MA: M.I.T.

Nadel, L. and M. Moscovitch. 1998. "Hippocampal Contributions to Cortical Plasticity". *Neuropharmacology* 37(4–5): 431–439.

Natsoulas, Thomas. 1978. "Consciousness". *American Psychologist* 33: 269–283.

Neumeister, Alexander, Erick Turner, Jeffrey Matthews and Teodor Postolache. 1998. "Effects of Tryptophan Depletion vs Catecholamine Depletion in Patients with Seasonal Affective Disorder in Remission with Light Therapy". *Archives of General Psychiatry* 55(6): 524–530.

Newman, James. 1997. "Putting the Puzzle Together. Part I: Towards a General Theory of the Neural Correlates of Consciousness". *Journal of Consciousness Studies* 4: 47–66.

Nobili, Lino and Walter Sannita. 1997. "Cholinergic Modulation, Visual Function and Alzheimer's Dementia". *Vision Research* 37(24): 3559–3571.

O'Donnell, Patricio and Anthony Grace. 1998. "Dysfunctions in Multiple Interrelated Systems as the Neurobiological Bases of Schizophrenic Symptom Clusters". *Schizophrenia Bulletin* 24(2): 267–283.

Olds, James and Peter Milner. 1954. "Positive Reinforcement Produced by Electrical Stimulation of Septal Area and Other Regions of the Rat Brain". *Journal of Comparative Physiological Psychology* 47: 419–427.

Parkin, Alan. 1998. "Central Executive Does Not Exist". *Journal of the International Neuropsychological Society* 4(5): 518–522.

Perrett, David, E. T. Rolls, and W. Caan. 1982. "Visual Neurones Responsive to Faces in the Monkey Temporal Cortex". *Experimental Brain Research* 47: 329–342.

Perrett, David. 1999. "Neurons that Perceive Faces: Face Perception and Responses of Brain Cells". Paper at *3rd Annual Meeting Association for the Scientific Study of Consciousness.* London, Ontario.

Persico, Antonio, Steven Riech, Jack Henningfield, and Michael Kuhar. 1998. "Parkinsonian Patients Report Blunted Subjective Effects of Methylphenidate". *Experimental and Clinical Psychopharmacology* 6(1): 54–63.

Peyron, R., L. Garcia-Larrea, M. C. Gregoire, and P. Convers. 1998. Allodynia after Lateral-medullary (Wallenberg) Infarct: PET Study". *Brain* 121(2): 345–356.

Pinel, J. P. J. 1997. *Biopsychology* (3rd ed.). Boston: Allyn and Bacon.

Pivac, Nela, Miro Jakovljevic, Dorotea Muck=Seler, and Zdravko Brzovic. 1997. "Hypothalamic-pituitary-adrenal Axis Function and Platelet Serotonin Concentrations in Depressed Patients. *Psychiatry Research* 73(3), 123–132.

Posner, Michael and M. K. Rothbart. 1992. "Attentional Mechanisms and Conscious Experience. In A. D. Milner and M. D. Rugg (eds), *Neuropsychology of Consciousness.* London: Academic Press.

Reiman, Eric. 1997. "Application of Positron Emission Tomography to the Study of Normal and Pathologic Emotions". *Journal of Clinical Psychiatry* 58(Suppl 16): 4–12.

Robbins, Trevor. 1997. "Arousal Systems and Attentional Processes. *Biological Psychology* 45(1–3): 57–71.

Rolls, E. T., S. J. Judge, and Manjit Sanghera. 1977. "Activity of Neurones in the Inferotemporal Cortex of the Alert Monkey". *Brain Research* 130: 229–238.

Salamone, J. D., M. S. Cousins, and B. J. Snyder. 1997. "Behavioral Functions of Nucleus Accumbens Dopamine: Empirical and Conceptual Problems with the Anhedonia Hypothesis". *Neuroscience and Biobehavioral Reviews* 21(3): 341–359.

Sander, Thomas, Helmut Harms, Peter Dufeu, and Silke Kuhn. 1998. "Serotonin Transporter Gene Variants in Alcohol-dependent Subjects with Dissocial Personality Disorder". *Biological Psychiatry* 43(12): 908–912.

Sara, S. J. 1985. "Locus Coeruleus and Cognitive Function, Attempts to Relate Noradrenergic Enhancement of Signal/noise in the Brain to Behavior". *Physiological Psychology* 13(3): 151–162.

Schiff, Nicholas. 1999. "Neurological Disorders of Consciousness". Workshop at *3rd Annual Meeting Association for the Scientific Study of Consciousness.* London, Ontario.

Segal, M. 1985. "Mechanisms of Action of Noradrenaline in the Brain". *Physiological Psychology* 13(3): 172–178.

Silver, Henry and Natalie Shmugliakov. 1998. "Augmentation with Fluvoxamine but not Maprotiline Improves Negative Symptoms in Treated Schizophrenia: Evidence for a Specific Serotonergic Effect from a Double-blind Study". *Journal of Clinical Psychopharmacology* 18(3): 208–211.

Soloff, Paul. 1998. "Algorithms for Pharmacological Treatment of Personality Dimensions: Symptom-specific Treatments for Cognitive-perceptual, Affective, and Impulsive=behavioral Dysregulation". *Bulletin of the Menninger Clinic* 62(2): 195–214.

Sonino, Nicoletta and Giovanni Fava. 1996. "Serotonin, Cushing's Disease, and Depression".*Psychotherapy and Psychosomatics* 65(2): 63–65.

Spiegler, Brenda and Mortimer Mishkin. 1981. "Evidence for the Sequential Participation of Inferior Temporal Cortex and Amygdala in the Acquisition of Stimulus-reward Associations". *Behavioural Brain Research* 3: 303–317.

Steriade, M. and R. W. McCarley. 1990. *Brainstem Control of Wakefulness and Sleep.* New York: Plenum.

Thompson, Richard. 1993. *The Brain: Neuroscience Primer* (2nd ed.). NY: Freeman.

Tiitinen, Hannu. 1995. "Selective Attention, 40-Hz Oscillations and Mismatch Negativity (Finn).*Psykologia* 30(4), 284–287.

Turner, Blair, Mortimer Mishkin, and Margaret Knapp. 1980. "Organization of the Amygdalopetal Projections from Modality-specific Cortical Association Areas in the Monkey". *Journal of Comparative Neurology* 191: 515–543.

Ursin, H. and B. R. Kaada. 1960. "Functional Localization within the Amygdaloid Complex in the Cat". *Electroencephalographic Clinical Neurophysiology* 12: 1–20.

van Hoesen, Gary and Deepak Pandya. 1975. "Some Connections of the Entorhinal (Area 28) and Perirhinal (Area 35) Cortices of the Rhesus Monkey. I. Temporal Lobe Afferents". *Brain Research* 95: 1–24.

Viinamaki, Heimo, Jyrki Kuikka, Jari Tiihonen, and Johannes Lehtonen. 1998. "Change in Monoamine Transporter Density Related to Clinical Recovery: Case-control Study". *Nordic Journal of Psychiatry* 52(1): 39–44.

Vogeley, Karl 1999. Association for the Scientific Study of Consciousness e-seminar on Consciousness and Self.

Waldinger, Marcel, Hemmie Berendsen, F. M. Bertil, and Berend Olivier. 1998. "Premature Ejaculation and Serotonergic Antidepressants-induced Delayed Ejaculation: Involvement of the Serotonergic System". *Behavioural Brain Research*, 92(2): 111–118.

Watt, Douglas. 1998. Association for the Scientific Study of Consciousness e-seminar on Emotion and Consciousness

Weiskrantz, Lawrence and Elizabeth Warrington. 1979. "Conditioning in amnesic patients". *Neuropsychologia* 17: 187–194.

Wilson, C. L., T. L. Babb, E. Halgren, and P. H. Crandall. 1983. "Visual Receptive Fields and Response Properties of Neurons in Human Temporal Lobe and Visual Pathways". *Brain* 106: 473–502.

Wilson, Kevin. 1998. "Issues Surrounding the Cognitive Neuroscience of Obsessive-compulsive Disorder". *Psychonomic Bulletin and Review* 5(2): 161–172.

Wise, R. A. 1996. "Addictive Drugs and Brain Stimulation Reward". *Annual Review of Neuroscience* 19: 319–340.

Woolf, Nancy. 1996. "Critical Role of Cholinergic Basal Forebrain Neurons in Morphological Change and Memory Encoding: A Hypothesis". *Neurobiology of Learning and Memory* 66(3): 258–266.

Yamaguchi, Shuhei and Shotai Kobayashi. 1998. "Contributions of the Dopaminergic System to Voluntary and Automatic Orienting of Visuospatial Attention". *Journal of Neuroscience* 18(5): 1869–1878.

CHAPTER 4

Conscious Emotion in a Dynamic System

How I Can Know How I Feel

Natika Newton
Nassau Community College

A dynamic model of brain mechanisms of consciousness and emotion offers more comprehensive and coherent solutions than the traditional Cartesian model to many traditional puzzles in philosophy of mind. One of these is self-awareness: how is it possible for a conscious being to be reflexively aware of its own consciousness? In this chapter I discuss specific ways this question can be treated using a dynamic model. The discussion has two parts. First, I propose, in general terms, a way in which familiar aspects of conscious emotional states can be viewed as elements of a dynamical system. Second, I show how, on this model, one can be consciously aware of one's own emotions.

1. Self-organizing systems

Ellis (this volume) has given us a broad picture of how dynamical systems theory can account for both the objective and the subjective aspects of conscious mental states without a need for positing nonphysical processes. Such an *enactive* theory views consciousness in terms of natural self-organizing properties of complex systems. The basic idea is that all life processes incorporate a natural tendency toward order, which arises spontaneously among molecules when in sufficiently complex groups, in a way that can be explained entirely by physical mechanisms and involves no mystery. It does, however, allow the emergent order to be conceptually distinguished from the substratum in ways that appeal to some theorists. Mental states are not reducible to the individual states of the substratum, but they are physical states nonetheless, and obey physical laws. On this view, consciousness is a type of organismic activity. The standard *reactive* approach, in contrast, presents consciousness in general, and conscious emotion

in particular, as subjective states of awareness produced in a passive subject by internal or external stimuli. The analogy on which this reactive model is based is a perceptual one, with perception understood in the most naive terms as a perceiver being acted on by perceptual stimuli, the way a mirror passively reflects the objects.

An enactive model is more consistent than the reactive, perceptual model with the way we view other aspects of a living organism. We think of living systems in terms of goal-oriented processes, from the activity of chromosomes all the way to the major metabolic systems and to behavior-controlling systems in the brain. The perceptual model has taken consciousness out of this context, and made it into a detached observer of the organismic activities that feed 'experiences' into consciousness; one result is a difficulty of understanding the purpose of consciousness. On the enactive approach the purpose of consciousness is clear: to organize the component elements of the conscious organism into patterns that advance the overall goals of the organism.

This answer is a useless tautology, however, if we cannot specify how consciousness does this, and how the structures of self-organization can be applied at this high, almost abstract level. Sections 1 and 2 of this chapter will briefly outline the basic elements in self-organizing systems. Then the remainder of the chapter will show how aspects of consciousness and emotion might be mapped onto these.

In the most general terms, a self-organizing system is a system of individual entities that interact in such a way as to maintain the existence of the system as a whole over time. A system is a group of entities with some collective property, e.g. a type of interaction of which members are capable, such that there is a well-defined surface separating the system from everything else. Maintaining the system is thus maintaining the collective property.

The property of self-organization with which we are concerned applies only to open thermodynamic systems, which exchange matter and energy with the environment. A thermodynamic system is one in which interactions of the entities involve the production and consumption of heat energy. In closed systems, isolated from the environment, the group eventually succombs to entropy: the interactions are random and the group is in a state of stable equilibrium — a state that does not change with time. In open systems, by contrast, there are three general possibilities with regard to order. First, the system could become chaotic: the interactions within the system could occur with with increasing randomness. The system would eventually reach total entropy, or equilibrium with the environment, and disintegrate because of environmental invasions (such as when a leaf decomposes in a compost heap). When an organism is at equilibrium with the surroundings, it is dead (Rawn 1983: 24). The second possibility is that the system could be frozen in a single state in which

interactions cease altogether (a diamond is such a system). Third, the system could reach a sustained, nonequilibrium state of homeostasis, in which stability of the system as a whole is maintained by means of continual adjustments to environmental perturbations. (Homeostasis is sometimes referred to as 'dynamic equilibrium' (Fruton and Simmonds 1960: 730), but that term can be misleading and I shall not use it.) All biological systems are homeostatic; absence of homeostasis, or equilibrium with the environment, is death.

With homeostasis, a system maintains itself in one or more ordered states consisting of repeating patterns of interactions, called 'steady states.' The pattern could take the form of a single configuration, enduring until some outside force caused the pattern to be destroyed; or the entities in the group could fall into a pattern of distinct state-cycles, in which a series of different states is repeated. In a homeostatic system, the group will eventually return to the pattern even when irregular behavior of individual members, or an outside disturbance, perturbs the system as a whole. To the degree that the group tends to return to the pattern after disruptions, it is resistant to disintegrating forces. No living system is completely immune to such forces, or it would be immortal.

Homeostasis is not a new concept. What is new is the exploration of the ways homeostasis can emerge spontaneously in nonliving collections of molecules. Self-organization can be demonstrated computationally to occur spontaneously as a result of variously naturally occurring parameters affecting the interactions of the entities in the groups. According to Kauffman (1993), the relevant parameters are the number of entities, the number of potential interactions among them, and the Boolean structure of their interactions. Specifically, given a sufficiently large network of entities capable of diverse interactions, a self-sustaining pattern of interactions is almost certain to arise.

2. Life and self-organization

Kauffman argues that such naturally arising order must precede the evolution of life forms, in order for there to be systems stable enough, and flexible enough, to evolve. But how could an initially disordered system order itself? He argues that the sort of interaction at the root of emergent life processes is *catalysis*: a molecule acts as a catalyst if it speeds up (or makes more probable) a reaction among other molecules that would otherwise occur much more slowly. A self-organizing system that can catalyze the reactions that maintain its own existence, or reproduce its own states, Kauffman calls a 'collectively autocatalytic system.' A living organism is such a system.

While the laws governing the behavior of living systems even as simple as single cells are not known, Kauffman shows how autocatalysis is biochemically

possible. Protein enzymes and ribozymes are examples of catalysts in living cells. An enzyme catalyzes reactions, such as the conversion of a molecule A into a molecule B, by reducing the activation energy of the transition state between the two forms and making it more likely to occur. This makes it easier for As to convert to Bs (Darnell *et al.* 1986: 48). In principle, autocatalysis would occur in a system with a critical diversity of molecules, such that the probability of molecules capable of catalyzing all the kinds of molecules in the system, including the catalysts themselves, is sufficiently high. Computer models employing a variety of different catalyst rules show that at certain points of diversity, collectively autocatalytic sets 'emerge.'

In order to evolve, an autocatalytic system must not be so rigidly ordered that it remains in a single state or cycle of states regardless of any perturbations; it also must resist becoming chaotic with small perturbations. The ideal system is orderly yet flexible; Kauffman says that it is 'poised between order and chaos.' Internal or external disturbances can cause the system to change from one cycle of states to another, but will not produce complete chaos.

The concept that we need in order to understand that type of order is that of an 'attractor.' An attractor is a pattern of behavior into which the system has a tendency to settle. There are two parameters that determine the collective effect of attractors upon a system. The first is number. The fewer the attractors, the higher the degree of order: a system with only one attractor will be frozen into that pattern of behavior; a system with vast numbers of attractors will be chaotic, slipping from one state into another at the slightest perturbation. Second, there is what we might call the range of the attractor. A system will remain in a wide-ranged attractor despite slight perturbations, which would knock the system out of an attractor with a narrower range. A useful model of an attractor is that of a basin on a plane, draining into a hole, or state cycle. The width of the basin determines how easily the attractor can capture and retain the system. If the basin is wide, then it includes many states of the system, and thus small perturbations will tend not to knock the system out of the state cycle. A homeostatic system has wide-ranged attractors. The size of the attractors also has an effect that is crucial for human recognition of patterns. Size is determined by the number of states that make up the state cycles, or pattern of behavior. A very large attractor could comprise so many states that one cycle of the pattern could not be repeated in a human lifetime. Small attractors are therefore necessary if the ordered system is to survive and interact with other living things.

3. Consciousness as a self-organizing system

To apply the self-organization model to consciousness we can consider a

particular mapping of the components of consciousness onto such a model, using the elements outlined above (e.g. self-organizing system, attractors) as a basis for organization. How do we determine the elements of consciousness? I begin with the assumption, supported by the arguments of the previous chapters in this volume, that consciousness viewed as a kind of activity must involve emotion. I will use our introductory characterizations of both conscious and unconscious 'motivation' and 'emotion.' A motivation must, it seems, be associated with a conscious emotion in order to be the direct cause of *voluntary* activity. Only if I am aware that I want to go to the post office will I gather my letters and get into the car, with the conscious intention of going to the post office. Goal-directed activity in sentient beings is driven in a general sense by hope for reward. The 'reward' may be a particular consummatory pleasure, in very simple cases; more commonly it will be a positive value associated with a certain type of action, e.g. exploratory activity (Panksepp, this volume) or social interaction without any particular concrete goal. Whatever the level of generality of the 'reward,' if the activity is voluntary and the being is conscious, then it is represented in emotion-laden sensorimotor imagery (Damasio 1994; Platt and Glimcher 1999).

For present purposes we can think of conscious emotion as a 'feeling' in Damasio's (1994) sense. The following mapping is only one possible one among many; its purpose is primarily to exemplify the way an enactive model can be applied to conscious emotional states. In this mapping, the order of the elements corresponds to the degree to which each occurs as an introspectible component of conscious experience. Individual neurons or neuron-groups, e.g., are not intentional objects within experience; we are aware only of the states of afffairs they represent. Similarly, activation and inhibition manifest themselves as vividness or salience properties of these states of affairs, not as neuronal properties. State cycles are experienced as sensorimotor images. Attractors, however, are directly represented as valenced goal-states, and catalysts as feelings, prompting and enabling us to act to bring about certain of these goal-states. The necessity of including both conscious intentional objects and unconscious substratum systems in the same mapping reflects the interdependency of conscious and unconscious properties within the organism.

The first two correspondences are the clearest. A phenomenally conscious brain is organized for certain specific kinds of experiences. The data for these experiences are the results of the processing of components in the form of lower-level brain structures such as sensory systems, motor control systems, and bodily representations. Since some information-processing can be assumed to occur unconsciously much of the time within these systems, the role of consciousness would be to coordinate their interactions through selective activation and inhibition of intergroup signalling.

Table 1. *Mapping of enactive model to emotional states*

Self-organizing system	**Phenomenally conscious brain**
components of system	neurons, neuron groups, information-processing modules
interactions	activation, inhibition
state cycles	working memory/extero- and intero-sensory loops (imagery states)
attractors	bodily states vis-à-vis the environment
catalysts	feelings (conscious emotions)

The correspondences to state cycles and attractors should also be clear. The goal of consciousness is to achieve and maintain certain bodily states, given specific environmental conditions, in order to promote the general interests of the organism. An attractor for a conscious organism might be, for example, a state of calm self-confidence in the presence of a potentially threatening social group. The attractor consists of a positively-valenced pattern of state-cycles (repeated series of different organismic states). In this case, this might be a series of states producing alertness to activities of the group, evaluation of the input, and appropriate behavioral responses, all combined with continual monitoring and adjustment of various bodily systems to maintain a type of overall comfort appropriate to the context.

Interpreting the state cycles as images is a proposal based on recent work on action-planning in the brain (Jeannerod 1994; Damasio 1994). In voluntary activity there is a representation, or motor image, of the desired goal-state, and this image plays a role in initiating and controling the action, and in determining ultimate success or failure (Blakemore *et al.* 1998; Georgieff and Jeannerod 1999). Note that 'image' here does not mean 'visual image.' A motor image is a representation of what it would *feel like* to perform a certain action.

The most important contribution of the self-organization model lies, I believe, in the notion of catalysis. In the case of simple self-organizing systems, a catalyst can be a single molecule that enhances interactions among other molecules (Kauffman 1993). While a conscious emotion cannot be reduced to a single molecule or even to a type of chemical such as serotonin, nevertheless it can play a catalyzing role by utilizing such chemicals to coordinate and focus the activities of various brain systems. It is here that the familiar phenomenal characteristics of feelings have their role. An emotion is a pattern of bodily states

comprising the various motivations or drives of the organism's component systems, organized into a coherent goal-directed state that includes a potentially-conscious representation of the goal. While subsystems of a conscious organism have distinct motivations, only the whole organism has emotions. The emotion understood as a *feeling* acts as a catalyst with regard to the motivated subsystems by enhancing the valences of represented actions and goals. This enhancement speeds up the competition among the various motivated subsystems, with the result that they become rapidly organized and directed toward a single goal.

It might seem that if a feeling plays the role of a catalyst in the sense defined above (an element that speeds up or makes more probable a reaction that would otherwise occur much more slowly), then the role of consciousness is significantly diminished. If the reaction would tend to occur anyway, albeit more slowly, why is consciousness important?

Faw (this volume) states that some motivational systems "*utilize the production of conscious feelings*–such as hunger, thirst, feeling cold or hot, and sexual arousal — to aid in the *production of more precise motivated responses*." Consider the case of hunger. If a motivation to find food is unconscious, it may still prevail in competition with other motived activities, such as to find warmth or sexual satisfaction, but its probability of winning will be decreased without the enhancement it obtains from the organism's consciousness of hunger. Conscious feelings, we might say, are like dirty tricks played by motivational systems on each other. The system that gets control of consciousness need not fight it out alone with its rivals at the street level; it has attracted the support of the Big Man, who sends reinforcements, intensifying the power of the fortunate system. The response of the organism as a whole will be more precise in that it is more focused on a single goal, using resources more efficiently than if each independent system has to compete using only its own resources. And in an environment in which quick decision-making enhances survival, consciousness will play a vital role.

On any level below the conscious one, the role of catalyst is played by physical substances such as neurotransmitters and hormones. On the conscious level, emotion of course makes use of such substances to play its role. Conscious emotion itself, moreover, has specific physical substances and interactions as its substrata; a metaphysical reductionist might refer to the disjunction of those as the catalyst. But it is important for our purposes to take conscious emotion in its functional role, so as to show how it might be that the elements of consciousness — experiential elements of which we are conscious — form components of a self-organizing system. The suggestion here is that feelings and sensorimotor images play crucial roles in their own right, and not just in terms of their substrata. It is because a feeling is conscious, not because it is subserved by

particular molecules, that it makes certain complex actions more timely and more probable.

Here, in an oversimplified sketch, is the way conscious emotion might act as a catalyst to move a self-organizing system to an attractor:

Conscious emotion as a catalyst for coordinated organismic behavior

1. Unconscious biological processes and systems compete to control the organism through drives and instincts that generate behavior directly or indirectly through various physiological states.
2. These systems receive sensory input from the environment, and their activity is shaped by this input. Specific environmental stimuli can cause one system to react more strongly than competing ones, temporarily achieving a state of activation that allows it to dominate the body.
3. This activition produces feedback (via cortico-thalamic and limbic-thalamic loops) that enhance the activity of the sensory input, resulting in coherent representations of the external stimulus together with the bodily response.
4. The resulting representation, consisting of juxtaposed proprioceptive and sensory imagery, becomes available to the cortex for longer-range planning in response to the new bodily/environmental state of affairs.
5. Planning activity in the cortex involves the constructing of imagery of hypothetical responses and probable consequences of those responses, which imagery is itself juxtaposed with imagery of probable bodily responses to those alternative scenarios.
6. Imagery of the bodily response to the most positive action plan affects the current bodily state, enhancing the activity of the subsystems that would be involved in carrying out the plan, and inhibiting others.
7. The entire organism mobilizes around the chosen plan of action.

Items 4–7 involve what I am calling emotional experience. A strained account of conscious emotion in the terminology of self-organization might go as follows:

Phenomenal consciousness is the autocatalysis of imagery-states representing alternative potential attractors for purposes of (a) selection and (b) generation of emotional catalysts appropriate to the production of the selected attractors.

4. Knowing our own minds: The inadequacy of the perceptual model

A central part of the traditional view of consciousness is that we have 'privileged access' to our own conscious states. I am an authority on the states of my own mind. Privileged access has been called into question in recent years, on various

grounds. Psychological experiments have shown that people misunderstand the bases for their own judgments (Nisbett and Ross 1980); Dennett (1987) has argued that our 'purposes' are determined by evolutionary processes and not by our conscious intentions (see also Millikan 1984). But these arguments attack only the causes of our conscious mental states, and not the fact that when we have them, we can be aware of them. I know not why I am sad, but I know I feel sad. In what follows I discuss the application of the enactive model to this question: *How can I know how I feel?*

The traditional Cartesian assumption behind this question is that I and only I can know for sure how I feel because feelings are subjective states, accessible only to the conscious subject 'from the inside.' In a very real but very mysterious way, my feelings are things that can be known, but *only* by me, their subject, because of their very nature. Even vivesection or technological imaging of my conscious brain would not reveal them to the scientific observer. An outside observer can make good guesses about how I feel based on (a) what went in, (b) what came out, and (c) how the observer himself feels under similar circumstances. But not only can the observer not observe my feelings, he cannot explain how it is that I *can* 'observe' my feelings. The nature of subjective knowing is as mysterious as the feelings that are subjectively known. Thus the Cartesian answer to my question can be only this: it is the essential nature of consciousness to know mental states. This answer is clearly inadequate.

Recall the advice of the enactive approach: give up the static, passive, stimulus-receiving model of consciousness, and adopt instead a dynamic, active goal-seeking model. Such a model would apply not just to the body, but to the whole person as a 'mind/body' unity. On this model, there is no mysterious substance that makes up my 'feelings,' and no mysterious 'knowing' or 'observing' relation between those feelings and my subjective 'self.' There is only my dynamic interaction with my environment.

5. The nature of feelings

On this model, what are feelings? They are bodily states that are experienced as desires satisfiable by action (and that promote those actions by catalyzing appropriate organization at the level of unconscious motivations). On the level in which we are interested, that of conscious feelings, these desires and the actions that could satisfy them are ones that I can be aware of. I can be aware of them by forming representations of them in neural activity, in such a way that these representations participate in guiding behavior that can satisfy the desires.

It may be hard to see how this view differs from the traditional one in which mental states are static entities that I, as a subject, passively 'perceive.'

But it does differ, because there is no perceiver on this view; instead, there is an actor. The actor is acting in creating the representation of the goal-satisfying behavior, because such a representation is an essential part of the kind of intentional action necessary to achieve the goal. As we have seen, actions cannot be planned without the generation of conscious (or potentially conscious) imagery of what it would be like for the organism to perform the action, and this image serves as an initiator and guide in the actual performance of the action. Thus the subject is not sitting back and enjoying the imagery in the role of an onlooker; instead, the imagery is a part of the subject's activity. (The idea that action images play a central role in intentional action goes back to William James, who claimed that unless an additional act of inhibition occurred, the image would trigger the action.) In forming representations of goal-satisfying actions, I am getting ready to act; I am not 'just thinking' about it.

It is important not to oversimplify the dynamical model. For present purposes there are two levels on which 'dynamical activity' take place. One level is that of the self-organizing system as a whole. All living beings are such systems. A self-organizing system seeks internal homeostasis while also seeking, in various ways, interaction with the environment for purposes of achieving organismic goals (nourishment, growth, reproduction, etc.). This 'level' of activity comprises many sublevels; the point here is that none of them needs be conscious. The other level of activity is that of consciousness. From the point of view of life as a whole, this level may be of minor importance, but from the point of view of human psychology it is central. Consciousness is itself a self-organizing activity, but it involves only certain components of the living organism, and it involves activity on a macroscopic, not just a microscopic, scale. The type of activity centrally involved in consciousness is what we think of as *intentional actions* (in contrast, for example, to reflexive movements): goal-directed motor behavior driven by emotional states of the person. The paradigm example of such an action is reaching for a coffee-mug because one wants a drink of coffee. But it is important to recognize that intentional goal-directed actions can be less visible than that. Moving one's eyes to focus on an object in the visual field is an action, and so is deliberately *attending to* an object, state of affairs, or aspect of an object, even without overt bodily movements. Attention can be captured involuntarily but can also be under voluntary control; the same thing is true of other bodily movements (Posner and Rothbart 1992).

When we talk about conscious experience, we often use locutions such as 'There is something "it is like" to have "an" experience,' or 'Science cannot tell us what consciousness "is like."' This terminology could lead one to suppose that just as what a rose 'is like' is defined in terms of its sensory properties that we can observe, so what consciousness 'is like' is defined in terms of its phenomenal or qualitative properties that we can mentally 'observe' or 'introspect.' But

that would be to revert to the perceptual model. We know what our experiences 'are like' by *enacting* them, not by *observing* them.

6. Knowing, desiring and enacting

If feelings are actions in the way I have outlined, then what is the answer to my original question: HOW CAN I KNOW HOW I FEEL? The short answer is: BY WHAT IT IS THAT I WANT TO **DO**. The reason this is the answer is that what I want to do is the expression of how I feel, and hence a representation of what I want to do is a representation of my feelings. My 'feelings' are desires satisfiable by actions that I can imagine performing. The desires are *necessarily* associated with actions that I can imagine performing. They are not isolated from such actions, even conceptually. A desire is a desire for a certain kind of activity or interaction with something. If some kind of action were completely unknown to me, a desire for a goal achievable only by such an action would be inconceivable to me. For example, a longing to hear the voice of Maria Callas would be impossible for me if I were congenitally deaf. I would not know what to long for. A desire to love another person is impossible for a victim of autism. (The deaf person and the autistic person could desperately wish to be like others who are not afflicted, but the objects of these wishes would necessarily be general and abstract.) On the other hand, if I have been sighted and am now blind, I can desire to see, even though it is now impossible; if I am not autistic I can long to love another even if I find no one loveable.

But then one can ask: HOW DO I KNOW WHAT IT IS THAT I WANT TO **DO**? This is not difficult to answer, in principle, although in practice the brain mechanisms are far from completely known. I know what I want to do by what it is that I am trying to do. And I know *that* because I am imaging myself doing it without inhibition, or I am imagining ways I can overcome obstacles that prevent me from doing it, or I am imagining myself doing it without overwhelming imagery of unavoidable pain and suffering accompanying the doing of it, or without at the same time sacrificing some other important organismic goal. If I imagine doing something in these ways and there is no competing imagery of alternative actions that I can imagine even more vividly in these ways, then I will do what I want to do. Normally, there are many things that I want to do, and many things that I do not, and one of the things I want to do will prevail in a kind of competition of imagery. The winner of the competition will be the action whose imagery is least accompanied by unavoidable obstacles of the kind described above. Addiction occurs when the desire for the larger holistic balance of the system fails to generate imagery vivid enough to compete with that of more specific desires. Not all of this activity need be vividly conscious; much of

it may be preconscious, unattended to. But to the extent that the competing actions are voluntary, their representations are potentially conscious.

Consider a simplistic example. I want to keep writing this paper; I also want to stop and take a nap. Both images are attractive, in that there are no vivid obstacles accompanying the imagery. But I cannot image doing both at the same time. If I imagine taking a nap now, then that imagery activates pleasurable images of drifting off to sleep, but also other imagery of resuming the writing later with great difficulty. If I imagine continuing to write, that imagery activates anticipatory imagery of increasing discomfort, but also other imagery of being able to sleep later without worrying about the paper. The latter imagery wins the competition, at least on this occasion.

The above examples are very specific, but we have more nebulous emotional feelings as well: restlessness, pleasant or unpleasant moods, general feelings of well-being. Perhaps most commonly, we have desires for a particular type of overall organismic harmony, and we pick out a specific object or state of affairs to represent the desired harmony. But the desired organismic state may be achievable by any number of different objects, as Ellis notes. When I need food the sight of a cookie may make me want it, but the cookie may be only one way, and not even the ideal way, my desire for nourishment could be satisfied. Conscious emotion is a continuum between highly specific desires and generalized mood-like states. At any point on the spectrum, I can know how I feel by imagining performing various activities, and assessing the distance between how they would make me feel and how I want to feel. A state of emotional restlessness will make quiet inaction of any kind less appealing than greater activity; a feeling of well-being will make most contemplated activities more attractive than usual, and I will be free to select one on the basis of pragmatic considerations. The vagueness of the emotion will be matched by the vagueness of the activities that it encourages and discourages. In a state of complete depression, not only will I want to do nothing at all, but I will also be indifferent to external stimuli, and to events that might affect me in any way. Profound states of depression, as is widely recognized, make even active suicide difficult to undertake, even though the depressed subject may find the prospect of death itself more attractive than that of continued life.

The idea is that we know our mental states not by 'observing' them in a mysterious way, but by enacting them. This may seem to leave it mysterious how we are able to attend to ourselves in a reflective mode, instead of just attending nonreflectively to the actions of our bodies in the world. We do, after all, make conceptual distinctions between our internal states and those of the physical world (which includes our bodies). How could anything like introspection occur on the current view? Clearly, we do distinguish between how we feel and how the physical world is, or seems to be.

Gareth Evans describes the activity of making judgements about one's own mental states that are different in content from experiential (often nonconceptual) states about the world that is the *object* of those mental states. Evans is not discussing feelings, but cognitive states only. Nevertheless, his point is relevant, since the basis for the method he describes is the same as the basis for the method by which we know our feelings, on the present account. We do not observe 'the feelings' themselves — they are not 'objects' for us in themselves — but they affect what are objects for us in such a way that we know what the feelings (or, in Evans's case, the cognitive states) are:

> The process of conceptualization or judgement takes the subject from his being in one kind of informational state (with a content of a certain kind, namely, non-conceptual content) to his being in another kind of cognitive state (with a content of a different kind, namely, conceptual content). So when the subject wishes to make absolutely sure that his judgement is correct, he gazes again *at the world* (thereby producing, or reproducing) an informational state in himself); he does not in any sense gaze at, or concentrate upon, his internal state. His internal state cannot in any sense become an *object* to him. (He is *in* it.)
>
> However, a subject can gain knowledge of his internal informational states in a very simple way … Here is how he can do it. He goes through exactly the same procedure as he would go through if he were trying to make a judgement about how it is at this place now, but excluding any knowledge he has of an extraneous kind. (That is, he seeks to determine what he would judge if he did not have such extraneous information.) The result of this will necessarily be closely correlated with the content of the informational state which he is in at that time. Now he may prefix this result with the operator 'It seems to me as though …' What this means is that there is no informational state which stands to the internal state as that internal state stands to the state of the world. (Evans 1982: 227–228.)

Applying this account to the issue of knowing how I feel, I would say: 'I feel a reluctance to go on working, but it makes me anxious to think of taking a nap before I am finished, so I basically feel as if I want to go on until the paper is done.' Here I conceptualize my state in terms of feelings, but I do not 'observe' feelings; rather I imaginatively perform alternative actions in an effort to identify the emotion that they best satisfy.

In cases where my emotion is difficult to identify I can intensify it, as Ellis notes, by imagery that evokes intensified desires. A piece of music may stir me to recognize my hitherto unnoticed longing for the presence of a loved one; thinking about the unbearable constraints of my job, I may visualize a caged tiger in a zoo. My emotions can also be symbolized by imagery that does not literally represent what I want to do, but rather an action of the general type that my organism needs or one that would satisfy the same desires: wanting to achieve reknown in my field, I picture climbing to the summit of Mt. Everest.

The point is not that any particular imagery must accompany a particular emotion in order for me to identify that emotion. It is that whatever the emotion, it manifests itself in interactions with the world. Consciously experiencing the emotion is enacting it: peforming, in imagination or in reality, activities of the appropriate type.

7. Conclusion

The language of inner states as entities observed privately and passively by a subjective inner 'eye' has been well learned since the writings of Plato. Like many other ancient traditions, it has worked well, and continutes to be harmless in ordinary contexts. But for philosophy and psychology it has been disastrous. Since it posits nonphysical but substantive entities that are not only (in the author's view) nonexistent, but arguably logically impossible, it has led to extremes of acceptance and rejection: from Cartesian dualism to behavioristic eliminativism, with attempts at compromise such as epiphenomenalism in between. None of these solutions has been successful because none of them has been able to replace the perceptual model with a workable alternative. Conscious states exist as subjectively observable mental entities with full causal powers, or they exist as partially observable entities with limited causal powers, or they do not exist at all, and consciousness is a myth. The proposal of this chapter (and of many others in this volume) is that if we conceptualize conscious states as emotionally-driven attentional responses of a goal-oriented agent to its body and its external environment, we will then be able investigate their nature without continually running into metaphysical brick walls.

References

Blakemore, S. J., Goodbody, S. J., and Wolpert, D. M. 1998. "Predicting the Consequences of Our Own Actions: The Role of Sensorimotor Context Estimation. *Journal of Neuroscience* 18 (Sept. 15): 7511–18.

Damasio, A. 1994. *Descartes' Error.* New York: Putnam and Sons.

Darnell, J., Lodish, H., and Baltimore, D. 1986. *Molecular Cell Biology.* New York: Scientific American Books.

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

Ellis, R. 2000. This volume.

Evans, G. 1982. *The Varieties of Reference.* Oxford: Oxford University Press.

Faw, B., 2000. This volume.

Georgieff, N. and Jeannerod, M. 1999. "Beyond Consciousness of External Reality: A 'Who' System for Consciousness of Action and Self-Consciousness". http://www.isc.cnrs.fr/wp/wpjea9805.htm

Fruton, J. S. and Simmonds, S. 1960. *General Biochemistry.* New York: Wiley.

Jeannerod, M. 1994. "The Representing Brain: Neural Correlates of Motor Intention and Imagery". *Behavioral and Brain Sciences* 17 (2): 187–244.

Kauffman, S. 1993. *The Origins of Order.* Oxford: Oxford University Press.

Millikan, R. 1983. *Language, Thought and other Biological Categories*. Cambridge, Mass.: MIT Press.

Nisbett, R. and Ross, L. 1980. *Human Inference: Strategies and Shortcomings of Social Judgement.* Englewood Cliffs, N. J.: Prentice-Hall.

Panksepp, J. 2000. This volume.

Platt, M. L. and Glincher, P. W. 1999. "Neural Correlates of Decision Variables in Parietal Cortex". *Nature* 400 (July 15): 233–238.

Posner, M. and Rothbart, S. 1991. "Attentional Mechanisms and Conscious Experience". In *The Neuropsychology of Consciousness,* Milner, A. D. and Rugg, M. D. (eds), Academic Press, 91–111.

Rawn, J. D. 1983. *Biochemistry*. New York: Harper and Row.

Part II

Toward an Ecological Science of the Affective Sphere

CHAPTER 5

The 'Mind'/'Body' Problem and First-Person Process

Three Types of Concepts

Eugene T. Gendlin
University of Chicago

1. Distinguishing the person from 'being in control'

There is now an ongoing discussion toward a 'science of consciousness,' including several journals, consisting of neurophysiologists and philosophers, a group which is encountering major difficulties around the old mind/body problem. This discussion needs a wider philosophical approach that can be related to the more recent psychological findings. Up to now, there has been a strong tendency to fall into the pitfalls of the most traditional formulation of the mind/body problem in which 'the mind' is the consciousness, and 'the body' is taken as represented in neurophysiology.

For example, much is made of the finding that when a person decides to move, the neurophysiological measures pick this up before the conscious person has deliberately decided to move (Libet *et al.* 1983). So it seems as if the person is unnecessary — free will reduces to bodily, i.e. neurophysiological factors.

What is missing here is that *'the person'* is much larger than our deliberate conscious capacities. The person includes the body-as-internally-sensed, and it also includes a much greater variety of reflectively aware experiencing than merely the deliberately controlled. What we control is a relatively small part of our sentient experiencing and experienced internal and external actions. Of course it is an error to split mind from body, but this is controversial. It is a more immediate error to define control as marking 'the mind' as against the body. This is a pitfall in which the person drops out and one defines 'the conscious mind' as control, reducing the whole person to 'the body' considered only as the mutually independent components such as neurons transmitting linear and ultimately inorganic forces.

The puzzle about the body knowing our decisions before we consciously know them might make us miss the fact that there is an *inwardly experienced body*, and that the reflective and bodily-sentient person is much wider than conscious control.

Freud, Jung, and many others (Gendlin 1962/1997, 1981, 1992a, 1992b, 1999) have shown that the person far exceeds the deliberate consciousness, both in terms of its control over behavior, feelings, and interpretations of situations, and in terms of 'sub-parts' of the personality. Around many kinds of processes and practices an inner geography has long been well-known to specialists. In some cases this has been formalized, for example in meditation, and in the teaching of musical performance. In others it is left up to the individual. For example, philosophers and scientists can describe a great deal of intricate detail concerning how we manage to obtain and maintain the peaceful and yet excited state we need in order to think and write. There are great differences among individuals as to how much they have explored their 'inner territory.' But everyone is familiar with such a realm, where more happens than one can control.

Both everyday experience and inner exploration reveal a gamut of responses, feelings, physical acts that 'can come,' which one does not simply control. One becomes familiar with this aspect of 'oneself' and the need to do something more intricate than either control or passivity — rather disposing oneself toward the coming, training habits, awaiting the coming, or developing friendly attitudes that maximize the chances of the given condition's coming (Gendlin 1981).

For example, we seek a certain sense of confidence when about to give a talk to a group. We hope the right words will 'come' as we need them. Or, for example, we cannot write down in advance all the sentences for what we will say because reading them off isn't appropriate in many situations — for instance if we are preparing for a tense meeting with the boss, or speaking to a small group. We know that under some conditions what to say will come even better on the spot, than beforehand in preparation.

This coming may seem mysterious, but we can notice that even in ordinary conversation we must let our words come. When they come smoothly we needn't notice, but everyone notices this keenly when it is even a little bit difficult to express something precisely, or to communicate a condition for which there are no ready-made phrases.

When we think, including this case now about this philosophical question, we like to own the products, our own ideas, nobody else's, but actually 'our' ideas 'come' only if they like, so to speak, as the phrase 'the muse' is meant to suggest.

Countless other everyday circumstances can let us recall how much more 'we' are, which we live from, aim for, sometimes access, but do not consciously simply control.

But now, what is the reason (sometimes distinguished from 'the cause') why our words might not come when we wish? It might be anxiety causing tension which reduces our inner access, as everyone knows. On an exam we cannot always have what we know come to us as it would at home. There are also complex psychological reasons why we might not easily overcome old experiences of rejection. Or, we become constricted for unique personal reasons, the complexity of which cannot possibly be represented physiologically although there might be simple indices of some of it.

For example, if under dangerous conditions an experienced pilot says not to fly, we do not require a complete explication of the reasons, nor do we insist that the reasons even be explicable.

So the first point I want to make is that as inwardly experienced persons we are far larger and include much more than we are (or can be) aware of, which we come to know and toward which we dispose ourselves in complex ways that far exceed simple direct control.

Concerning the Libet experiment we must not miss the fact that it is still *the person's* motions, both if the moment of moving is first consciously decided and if this is first picked up physiologically. What it is that moves the person to pick a particular moment need not be in the person's control.

2. Distinguishing the person from indices

Various specialists know very well that there are physiological indices of what persons are going through in their lives. For example, a gynecologist has told me that when his patients are in psychotherapy he can always recognize from the condition of the vagina whether the therapy is going well or not. He regularly checks this out with the therapist. A doctor or body-worker who has many years of experience with people's feet can tell from the sole of the foot whether the person is just then living tensely, anxiously, on the edge of things, or is doing comfortably well.

In the case of feet this has given rise to a small 'science' in which the principle is enunciated, that one can diagnose people from their feet. However, they have not yet made the error of assuming that the personality is unnecessary, that people can be reduced to their feet. Yet this is the error made by many reductive philosophers of mind in regard to the neurophysiological body, when it is assumed that the 'real story' of what causes the higher-order processes that include consciousness is all to be found at the level of individual neuron activity and additive combinations of those, as if self-organizing and multiply realizable patterns of activity had no ontological status except as epiphenomena of the linear neuronal activity. That rendition of the body is even narrower and more

limited than the specialist's broad impressionistic observation of feet.

We would not be at all surprised if an expert could recognize when a person's foot tightens as the person gets ready to decide to move. The danger both to us as humans and as philosophers now is that our best knowledge — what social policy-makers have to go on, will be considered to be only what registers on the sole of the foot, or in the narrow range of what registers in physiology. And this will seem convincing because, of course, *something* does register there.

My second point is that the scope and intricacy of the person cannot be represented by, or reduced to indices, although there are undoubtedly many kinds of those. This is because the person is a self-organizing process, with both conscious and unconscious aspects, that appropriates, rearranges, and replaces its own needed components. The process as such does not reduce to those components alone, but rather *uses* them to maintain a certain pattern of organization.

3. Process concepts *versus* constituent-unit concepts

The traditional 'mind/body' problem concerns neither the person nor the body. This is not only because the far larger part of both is left out, including what cannot be split. There is another quite different reason why the traditional problem concerns neither mind nor body even were we to accept the narrow definitions of both.

The traditional problem concerns the incompatibility of two vocabularies, two *kinds of concepts,* that of physiology vs. psychology (whichever of the variety of kinds of concepts one chooses to place here). There seem to be only very few relationships between our knowledge of the psyche and our knowledge of the body, but this is not because psyche and body are unrelated. In the human being they are intimately related. It is not at all the physiology which has difficulty relating to the psyche. The human physiology is through and through the physiology of a psychological person. Nor is the lack of relation generated by the psyche which is inherently physical. It is rather the fact that we happen to have developed studies of the body in terms of mechanistic, third-person space-time-grid precise assumptions which create a *kind of concept* that is or seems utterly irreconcilable with the imprecise, holistic, first-person-involving *kind of concept* we have been using to formulate most of what we know in psychology.

There is no doubt that this difference in kind of concepts is due to some difference in what is being studied. Certainly one can say that 'mind' and 'body' are *indirectly* responsible for the problem of finding few relationships. Certainly the two realms of subject-matter cause us to employ two different kinds of concepts, and we can look into this further. I will return to this question. But

what is discussed as the traditional mind/body problem is the intractable relationship, not between mind and body, but rather between the different kinds of concepts which are currently in use.

If this is not recognized, an unresolvable mystification results. The problem then becomes ideological and political. People will line up on either side, as they do when complex political questions are divided into two sides, so that the complexity of the philosophical issue is lost.

The philosophical issue has to be sorted out into two related but very different questions: The first concerns relations between kinds of concepts. Only the second question concerns what it is that has led to the development of two or more bodies of knowledge couched in the different types of concepts.

I will present my own views about different kinds of concepts separately below, because those are controversial. What seems to me most important and also beyond question, once it is noticed, is that the traditional and currently discussed 'mind/body problem' concerns the difficulty of relating different kinds of concepts — holistic concepts *versus* constituent-unit concepts. In the realm of the physical sciences, much success has been obtained up to now by using constituent-unit concepts; we explain an overall process by dividing it up into constituent units and then explaining how those units can combine to make that process. By contrast, first-person psychological descriptions can be best understood neither in terms of wholes nor constituent units. I shall present a third way of understanding them, in terms of what I shall call 'process' concepts.

4. Why psychology has failed to emulate the physical sciences

For more than a hundred years now there have been constant efforts to employ in psychology the same kind of concept which has been so successful in physiology. This attempt regularly fails. For example, many efforts to define emotions by physiological measures have found no differences among them. All one gets is a difference between 'arousal' and 'non-arousal,' alike for love and hate, and for anger or happy excitement. Only a few poor variables have come from attempts to define psychological variables by beginning with machine-defined physiological correlates.

The effort has therefore regularly turned to defining psychological variables freshly as such, but still dividing psychological aspects into precise entities, factors, traits, units that one could count and treat as the natural sciences treat their variables. Many theories but very little successful operational research has come of this. Some successes have been achieved, but only where the topic is highly specific, such as specific phobias, specific events of trauma, or changing a specific behavior. Factor-analysis and psychometric measures do produce

personality profiles that have some uses, but they have not become superior to clinical impressions. (Asked to explain his definition of a patient's psychometric tests as indicating 'psychosis,' the psychometrist first argued for his measures, but then said "After all, the man is in the hospital.")

There are many measures that produce quantitative results with excellent and many-times replicated reliability, but what the numbers mean is not very precise. The names of the measures are not univocal. Measures of 'anxiety' by skin resistance (sweating) do not correlate well with measures by Rorschach and TAT, nor those with measures by paper and pencil questionnaire, or with evaluation by clinicians. The various measures of 'anxiety' get at different first-person experiences. First-person process is much more differentiated. When students choose a measure, they need to administer it first to themselves. Only so can they tell what it measures, and whether their prediction should really apply to it. The attempt to use in psychology the kind of concept that was developed for use in the physical sciences, and thus could simply conform to physiology, has largely failed.

It is now time to reverse the implication and to examine the assumption that a psychology ought to be successful with the kinds of concept used in physiology. Why should it be the case that we can best understand people by plotting them within the model of the space-time-locations, the model with which we chop everything into precise units out of which we can then reconstruct whatever we study? This is the most successful model in history so far. But does it follow that it must be the best model for studying those beings who devised this model? There are other models, for example, ecology employs the holistic model. Below I will sketch out a third alternative. Right now I want first to argue that its success in so many fields *does not necessarily imply* that the unit-model should be successful in this one.

Once we have freed ourselves from this metaphysical assumption, we can also see many reasons why this kind of concept is unlikely to succeed here. I will first discuss what is limiting in the unit-making kind of concept, and then I will show why another kind of concept is more likely of success.

5. Both process and constituent-unit concepts apply to both nature and persons

A large branch of philosophy, although admittedly difficult to read, has now developed a number of deep-going critiques of the time-space-unit type of concept. This trend follows in the footsteps of Dilthey, Husserl, Heidegger, Derrida, and Wittgenstein (See Gendlin 1962/1997). It seems obvious that since

we humans created this model, we could be larger than what can be represented within it.

We can recognize that the unit-model drastically narrows anything that is studied within it. We can with certainty assert that nature is much more than what the 'natural sciences' could possibly render. We see this easily with other models. For example, ecology regularly predicts future events better than the sciences that use reductionistic-unit concepts. Because of this, ecology has now established itself next to the regular science, and it is taken into account when policy-makers must base their decisions on the best available evidence.

Do not say that 'nature' lends itself to the reductive model while humans somehow do not. Say rather that nature as a whole has only recently talked back, while as humans we can readily recognize that the reductionistic sciences leave most of us out. Where I said above that our psychometric measures have not yet equaled 'a clinician's evaluation' — there is nothing parallel I can say from nature, except what nature can say back via ecology.

To use the reductionistic kind of concept, we divide what we study into units. We then attempt to reconstruct what we study out of those units. So a picture can be reduced to dots and reconstructed elsewhere out of the dots. Is it 'the same' picture? With 'high resolution' it will be sharper than what you ordinarily see. No one misses what the dots leave out.

This procedure is successful with anything considered as something over there, something we merely perceive like a picture, something that exists alone without us, that is to say anything considered as a third-person thing. But this no longer holds as soon as we turn to first-person process, how we live as wide persons-and-bodies.

I told one film-maker my amazement that all the old films are back. For years a film would date a person. You had seen it only if you were old enough. Now they are all contemporary with each other. My daughter knows them all. "Yes," he said "but they were made for people sitting in the dark and looking straight ahead at a large screen. The effect is utterly different on the little TV at home. We now use a lot of ways to take into account that the films will appear on little TVs." So it is not always 'the same' picture. One can probably measure the difference in impact by physiological arousal, but this will not contain the aesthetic judgment of the film-maker.

I have been arguing that we can recognize by inspection that the time-space-unit type of concept places drastic limits on anything represented in them.

6. We need both types of concepts

That the film-makers have developed a body of knowledge is an example of my

next point. What kind of knowledge is this? It is a very important type because we have it in many areas of human life.

With the passing of primitive cultures we have largely lost their knowledge of drugs in plants. Any large pharmaceutical company today would offer a billion to buy it, if this knowledge still existed, but it was only taught from teacher to student, and was not written. Bodies of knowledge of that kind exist today in many people from computer hackers to four-star chefs, from pilots to therapists, from gynecologists to artists. We have much process-knowledge. This comprises much that has been differentiated and written-down, but it always also includes the person for whom this is written, the person who reads the instructions and must locate them within a much richer and more intricate experiencing.

This kind of knowledge seems unscientific if we judge it by the precision of the distinctions between units in reductionistic physics, but if you examine it in terms of *kinds of processes*, then its distinctions appear well-precisioned. There are distinctions between processes and sub-processes, and between desired and undesirable variants. Pitfalls are often well-defined. The conditions for obtaining a process, and the definitions of outcomes are often precise.

What we need and do not yet have is a way for this kind of knowledge to take its place beside ecology as a kind of knowledge to be developed and consulted, rather than considering reductive science as capable of standing alone. I think this could be achieved if people begin talking about various kinds of concepts, various models, so that it becomes clear that there is different knowledge inherent in them, and that we need at least all the ones that we have.

But what is this third model beyond reductive science and the holistic ecology with which we are familiar? Let me move from the above examples to the kind of model that could accommodate their kind of knowledge. This is a *first- person model* which I call *'a model of process,'* a model in which each next bit is a newly-made whole. It does not consist of predictable units in a Laplacian universe assumed to be a mathematical grid which we only observe. Rather, it always involves us as ongoing persons. The next bit of process arrives only for someone, and sometimes only if that someone looks for it to arrive. This may make those who are used to units despair, but this model does offer equally precise distinctions in other regards, as I said above.

With this model we can re-understand biology as a life-process that generates and organizes its own next steps. The crucial self-organizing character of life process also shows a capacity to produce quite novel life-forwarding steps (Monod 1971; Gendlin 1981; Kauffman 1993). Self-organizing cannot be represented in the unit-model in which all points and units are passive and only an external observer relates them.

7. We need all three models

The political side of the issue must be looked at without flinching. The reductionists want to think that they have earned the right to treat people as mere physiological and neurophysiological bodies, i.e., as collections of separately-explainable components interacting, because of the general success of reductionistic science, and the habit of considering this reductionistic science as standing alone. Although the medical and biological sciences involve large uncertainties and are admittedly not at the stage of physics and chemistry, recent progress has been so intense that they have all but earned this 'right' to stand alone as 'the science' of human bodies. Our society seems to recognize them as such. Official grant money flows to them, and social policy is based on them.

This demand to be considered as a sufficient science alone, based on the general success of reductionistic science, leads the other 'side' to attack science as a whole. The other side includes many groups including holistic physicians and ecologists. The opponents of the reductionist basis for policy decisions attack science as a whole. Thus they play into each other's hands, and the issue is not faced directly.

What is at stake here can easily get lost. It is what we say to the public, to our society, and to those who serve on policy-making committees. How can they decide other than on the best available knowledge and evidence? The question is whether that is neurophysiology and other information sciences *standing alone.* It has seemed so up to now, and that must now be broadened to include more, but we cannot expect anyone simply to ignore the hard evidence of the reductionistic sciences. We must show how that evidence can be placed within a broader context.

In philosophy the assumption of reductionism retains a great attraction because it is so neat and simple. If one is concerned only with broadly general terms, then 'the unity of the sciences' seems like something that ought to be achieved. But this assumption is not held by philosophers who like to look at actual science.

The assumption that everything should reduce to physics has only a very limited application in chemistry. That life process should reduce to chemistry is not actually used to limit organic chemistry. And so on up. It is merely a metaphysical assumption, not what is employed in practice. The opposite is much more striking. Each specialty creates its own terms and as the years go by almost every speciality develops more and more terms. Only 'in principle,' not in any practice, do those reduce to physics or inorganic chemistry. So we must not turn physiological reductionism into some kind of ideology to be defended by general arguments that have no relation to scientific practice. From success in physiology it does not follow that psychology should be capable of being reduced to it, nor

that it should employ the same kinds of concepts.

We would not expect a first-person process science to contradict the genuine findings of reductionistic science, any more than ecology does. But it places those findings within a broader perspective that can be more useful for certain purposes. One of those purposes is to understand first-person processes in such a way that 'the person' does not drop out. Our first-person experience offers a knowledge basis with useful practical applications, just as does ecology, even though neither can be translated into the concepts of the unit-component types of sciences.

References

Gendlin, Eugene. 1962/1997. *Experiencing and the Creation of Meaning.* Evanston: Northwestern University Press.

Gendlin, Eugene. 1981. *Focusing.* New York: Bantam.

Gendlin, Eugene. 1992a. "Thinking Beyond Patterns". In B. den Ouden and M. Moen (eds), *The Presence of Feeling in Thought.* New York: Peter Lang.

Gendlin, Eugene. 1992b. "The Primacy of the Body, Not the Primacy of Perception". *Man and World* 25: 341–353.

Gendlin, Eugene. 1999. "A New Model". *Journal of Consciousness Studies* 6: 232–237.

Kauffman, Stuart. 1993. *The Origins of Order.* Oxford: Oxford University Press.

Libet, Benjamin, A. G. Curtis, E. W. Wright, and D. K. Pearl. 1983. "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). The Unconscious Initiation of a Freely Voluntary Act". *Brain* 106: 640.

Monod, Jacques. 1971. *Chance and Necessity.* New York: Random House.

CHAPTER 6

Dissolving Differences

How to Understand the Competing Approaches to Human Emotion

Valerie Gray Hardcastle
Virginia Polytechnic Institute

Quick pharmaceutical fixes are increasingly the treatment of choice for emotional problems. Depressed? Take Prozac. Anxious? Take Xanax. Fearful? Take Paxil. It's an anti-depressant, but it also cures some phobias. Sexually inhibited? Try Wellbutrin. That sometimes works.

A friend of mine had a dog that was depressed. The doctors diagnosed the pet with a mood disorder and prescribed medication as a solution. This is how much biomedical views about emotions and how they should be controlled have infected our culture. My point here is not to knock modern pharmaceuticals, nor is it to claim that, in our hurry to find quick fixes, we over-prescribe mood altering drugs. But the fact is that popular culture now sees our emotions as simple brain twitches; to change our feelings, we just change the twitch. Depression and anxiety are analogous to headaches or heartburn; they are all simple bodily responses to various stressors.

It would belabor the obvious to point out that our emotions are much more complicated than acid indigestion. At the same time, wondering how it is we should understand human emotions is a legitimate theoretical question. Popular culture and the popular press don't have it quite right, that is certain, but how far off are they?

This essay will try to address that question. I propose a view of emotions that pays homage to their complexity as well as to their basic neurobiological roots. In brief, my main points are these: There are currently two dominant traditions in the study of emotion, although not all approaches fit neatly into either category: the constructivists and the reductionists. The constructivists hold that human emotions are constructed out of our social interactions with others. In contrast, the reductionists believe that our complex emotions reflect common affective responses found across the animal kingdom. Here, I try to show — perhaps paradoxically — that both camps are largely right.

1. The social construction of emotion

For most of our psychological properties, if not all of them, one can find a social-constructivist hypothesis arguing that that property comes more or less directly from our complex social nature. For example, in the late 1920s and early 1930s, the Russian psychologist Lev Vygotsky argued that linguistic meaning comes out of our cultural interactions with others (Vygotsky 1962, 1978). A pointing finger means nothing until it is embedded in a particular cultural-historical environment. Similarly, we do not know what a pointing finger means unless we are embedded as well. Culture teaches us the meanings of all sorts of arbitrary symbols. As it does so, it also teaches us how to think, for meaning per se cannot exist apart from linguistic practices, and linguistic practices cannot exist apart from a social group.

The study of emotion fares no differently. The social-constructivist move is quite popular these days in anthropology, sociology, and all manner of cross-cultural studies. For example, we have learned through these studies that Chinese infants do not express their emotions facially as much or in the same manner as either Anglo American or Japanese babies. Japanese infants are more expressive than Chinese, but less than Anglo Americans. The differences manifest themselves in quite specific ways; minute differences in facial actions across the three cultures have been identified (Camras *et al.* 1998).

In addition, there is cultural variability in reading facial expressions, even the expressions of the supposed universal emotions of anger, contempt, disgust, fear, happiness, sadness, and surprise. Japanese subjects misrecognize facial expressions in standard photo arrays. Interestingly, neither the ethnicity nor the gender of the models has any impact on how well Japanese do (Shioriri *et al.* 1999).

Moreover, we have learned that different cultures not only express their emotions differently, but they apparently feel different emotions. The Japanese experience a familial shame quite foreign to Westerners. The Ifulak experience a grieving love entirely foreign to Westerners. However, it is very difficult to separate genuinely different emotions from different ways of describing the same or similar emotions. As long as we believe that different cultures can have embedded in them different ways of defining, expressing, and understanding emotions (and we believe that we really only have good access to one another's complex conscious states via self reports), then we will not be able to differentiate different descriptions of the one thing from different descriptions of different things. Koreans manifest depression in terms of somatic symptoms instead of psychological ones (Pang 1998). Do the somatic symptoms that Koreans exhibit reflect the same underlying depression that Westerners experience, even though the linguistic descriptions of the feelings, their behavioral

manifestations, and their psychosocial historical stories differ? That question as yet remains unanswered.

Nevertheless, the differences among cultures we can identify apparently begin quite early. Chinese toddlers show different affective responses from Anglo Canadians. In particular, Chinese toddlers are more inhibited than Canadian. How these differences get embedded in the psyches of the infants is no mystery, for Chinese mothers reward behavioral inhibition with acceptance and encouragement, while Canadians reward extroversion (Chen *et al.* 1998). One implication of these facts is that the affective behavior itself is assigned different meanings across cultures. What counts as "good" conduct in China is "bad" in Canada.

One can see parallels between the view that emotions are fundamentally social constructions and the view William James advocated historically. James hypothesizes that

> *The bodily changes follow directly the perception of the exciting fact, and ... our feelings of the same changes as they occur IS the emotion.* Common-sense says, we lose our fortune, are sorry and weep; we meet a bear, are frightened and run; we are insulted by a rival, are angry and strike. [T]his order of sequence is incorrect.... [W]e feel sorry *because* we cry, angry *because* we strike, afraid *because* we tremble. (James 1890).

We feel what we do because we interpret our bodily states in a particular way. That is, the environment triggers a certain physiological response. We then feel an emotion as a psychological response to the alterations in our body. Like contemporary social-constructivists, James believes that we can only feel an emotion as a cognitive reaction to physiological changes. Social-constructivists go one step farther than James and claim that the physiological reaction itself is a learned reaction. In both cases, though, feelings follow bodily behavior. With James, the feelings follow as our brains' response to changes in our viscera. With the social-constructivists, the feelings follow as we learn the appropriate psychology to manifest in our particular cultural environment.

James's views are not without merit. We do need some autonomic arousal in order to feel full-blown emotions. George Hohmann has discovered that patients with spinal cord injuries simply do not experience the full range or depth of emotion that the uninjured do (1966). Moreover, the higher up the spinal cord the injury occurred, the more sympathetic arousal is disrupted, and the more attenuated the emotional response becomes. We need visceral reactions in order have feelings.

At the same time, physiological changes are not enough. A shot of epinephrine promotes sympathetic arousal, but most do not feel any particular emotion

associated with the changes (Marañon 1924). We need both an internal bodily change and some triggering cognitive event.

James Pennebaker gave subjects one of three consent forms: they either agreed to partake in an experiment which might result in a brief but intense (and harmless) pain, or one that would produce a pleasurable experience, or one in which the sort of experience they would have was not mentioned. In each case, however, the experiments proceeded in exactly the same way: subjects worked a series of mathematical problems and then placed their finger in a machine that contained a vibrating emery board. They were then asked to rank the sensation on their finger along a 16 point pleasure-to-pain scale. Remarkably, subjects who signed the forms that suggested they might feel pain reported the vibration as "stinging" or "burning" and indicated that the sensation fell on the pain end of the scale. Many believed that they had been shocked electrically. Subjects who signed the forms intimating pleasure reported a "tingling" feeling and ranked their sensations on the pleasure end of the scale. Subjects who received no indication of what the experience might be like reported the sensation as "vibrating," and ranked the experience as neutral. Even though the subjects all received the same input, their responses differed significantly, varying as their expectations did (Anderson and Pennebaker 1980).

The converse is also true. Expectations of having a particular feeling can override bodily reactions. For example, Nisbett and Schachter gave subjects a placebo pill and told half of them that possible side effects include hand tremors, heart palpitations, butterflies in the stomach and other symptoms of arousal. The other half they told the pill would cause itching, slight numbness, and other symptoms not associated with arousal. They then asked these subjects to report how much pain they experienced with a series of mild electric shocks. Of course, being shocked produces arousal in all of us. The question being investigated was how the subjects who believed that their autonomic response might be caused by a pill would interpret their pain. It turns out that these subjects were willing to tolerate shocks four times as intense as the subjects who believed that their pill resulted in no changes in arousal or in normal subjects who had not taken any pill whatsoever (Nisbett and Schachter 1966). Subjects suppressed otherwise normal emotional responses to electric shock if they could attribute their bodily reactions to other causes.

Our visceral reactions influence our affective responses. We can also, to a certain extent, detach our visceral reactions from our cognitive interpretations of what they might mean. Both of these claims dovetail with a social-constructivist's view. In both cases, how we interpret something determines what we feel. And how we interpret something turns on how we have learned to interpret the world through our interactions with others.

2. The continuum hypothesis

In contrast to the constructivist views of emotion, which stress complicated social reactions and interpretations, reductionist approaches focus almost exclusively on the seven basic emotions of anger, contempt, disgust, fear, happiness, sadness, and surprise. The guiding hypothesis is that our emotions are but human manifestations of these seven primitive affective reactions. All emotions exist on a continuum; we humans don't bring anything particularly new to the mix. We might be more complicated, but we aren't different in kind.

The West's current love affair with selective serotonin re-uptake inhibitors is one demonstration of the reductionistic continuum hypothesis. Things like depression and anxiety are just imbalances in a few (maybe even one) neurotransmitters. Increase the amount or the effectiveness of the neurotransmitter and you will thereby eliminate the feeling. (That we have hundreds of different types of neurotransmitters in the brain and that we have currently isolated only a score or so seem to bother no one.)

Of course, our emotions are not stereotypical by any means. Popular culture gets a lot wrong. Elizabeth Kübler-Ross's five stages of dying are one case in point (1969). Kübler-Ross and the popular press have it that we all pass through the same steps when grieving. In the beginning, we deny our tragedy, then are angry about it. Next we bargain with our god for another chance and become depressed when we have to accept the inevitable. Finally, we accept our lot with grace. These stages are now so ingrained in our world-view that we discuss the bereaved in those terms: "Don't bring up her loss; she's in denial." "Joe went very quickly and peacefully after he accepted his condition."

In the first place, we all do not react to crises in the same manner. Some show extreme emotions; others show very little (Burgess and Holstrom 1974, Natterson and Knudson 1960). Secondly and more importantly, we do not all pass through the five stages, even under our own schedule (Lawson 1976). Third, and maybe most important from a therapeutic point of view, we do not all 'accept' our lot, even after considerable time (Burgess and Holstrom 1978, Parkes 1970, 1975). A significant number of women who have undergone mastectomies to remove malignant breast tumors still feel anxious and depressed over their loss a full year after the operation. Almost half of bereaved individuals are still anxious two to four years after their loved one died. A quarter of rape victims still do not feel 'recovered' four to six years after their assault.

Many of us do not go gentle into that good night, no matter what our culture expects of us. Many of us do. There aren't clear generalizations we can make about the 'natural' stages of grief and sadness, even within a single social context. Our complexities translate into a variety of individual reactions, even when the triggering events are culturally recognized and anticipated.

At the same time, it is hard to claim that our emotions are merely products of our environment. Joe LeDoux has isolated the lower brain circuits used in processing fear (cf., his 1996). These circuits appear homologous across species, which gives us one reason to believe that at least some aspects of some fear processing are genetically determined.

Indeed, what most humans are typically afraid of makes sense only in an evolutionary context. Humans react strongly to snakes, for example. Many people fear snakes outright. Yet the chance of the average citizen in the Western world being bitten by a snake or having any other sort of unpleasant interaction with one is exceedingly slim. Most of us live significant portions of our lives without ever even seeing a snake in the wild.

If we were to form fears based solely on previous experiences and what we have learned about the experiences of others, it would make much more sense that we not worry so much about snakes, but be very afraid of guns, automobiles, and perhaps knives and other sharp objects. But we don't and we aren't. Phobias associated with any sort of human artifact are in fact quite rare. Most are of things we might have encountered long ago, things it would have been wise to fear back in the Pleistocine era. Human phobias concern things like snakes and spiders, open spaces and high places. Some take these facts as an indication that our affective system is hard-wired and slowly evolved by natural selection over evolutionary time.

But even though the subcortical fear responses have been largely isolated, that doesn't tell us too much about our other emotions. Fear could be special, since it would have played such a distinctive role in our evolutionary history. Certainly, what we do know about the other basic emotions tells us that they are housed in different circuits in the brain. We don't have a single emotion center, despite all the fanfare our limbic system receives. And what properties these different streams have are still unknown. Even if our fear system is largely innate, our other affective responses might not be.

Certainly, emotions associated with joy and happiness have large cortical components. It has been standard procedure during brain operations to stimulate the areas surrounding the region being lesioned to make sure that the least amount of damage is done to cognitive processing. Since individual brains vary considerably, surgeons have to map out the processing areas for each brain separately. In a recent procedure, neurosurgeons stimulated the region of a patient's brain that was obviously associated with laughter (this in itself was interesting, since the region being stimulated was in motor cortex and not in any of the more 'cognitive' regions). The patient laughed, which is not too surprising. But she also confabulated explanations for why she was so amused — she claimed the doctor had told a joke (when he hadn't) or that her situation was funny (which it wasn't). The patient was not only laughing, but she felt genuine amusement and

justified her reactions by analyzing her environment (Fried *et al.* 1998).

This reaction stands in stark contrast to other patients who experience various reactions when having their brains stimulated. Quite often, they are only too aware that they are seeing the spots or hearing the music because a surgeon was poking about in their cranium. My point with this example is not that we can locate centers outside of the amygdala that index emotions, but it is that the patient could only feel amusement with some sort of cognitive interpretation. She couldn't just laugh for no reason. Since she didn't have a real reason, she invented one. Amusement is much more cognitive than something like fear. If this conjecture be correct, then amusement would be much more subject to environmental influences. We learn what is funny from our social surroundings. Few human attributes are more culturally specific than humor.

So we have different processing streams and these streams vary in terms of how biologically hard-wired they are. Some affective reactions are heavily cognitive; others aren't. None of this is surprising. At the same time, things like SSRIs and other anti-depressants do work well. All we have to do is enhance the response of a single neurotransmitter and we can change our entire affective outlook. Even if our emotions are cognitive in a deep sense, they can still be altered by changing our neural firing patterns, and these patterns can be traced to homologous circuits in other animals.

Moreover, we can find striking commonalities across our most cognitive of emotions. Even though the contents of depressive thoughts are culturally influenced, the basic symptoms of depression aren't (Ebert *et al.* 1995). Depressed people are disinterested in life and find it difficult to move through their day. Similarly, even though society determines what amuses us, we all express joy in more-or-less the same way. All humans laugh. There is much variation in the details of our emotional responses, but just as much remains common across folk.

It might seem that we should divide our emotions along a continuum of hard-wiredness. Things like amusement and perhaps depression can stay at the more soft-wired, socially determined end and things like fear and phobias would remain at the genetically determined, hard-wired end. The simpler, more primitive emotions come fully formed in all humans; the more complex emotions develop during socialization.

This is a nice idea; it would make understanding emotions and the conflict between the two approaches easier if it were true. But it is false. For some of our most 'complex' and 'social' emotions are in us from the get-go. For example, empathy, our capacity to experience others' pain or pleasure as our own, looks to be inborn. It is "part of our native endowment" as William Damon claims (Damon 1999: 72). Even young infants cry when they hear others crying or coo when they see the happiness of others. Small children hug and otherwise try to comfort their loved ones. Alison Gopnik tells a delightful story about her

daughter's drive to console. Gopnik had had a hard and frustrating day and sat sobbing in frustration on her sofa one evening. Her daughter happened upon her and did the only thing she knew to help dry someone else's tears — she covered Gopnik in bandaids (Gopnik, personal communication).

Gopnik's daughter isn't unique; we all start out feeling one another's pain, though we do not all retain this characteristic. One legitimate social concern is that we are raising children to lose their empathy. A teen-age boy interviewed after he had just savagely beaten an 83-year-old woman replied, "What do I care? I am not her," when asked how he could do that to someone else (as reported in Damon 1999: 72). These sorts of incidents are not uncommon and have become more frequent over the past decades (Achenback and Howell 1993).

At the same time, it is clear that children from widely divergent cultures and backgrounds all start out the same with respect to certain core reactions. Some of these emotions are fairly non-cognitive and 'primitive,' such as fear of predators; others are extremely socially dependent and complex, such as empathy. The long and the short of it is that we can't divide our emotions along a simple/complex continuum that parallels the innate/acquired continuum. It might be that all our emotions, no matter how complex or socially constructed they seem, stem from our genetic heritage. We see individual variations in how these emotions are expressed or the cognitive contents associated with them, but there are distinctive cores around which we can group all our feelings (see also Griffiths 1997).

3. Multiple processing systems

We are confronted with two apparently incompatible alternatives when understanding emotions. Yet both seem compatible with known data. On the one hand, emotions appear socially constructed and defined. On the other, they appear to reduce to basic neural circuits preserved in our brain across evolutionary time. Which approach is the correct one?

I would like to make a radical suggestion: they both are. I shall have to digress a bit and talk about something else to explain what I mean. Facial recognition in humans appears to be a specialized process. I am going to duck all the complicated questions here concerning whether this processor is modular or informationally encapsulated, for the answers don't matter to our purposes. Suffice it to say that we are good at recognizing faces in a way that we aren't at recognizing brands of washers and dryers, or even the faces of other animals.

Tradition had it that we develop our talent and preference for face-gazing at around two to four months of age, when we start to prefer looking at faces over looking at line drawings of faces or at empty ovals. (I take my account of the

development of facial recognition from Johnson and Morton 1991. See also my 1995). Some experiments, though, did not support tradition. In particular, some experiments showed that infants as young as nine minutes prefer to look at faces (Goren *et al.* 1975, Ellis and Young 1989). At first one could claim that such results were anomalous or perhaps reflected greater sensitivity to infant desires. Then it became known that infants one month old had little inclination to prefer faces (Johnson and Morton 1991). Infants are born wanting to look at faces, then lose that impulse, then regain it again at two months, never to lose it again.

What could possibly be going on here? The obvious answer is that we have more than one facial recognition device that follow more than one time-course for development. As Johnson and Morton point out, "it [is] difficult for theories which assume a single mechanism to account for developmental U-shaped curves.... It is difficult to see how to account for the pattern of data observed in terms of graduate increases in perceptual abilities or learning" (1991: 36–37).

Several face processors have been isolated now, but only two large divisions need to concern us here. We have first a primitive subcortical processing circuit, available from birth, that is sensitive to the visual features of human faces. It is hard to know exactly what this circuit can process, since a newborn's visual system is still quite shaky. It can distinguish faces from other objects, but maybe not the individual features of particular faces, and it prefers to look at faces over other things. This system is later supplanted or enhanced by a cortical circuit, which comes on-line a few months after birth, and is probably trained up by the output of the more primitive system.

So we have two interacting systems, one of which affects how the other develops. This would explain the strange U-shaped learning curve. We start out being interested in faces and knowing a bit about how to recognize them. Then, as our cortex comes more on-line in our first few months of life, and begins to inhibit the subcortical mechanisms because the errors resulting from them are increasingly recognized, it suppresses and replaces some of the subcortical processes. However, this system doesn't quite know what it is doing yet, so we lose our competence at recognizing faces. But it learns quickly and pretty soon we are better than before at this task.

I submit that our affective systems might work in the same way. We are born with some basic affective circuits. Fear processing in the amygdala is one example; empathy might be another. However, several other circuits come on-line as our brain matures. Pride, shame, and indignation might be examples of these. These cortical circuits learn their trade and are sensitive to social and cultural norms.

This hypothesis would explain why some of our emotions seem so "hard-wired," such as a fear of heights, while other affective reactions mature over time, such as a sense of humor. It would also explain why some affective

reactions are hard to get rid of and why others can be altered fairly easily through talking therapies. Some emotions are grounded in subcortical processes and so are difficult to affect directly through cortical interventions. Others are housed more centrally in cortex and so are less resistant to brain changes. Finally, it explains how it could be that we could start out with the same primitive emotions, yet end up in such different places as adolescents and adults. We do begin with a basic core, but that core is very small, compared to what we get when we have our entire brain trained up and running. Nevertheless, even though our inborn emotions form a small core, they are remain quite powerful, for they remain largely non-cognitive.

4. Dynamic dual systems: A speculative hypothesis

My hypothesis is that we have two sorts of affective systems, one subcortical and non-cognitive and the other cortical and cognitive. These are not to be understood as absolute distinctions between the systems, but rather as differences in emphasis. Even our most primitive system has to react to some interpretation it has assigned to the world, no matter how crude. And even our most cognitive of emotions still has some affective flavor, else it would not be an emotion in the first place.

This hypothesis would explain how those interested in the social construction of emotions and those fascinated by their reduction to primitive circuits can seem to be at once correct yet talking about incompatible things. They are; each is focusing on a different type of affective processing in the human head.

If I am correct, and we have at least two different fundamental types of affective processors in the brain, then our emotions are even more complex than either perspective assumes alone. It is not just that each has only half the story, though each does. It is that both sides leave out how these two different processors can interact with one another. Just as in the case of facial recognition, how one system influences and teaches the other to behave is crucially important.

Perhaps it is best, if I am right, to think of our affective brains as a set of complex resonating circuits, constantly active as we are always reacting to our early, rough-and-ready impressions of the world as well as interpreting and predicting the world around us in a more sophisticated fashion. We have both innate reactions and learned ones and they constantly interact and affect one another. Subcortical neural firing patterns form one part of our overall affective activity, as do firing patterns in cortex. Different affective processors, each operating according to its own rules of engagement, but each interacting with the others (and with other regions in the brain) comprise our emotional self.

The set of circuits will naturally differ across individuals and within the

same individual over time, as we each mature and change in different communities and cultures. We start with a common core of affective reactors, just as we all start with a common face recognizer. Then each brain overlays this core with its own specialized reactors, winnowed and honed by individual experience and expectation, just as each brain overlays the common face recognizer with a specialized one, tuned to the faces in its life. The end products are affective brains very different from one another, each highly idiosyncratic, yet each maintaining a family resemblance with the others.

Without pushing this line too heavily here, I believe a dynamical systems approach gives us a useful approach in theorizing about the complexity of our emotions (see my forthcoming for more discussion). When the brain receives some valenced stimuli, it will seek to interpret that, just as it would any other stimulus. Subcortical activation is part of the meaning-making activity; so is what the cortex does. That is, both the subcortical and cortical areas are part of a larger resonating brain circuit relevant to emotion. These two regions work together to establish a larger coherent and cohesive response to input. Stimuli drop in, so to speak, onto on-going interpretive efforts in a brain shaped by experience and genetic endowment. How the brain has carved up its own circuitry will shape how the emotional stimuli are perceived and reacted to.

No two emotional responses are ever going to be exactly the same, since they are determined by the brain's current resonating activity. Still, family resemblances will exist among them, perhaps as seen in the established 'attractor basins.' Even though strict identity criteria do not hold, our emotions do partake in related activity configurations. Perhaps we can map the firing patterns of our emotional circuits to a trajectory in a multi-dimensional phase space. (Each dimension would represent each variable that determines how the firing pattern might go; the phase space then would be a region that includes all and only the possible activity patterns in the circuit.) If we could create a phase space of possible affective responses by the brain, we should be able to use it to identify regions corresponding to our various emotional categories.

My final highly speculative conjecture is that, given the dependence of the firing patterns on previous activity and the dependence of previous activity on the organisms' complicated and unique environmental history, the trajectories in phase space will most likely be well behaved but chaotic. That is, no activity pattern will precisely duplicate any other pattern, but the patterns will settle into attractor basins in the phase space. (I believe that we will find attractor basins instead of virtually random trajectories because organisms are relatively predictable. The behavioral regularities have to be caused by something other than near random events.)

These chaotic attractors give us a natural way to individuate our physiological or psychological events, for they identify common trends across firing

patterns without entailing that these trends contain isolable defining properties. Assuming we have identified the relevant variables for physiology or psychology, they will indicate how, speaking broadly, the brain interprets incoming stimuli emotionally. The chaotic attractors will show how the brain, in all its complexity, appreciates and understands its world.

Acknowledgments

Versions of this paper were presented to the Departments of Philosophy at the University of Cincinnati, Binghampton University and East Carolina University. I thank the people there for their hospitality and their insightful and stimulating comments. This paper was completed while I was a Taft Fellow at the University of Cincinnati. I thank the University and the Taft Foundation for its generous support.

References

Achenback, T. M., and Howell, C. T. 1993. "Are American Children's Problems Getting Worse? A 13-year Comparison". *Journal of the American Academy of Child and Adolescent Psychiatry* 32: 1145–1154.

Anderson, D., and Pennebaker, J. 1980. "Pain and Pleasure. Alternative Interpretations of Identical Stimuliations". *European Journal of Social Psychology* 10: 207–210.

Burgess, A. W., and Holsmstrom, L. L. 1974. "Rape Trauma Syndrome". *American Journal of Psychiatry* 131: 981–986.

Burgess, A. W., and Holsmstrom, L. L. 1978. "Recovery from Rape and Prior Life Stress". *Research in Nursing and Health* 1: 165–174.

Camras, L. A., Oster, H., Campos, J., Campos, R., Ujiie, T., Miyake, K., Wang, L., and Meng, Z. 1998. "Production of Emotional Facial Expressions in European Americans, Chinese, and Japanese Toddlers". *Developmental Psychology* 34: 616–628.

Chen, X., Hastings, P. D., Rubin, K. H., Chen, H., Cen, G., and Stewart, S. L. 1998. "Child-rearing Attitudes and Behavioral Inhibition in Chinese and Canadian Toddlers: A Cross-cultural Study". *Developmental Psychology* 34: 677–686.

Damon, W. 1999. "The Moral Development of Children". *Scientific American* 281: 72–78.

Ebert, D., Martus, P., and Lungershausen, E. 1995. "Change in Symptomatology of Melancholic Depression over Two Decades. Core Symptoms and Culturally Determined Symptoms". *Psychopathology* 28: 273–280.

Ellis, H. D., and Young, A. W. 1989. "Are Faces Special?" In A. W. Young and H. D. Ellis (eds.), *Handbook on Research on Face Processing*. Amersterdam: North-Holland Press.

Fried, I., Wilson, C. L., *et al.* 1998. "Electric Current Stimulates Laughter". *Nature* 391: 650.

Goren, C. C., Sarty, M., and Wu, P. Y. K. 1975. "Visual Following and Pattern Discrimination of Face-like Stimuli by Newborn Infants". *Pediatrics* 56: 544–549.

Griffiths, P. 1997. *What Emotions Really Are.* Chicago: The University of Chicago Press.
Hardcastle, V. G. 1995. *How to Build a Theory in Cognitive Science.* Albany, New York: SUNY Press.
Hardcastle, V. G. Forthcoming, 2000. "It's O. K. to Be Complicated: The Case of Emotion". *Journal of Consciousness Studies.*
Hohnmann, G. W. 1966. "Some Effects of Spinal Cord Lesions on Experienced Emotional Feelings". *Psychophysiology* 3: 143–156.
James, W. 1890. *The Principles of Psychology.* New York: Holt.
Johnson, M. H., and Morton, J. 1991. *Biology and Cognitive Development: The Case of Face Recogntion.* Cambridge, MA: Basil Blackwell.
Kübler-Ross, E. 1969. *On Death and Dying.* New York: Macmillan.
Lawson, N. C. 1976. *Depression after Spinal Cord Injury: A Multimeasure Longitudinal Study.* Doctoral Dissertation. University of Houston.
LeDoux, J. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life.* New York: Simon and Schuster
Marañon, G. 1924. "Contribution à l'etude de l'action emotive de l'adrenaline". *Revue Francaise d'Endocrinologie* 2: 301–325.
Natterson, J. M., and Knudson, A. G. 1960. "Observations Concerning Fear of Death in Fatally Ill Children and their Mothers". *Seminars in Psychiatry* 22: 456–465.
Nisbett, R. E., and Schachter, S. 1966. "Cognitive Manipulation of Pain". *Journal of Experimental Social Psychology* 2: 227–236.
Oltjenbruns, K. A. 1998. "Ethnicity and the Grief Response: Mexican American versus Anglo American College Students". *Death Studies* 22: 141–155.
Pang, K. Y. 1998. "Symptoms of Depression in Elderly Korean Immigrants: Narration and the Healing Process". *Culture and Medical Psychiatry* 22: 93–122.
Parkes, C. M. 1970. "The First Year of Bereavement: A Longitudinal Study of the Reactions of London Widows to the Death of their Husbands". *Psychiatry* 33: 444–467.
Parkes, C. M. 1975. "Unexpected and Untimely Bereavement: A Statistical Study of Young Boston Widows and Widowers". In B. B. Schoenberg, I. Gerber, A. Weiner, A. H., Kutscher, D. Perez, and A. C. Carr (eds.), *Bereavement: Its Psychosocial Aspects.* New York: Columbia University Press.
Russell, J. A. 1995. "Facial Expression of Emotion: What Lies Beyond Minimal Universality?" *Psychology Bulletin* 118: 379–391.
Shioriri, T., Someya, T., Helmeste, D., and Tang, S. W. 1999. "Misinterpretation of Facial Expression: A Cross-cultural Study". *Psychiatry and Clinical Neuroscience* 53: 45–50.
Vygotsky, L. S. 1962. *Mind and Language.* (Translated and edited by E. Hanfmann and G. Vakar.) Cambridge, MA: The MIT Press.
Vygotsky, L. S. 1978. *Mind in society: The development of higher psychological processes.* Edited by M. Cole. Cambridge, MA: Harvard University Press.

CHAPTER 7

The Effect of Motivation on the Stream of Consciousness

Generalizing from a Neurocomputational Model of Cingulo-frontal Circuits Controlling Saccadic Eye Movements

Marica Bernstein, Samantha Stiehl and John Bickle
Focused Research Program in Computational Neuroscience

1. Introduction

The intact adult human brain processes external information simultaneously along multiple pathways and through at least five anatomically distinct but distributed networks dedicated to spatial awareness, language, explicit memory, object recognition, or working memory (Mesulam 1998). However, cognitive output — selective sensory attention, speech production, language comprehension, and working memory — is sequential. For example, psychologist Bernie Baars (1997) writes, "[a]ll working memories operate serially, one thing at a time. They show a stream of events one at a time." Consciousness itself arises from the parallel processing of vast amounts of information bombarding the brain, but it is a *single stream* that we experience. To further complicate the issue of parallel processing and serial output, cognitive processes and consciousness are influenced by one's own internal state of affairs: by our recollections, our fears, our favorite color, the song we learned last week and who we are expecting for dinner tonight. In this chapter, we present a neurocomputational model that accounts for the sequential and affective features of some cognitive and conscious processes.

In his (1890) *Principles of Psychology* William James eloquently described sequential features of cognitive processing and the stream of consciousness:

> Into the awareness of the thunder itself the awareness of the previous silence creeps and continues; for what we hear when the thunder crashes is not thunder *pure*, but thunder-breaking-upon-silence-and-contrasting-with-it. Our feeling of the same objective thunder, coming in this way, is quite different from what it would be were the thunder a continuation of previous thunder.... [I]t would be difficult to find in the actual concrete consciousness of man a feeling so limited to the present as not to have an inkling of anything that went before (240–241).

According to James, conscious streams (1) are extended through time and (2) typically proceed in orderly fashion from one unified representation to another (*"thunder-breaking-upon-silence-and-contrasting-with-it"*). Also (3) the contents of later events depend upon and are affected by the contents of earlier events in the stream (*"Into the awareness of the thunder itself the awareness of the previous silence creeps and continues"*). These features generalize across presentation modalities and types of conscious experience (*"[I]t would be difficult to find in the actual concrete consciousness of man a feeling so limited to the present as not to have an inkling of anything that went before"*). Contemporary philosopher Daniel Dennett (1991) notes another feature: (4) Temporal limits often require that multiple steps in the sequence be computed in advance with little or no time for revision once the sequence is initiated. Many cognitive and conscious sequences are thus 'semi-ballistic' — to a certain extent *not* adjustable because there is no time for 'reflective feedback' once a sequence has begun. Consider quickly scrolling through a long list of names searching for your own. You often overshoot the mark by two or three lines before you are aware that "There it is!"

Concerning these sequential features of cognition and consciousness, we agree with Crick and Koch (1998): "[m]ost of the philosophical aspects of the problem [of consciousness] should, for the moment, be left on one side, and ... the time to start the scientific attack is now." Our methods are those of computational *neuro*science. We have previously proposed a biologically plausible neurocomputational model that accounts for these sequential features (Bickle *et al.* in press; Bernstein *et al.* 1999; Bickle *et al.* 1998; Bickle *et al.* 1997). Our model generalizes from known properties and connectivities of cells in the primate saccade generating system and associated prefrontal areas. Although saccade generation is not a "cognitive" process, the output of this system shares with higher order cognitive processes, and with the stream of consciousness, many of the sequential features listed above. We make the common assumption that there is "a basic common [neural] mechanism, or perhaps a few such [neural] mechanisms" underlying these shared features (Crick and Koch 1998; Crick and Koch 1990). Neurophysiological research has uncovered a great deal about the activity and properties of cells in the primate saccade generating system and the biological basis for our model is the way these cells and circuits

implement a powerful computational operation. Understanding this biological system may reveal a common neural mechanism for sequential processing.

However, this general model does not directly address one additional feature of sequential and conscious processes: *internal* motivations can contribute to determining the sequence of cognitive outputs (and the stream of consciousness) irrespective of the *surface* properties of individual points in the sequence. Again, we defer to James (1980) for a delightful description of this feature in sesory attention, "[t]hese things may be called the motives of attention.... A faint tap *per se* is not an interesting sound.... But when it is a signal, as that of a lover on the window-pane, it will hardly go unperceived." We are often motivated to attend to a particular stimulus. At other times, when we are otherwise motivated, that same stimulus is merely one in a seemingly endless string, none commanding more than fleeting attention, each disappearing, a faint forgotten tap. This feature is evident in other cognitive processes. Some sequences of linguistic utterances are filled with words 'purposefully' chosen, other times we mutter. Sometimes our thoughts blaze a trail toward a goal, other times they wander. How does this cognitive-affective "fusing" happen?

Again, we believe that this question is ready for neuroscientific investigation. Here we propose a neurobiologically plausible computational mechanism for the production of *motivated* sequences. In the next section we justify our choice of methods. In Section 3 we give an overview of our general neurocomputational model of sequential processing and its biological basis. In Section 4 we turn to the question of motivated sequences. We enumerate, from a computational perspective, the properties a system must possess to "fuse" internal motivation with sequences run off semi-ballistically. We then present cell properties and circuitries of anterior cingulate cortex (a region of the limbic system). These provide the biological basis for computational features of our extended neurocomputational model. In the final section we step back from the neurobiological detail to discuss implications of our model for the phenomenology of consciousness, cognition and affect.

2. Methods

If our objective is to understand *neural* mechanisms, why are we *computational* modelers? In the past few decades, neurochemists, neurophysiologists, molecular geneticists, neuroanatomists, neuropsychologists, behavioral pharmacologists, cognitive psychologists and others have provided a plethora of experimental data detailing everything from the way neural-glial signaling pathways regulate axonal membrane ion channels to the effects of drugs, lesions and knockout genes on behavior. Nonetheless, while the collection of neuroscientific detail proceeds at

an astonishing rate, neuroscientists admit that a completely neuromechanistic story of even one complex primate behavior is a long way off. (E.g., see the chapters by Newsome, Gallistel, and Carmazzo in Gazzaniga 1997) However, properly constrained, computational models can synthesize lower level (neuromolecular) detail with higher level (psychological) observation. According to James Bower and David Beeman (1995), developers of the popular neurocomputing software *GENESIS* (GEneral NEural SImulation System), computational neuroscience provides "a mechanism for generating new ideas based on the anatomy and physiology of the circuits themselves". The 'new ideas' include not only testable predictions about cell-biological and molecular mechanisms but also new ways of understanding how billions of neurons and their trillions of connections produce complex behavior. (See also Churchland and Sejnowski 1992.)

If parameters of the computational model are constrained by actual cell properties and connectivities of the system under investigation and cohere with data from imaging, behavioral, clinical studies and the like, the results are (1) a *biologically plausible* model of the specific system; (2) precise testable predictions about lesser known cell properties and connectivities that can guide continuing neurophysiological research; and (3) a *biologically implemented* model that can generalize to other systems or behaviors with similar features. Models that begin with and replicate to the deepest extent possible the "structure of the nervous system as a basis for exploring its computational features are more likely to uncover features which had been previously overlooked or unsuspected" (Bower and Beeman 1995).

One promising theoretical resource of computational neuroscience comes from the mathematics of dynamical systems. This resource characterizes neural representations as points (or sub-volumes) in high-dimensional vector spaces and neural computations as vector-to-vector transformations (Churchland and Sejnowski 1992). Figures 1a–1e illustrate several ways that this resource has been fruitfully applied. This approach provides natural mathematical interpretations of activity patterns and dynamics in neural networks, both biological and artificial. Individually, neurons are limited capacity computers. They compute their total input, their new activation states as a (nonlinear) function of their total inputs and previous activation states, and their total output rates from their new activation states (Figures 2a and 2b). Output rates reflect signal strengths serving as input to other neurons sharing active synapses. In biological networks, signal strength reflects a variable such as action potential frequency. Synaptic weight values reflect pre- and post-synaptic variables affecting neural transmission (e.g., amount of neurotransmitter released, availability of re-uptake enzymes, number of active receptors, presence or absence of neurohormones, etc.). Strung together in parallel networks, where patterns and strengths of synaptic connections determine a network's computational architecture, these limited capacity computers become

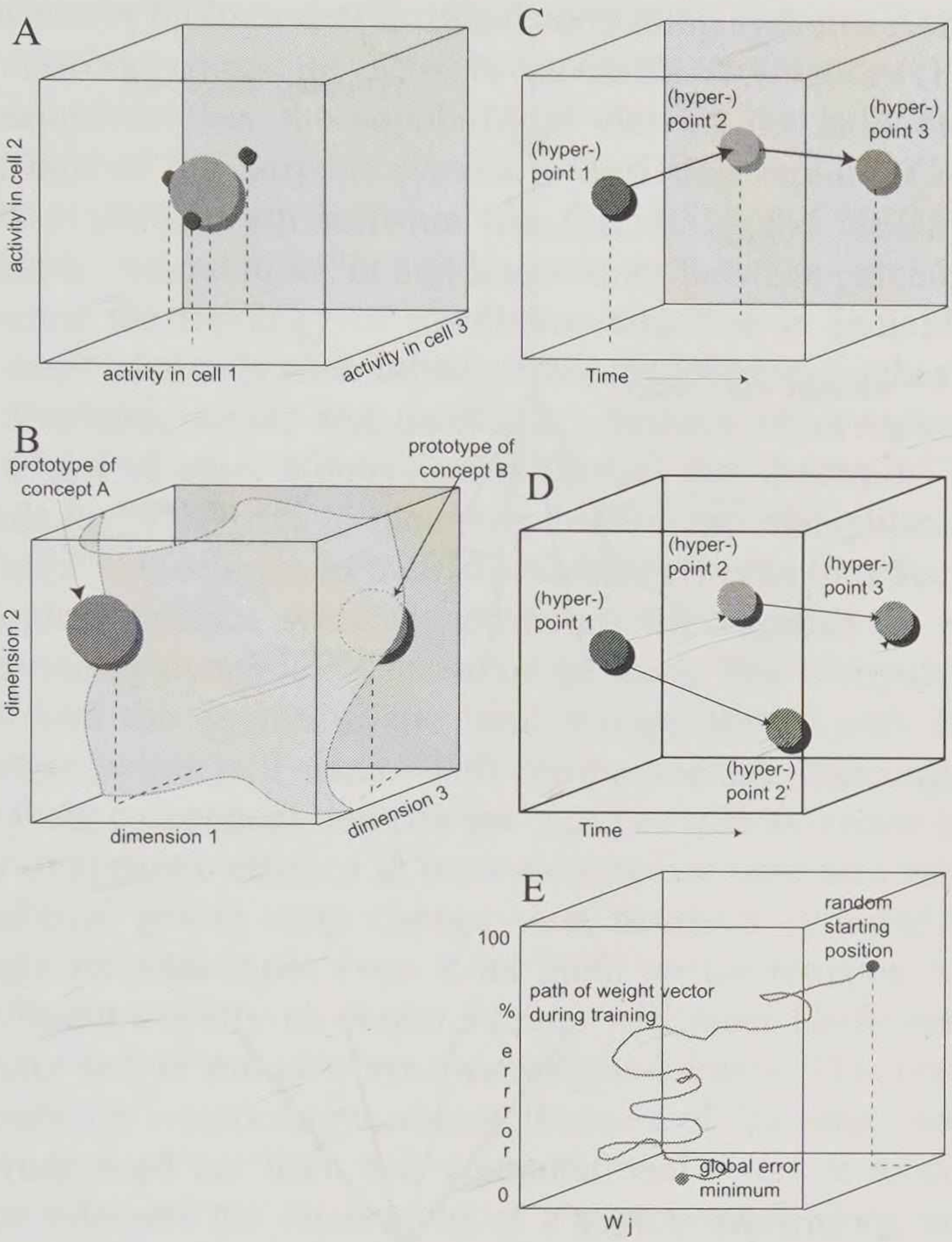


Figure 1. *The computational power of vector representations and operations*

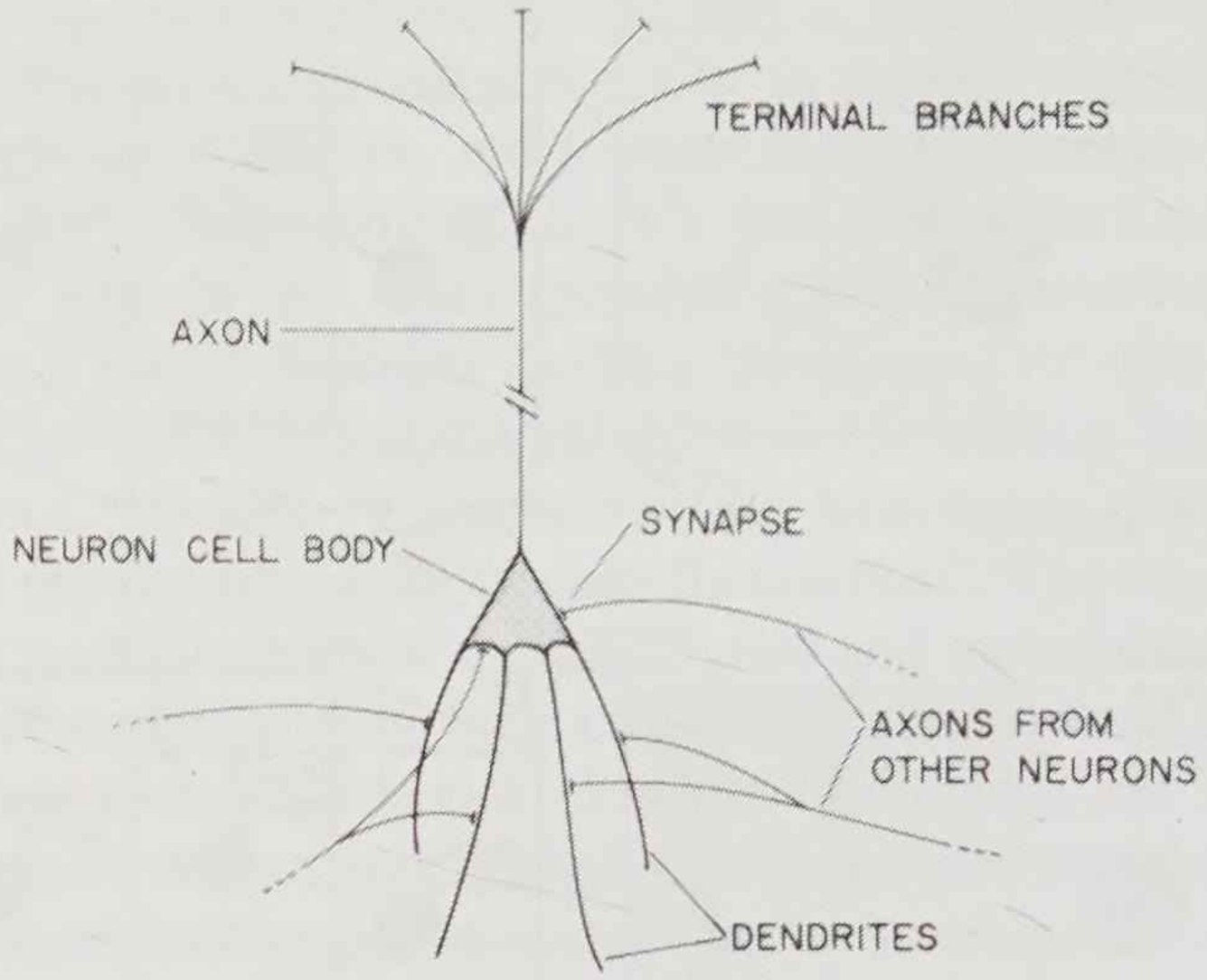
A. Concepts represented as prototypes (large subvolumne) and instances (points) in a unit-activation vector space. Prototypes constitute a set of network activity patterns generated by appropriate input vectors. Distance in the space between a point and the center of the prototype provides a measure of similarity.

B. Partitioning of concepts in a state space. (Adapted from Churchland 1989.)

C. A sequential process in vector space. The neural network implementing the process proceeded from the activity pattern at (hyper-) point 1, through those at (hyper-) points 2 and 3. Each (hyper-) point reflects the activity pattern across the network representing a cognitive content in the appropriate high-dimensional vector space. (Adapted from Bickle *et al.* in press.)

D. Alternative "paths" through vector space. As in C a sequential process (e.g., problem solving) is represented as a trajectory through vector space. Here two possible outputs result from activity at (hyper-) point 1, but the output of activity patterns at (hyper-) points 2 and 2′ both result in a pattern of activity represented by (hyper-) point 3. Changes in synaptic weights can produce alternative outputs, see text for discussion. (Adapted from Bickle *et al.* in press.)

E. Learning as a gradient decent through synaptic weight-error vector space. (Adapted from Churchland 1989.)



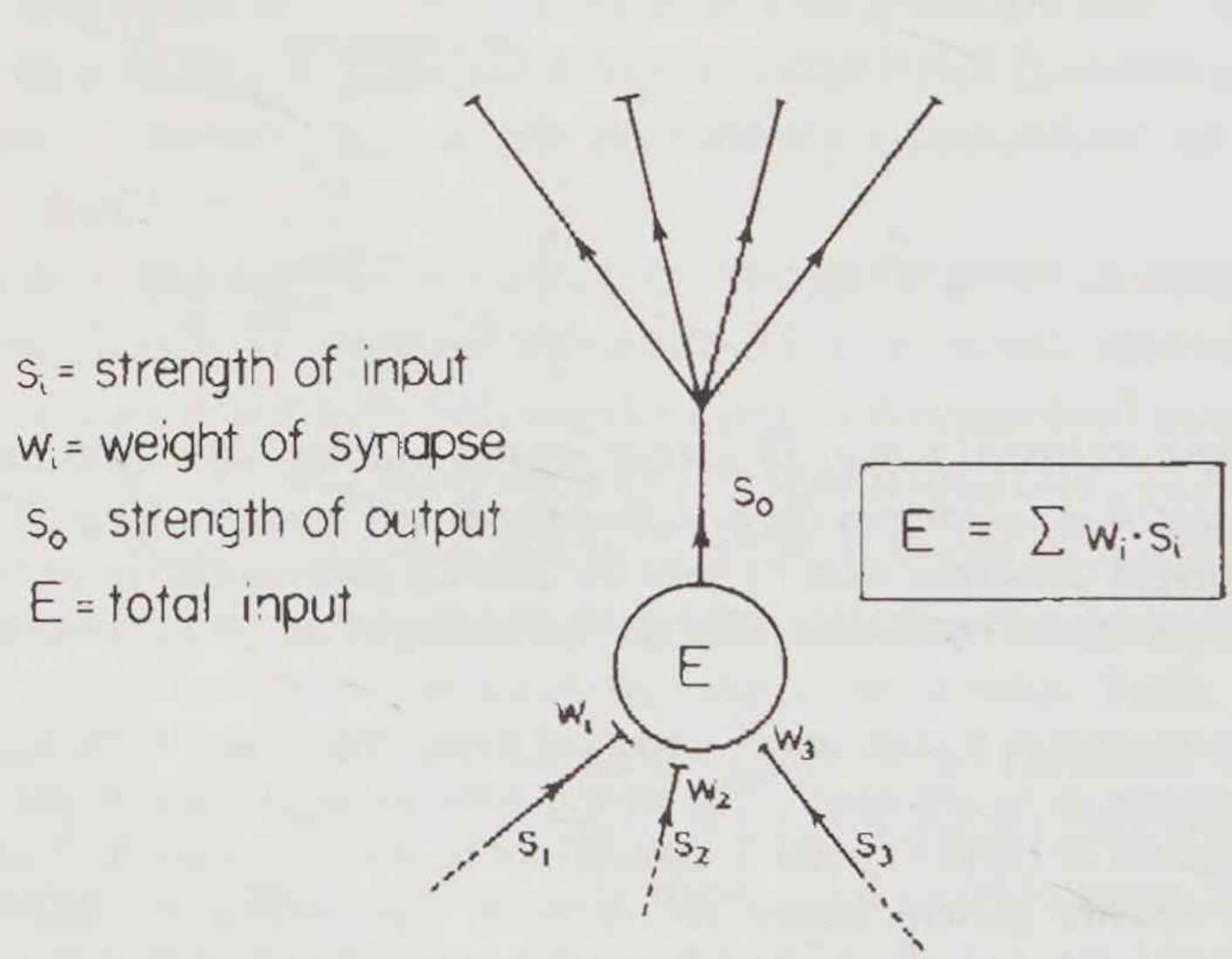


Figure 2. *Neurons as "limited capacity computers"*
A. Schematic illustration of neuron
B. Computational features of a neuron's activity (Both figures adapted from Churchland 1989.)

components of power biological or artificial computing systems. Activity patterns across neural networks constitute vector-to-vector transformations (Figures 3a, b).

Among *neuro*scientists, the popularity of viewing the individual neuron as the functional unit of the nervous system is declining rapidly. *Compartmental modeling*, now available with software like GENESIS and NEURON, enables modelers to mimic computations in and interactions between patches of neuronal membrane. Hence the topology of membrane structure in individual neurons, variations in ionic channels distributed across membrane patches, location of synapses on dendrites, soma, and axon (i.e., distance of synapses from axon hillock), and a host of other biophysical properties that determine frequency of action potentials are accessible to modeling control and manipulation. Modelers can 'custom-build' the neurons in their simulated networks to reflect any desired level of biological realism without sacrificing the capacity to study circuit properties and activity dynamics of the entire network. Few computational *neuro*-scientists still treat the neuron as the basic computational unit. However, the above interpretive points still stand. With compartmental modeling, both neural networks and their component neurons are interpretable as vector transformers.

Plasticity of synaptic efficacy in *biological* neural networks vastly increases their computational power. Any change to a neuron's synaptic weights will typically change its total input even if its input vector remains constant. This means that different activity or output vectors will most likely result from the same input vector before and after synaptic weight changes. This feature provides a powerful basis for neurocomputational theories of learning (see again Figure 1e). Theorists need not limit this computational resource to *activity* vector spaces. We can interpret the dimensions of a vector space along any biological or representational parameters. (See e.g., Churchland's (1995) discussion of Susan Brennan's interesting application of this approach to face recognition.) In addition, cellular mechanisms of real neurons governing action potential rate implement extensively nonlinear functions. Hence, biological networks interpreted as vector transformers can compute nonlinear input-output functions, dramatically broadening their computational range. Finally, biological networks also implement a variety of computational architectures, employing layers, columns, feedback, and a multiplicity of excitatory, inhibitory and modulatory synaptic connections.

Given these features, the vectorial account of the representational and computational power of the primate brain is inspiring. Paul Churchland (1989) elaborates some possibilities:

> Given high dimensional vector spaces, which the brain has in abundance, those spaces and prototypes they embody can encompass categories of great complexity, generality, and abstraction including those with a temporal dimension such as harmonic oscillator, projectile, traveling wave, Samba, twelve-bar blues, democratic election, six-course dinner, courtship, elephant hunt, civil

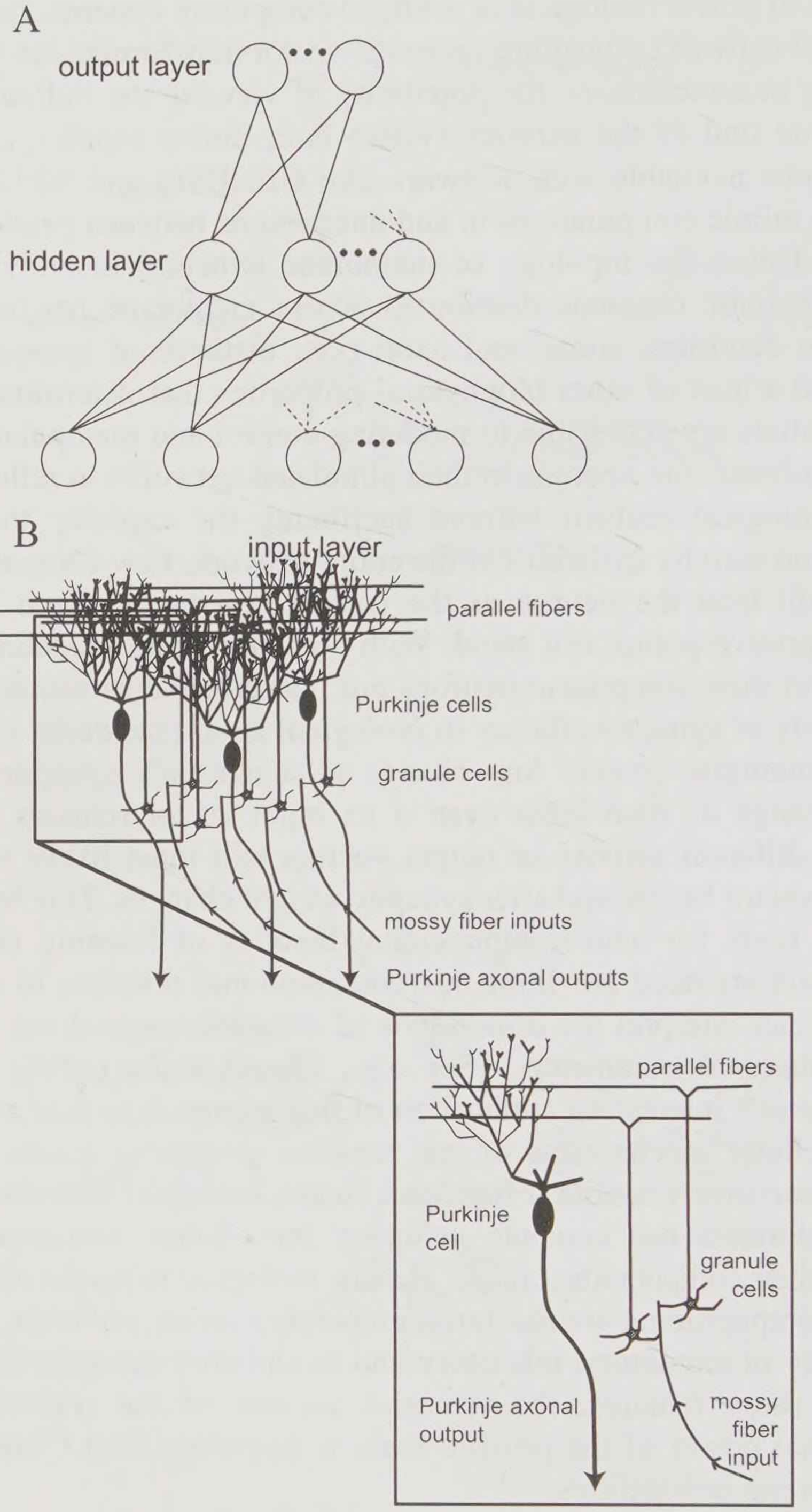


Figure 3. *Activity patterns across neural networks*

A. Schematic illustration of an artificial three-layered feedforward neural network.

B. Schematic illustration of a cerebellar network. Insert reveals the detailed structure of a single Purkinji cell.

> disobedience and stellar collapse.... In principle, then, it is no harder for such a system to represent types of processes, procedures and techniques, then to represent the "simple" sensory qualities. From the point of view of the brain, these are just more high-dimensional vectors (191).

Consider a single example. The position of the left hand and fingers, together with the movement of the right arm that combine to produce a particular note while playing a cello can be represented as a (hyper-) point or subvolume in a finger-position-and-arm-angle vector space. The *sequence* of finger positions and arm movements that generate a melody is a path through (hyper-) points or subvolumes in that vector space. Any sequence of representations that can be expressed as a path through points or subvolumes in a vector space is a process that can be implemented in an appropriately structured biological neural network. Given the known structure of the primate brain, the possibilities are vast.

What operations *compute* paths through vector spaces? Consider Figure 4a. Here, in a two-dimensional Cartesian space, is the path from a point of origin, **O**, ($\langle 0, 0 \rangle$), through **A** at $\langle 3, 10 \rangle$ and **B** at $\langle 12, 8 \rangle$. We can compute the dimensions of the second vector in the sequence (**A** → **B**) using the dimensions of the first vector (**O** → **A**) and of the second target relative to **O** (**O** → **B**):

$$\mathbf{O} \rightarrow \mathbf{B} = \mathbf{O} \rightarrow \mathbf{A} + \mathbf{A} \rightarrow \mathbf{B}$$

Rearranging and solving this equation yields

$$\begin{aligned} \mathbf{A} \rightarrow \mathbf{B} &= \mathbf{O} \rightarrow \mathbf{B} - \mathbf{O} \rightarrow \mathbf{A} \\ &= \langle 12, 8 \rangle - \langle 3, 10 \rangle \\ &= \langle 9, -2 \rangle \end{aligned}$$

These are exactly the dimensions necessary to get to **B** from **A**. Notice that the dimensions of the step **A** → **B** are very different from the step to **B** directly from **O**.

The operation here, *vector subtraction*, is an iterative operation. Its equation takes the general form:

$$\mathbf{(N\text{-}1)} \rightarrow \mathbf{N} = \mathbf{(O} \rightarrow \mathbf{N)} - \mathbf{(N\text{-}2} \rightarrow \mathbf{N\text{-}1)} - \ldots - \mathbf{(O} \rightarrow \mathbf{A)}$$

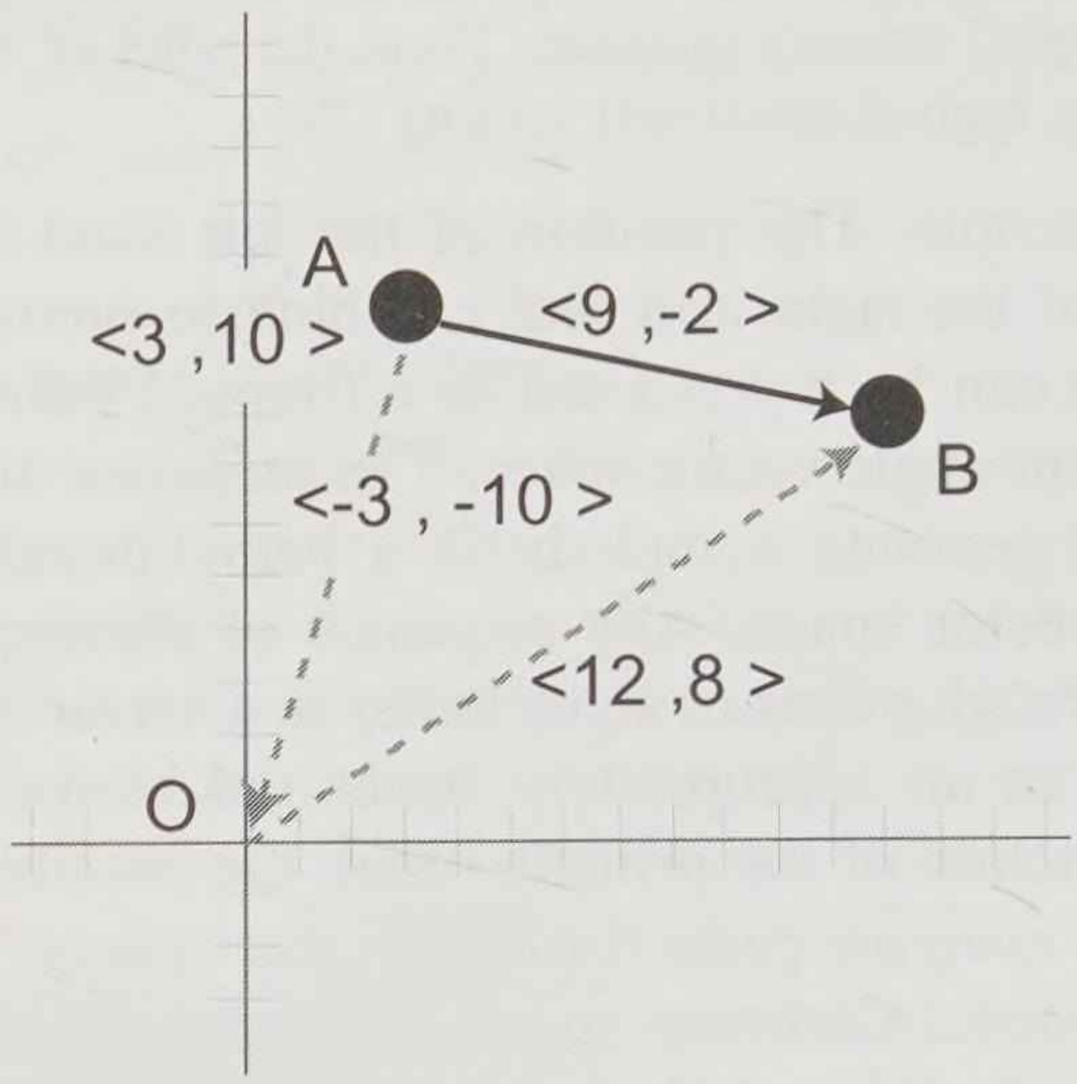
Referring to Figure 4b, we can compute the dimensions from **B** to **C** ($\langle 5, -4 \rangle$):

$$\begin{aligned} \mathbf{B} \rightarrow \mathbf{C} &= \mathbf{(O} \rightarrow \mathbf{C)} - \mathbf{(A} \rightarrow \mathbf{B)} - \mathbf{(O} \rightarrow \mathbf{A)} \\ &= \langle 5, -4 \rangle - \langle 9, -2 \rangle - \langle 3, 10 \rangle \\ &= \langle -7, -12 \rangle \end{aligned}$$

These are exactly the dimensions necessary.

Vector subtraction is thus an operation that computes sequential paths through vector spaces. Since vector subtraction uses only information about previous steps in the sequence and the location of the target relative to the origin,

A



B

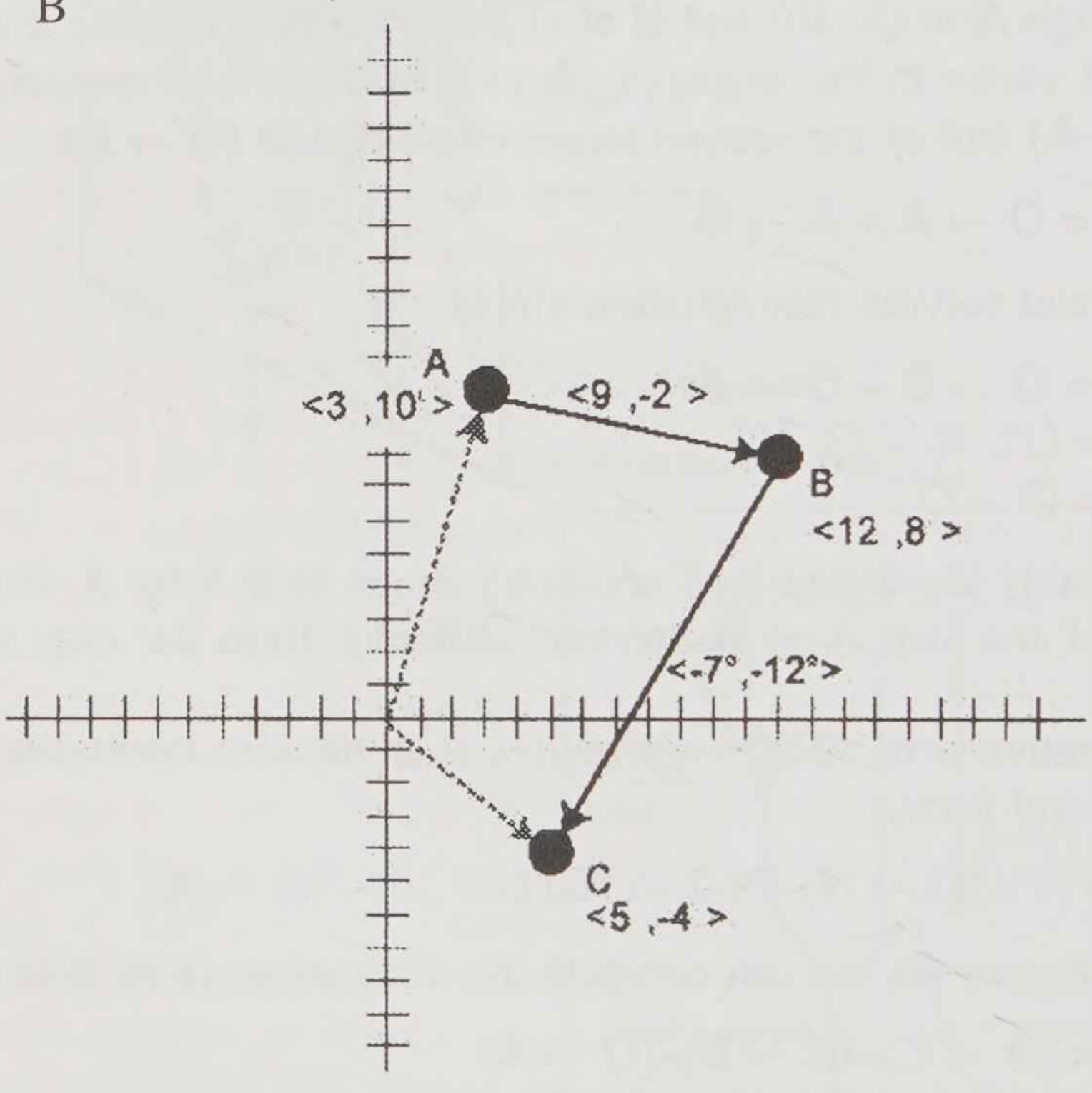


Figure 4. *Computing paths through vector spaces*

A. The dimensions of the path from a point of origin, O, through points A and B are computed using vector subtraction (see text). Solid line reflects vector between points as computed by this operation. Dashed lines reflect vector distances between origin and points.

B. Vector subtraction iterates to compute the dimensions of a path through *n* points as described in the text.

it can *compute* multi-step sequences through the space in advance and *execute* ballistically (without explicit 'feedback') the results of those computations. It uses the contents of earlier representations in the sequence (the dimensions of points or subvolumes in the vector space) to compute the dimensions of later steps. Within this vectorial approach to cognitive neuroscientific theorizing, vector subtraction is an operation that generates sequences possessing the features of cognitive processes and the "Jamesian" stream of consciousness.

However, biological plausibility demands more than a reinterpretation of biological systems in terms of a pre-existing theoretical model. We must examine biological systems to determine the computational operations implemented in real brains. Is there any evidence that vector subtraction is implemented neurally to generate sequential outputs? There is, and so we begin the next section with an overview of the properties and connectivities of cells in primate frontal eye fields.

3. A neurocomputational model of sequential processing

Previously we developed a neurocomputational model of how the brain generates sequential processes characteristic of higher cognition and consciousness (Bickle *et al*. in press). Our model generates output with the sequential features we listed in the introduction. Our methodology was to find a neural system that generates sequential outputs possessing these features, and whose cell properties and anatomical connectivities were known. The model neural system itself need not be involved in cognitive or conscious processes, but its basic anatomy and physiology had to be generalizable to other neural systems that are. Our guiding idea was that if we could understand how the model system computes its sequential outputs, we could predict the detailed cellular mechanisms by which other neural regions compute sequential cognitive and conscious processes. These predictions would be testable by anatomical and physiological investigation. The vector space characterization of neural representation and computation provided the necessary bridge between the neuroscientific detail and our computational model of cognitive and conscious processing.

We chose one component of the primate *saccade command system* as our model neural system. Saccades are a type of eye movement that locates the fovea, the area of the retina with the greatest visual acuity, on key aspects of visual stimuli. Humans and other primates saccade on average 3–5 times per second. Most saccades are 'involuntary,' but we can control saccades voluntarily and consciously. Non-visual factors can also initiate saccades, including noises, verbal commands, and memories (Goldberg *et al*. 1992).

The primate saccade command system involves frontal and posterior parietal cortical regions, the midbrain superior colliculi, burst circuits in mesencephalic

and pontine brain stem nuclei, the corpus striatum, and the substantia nigra pars reticulata (Goldberg *et al.* 1992). The areas most important for purposive, intentional saccades are the frontal eye fields (FEFs) in premotor (frontal) cortex (encompassing primarily Brodmann's area 8a). Over the past fifteen years, neuroscientists Michael Goldberg and Charles Bruce have provided a detailed account of FEFs' role in saccade control. FEFs receive retinotopically-coded visual information from the dorsal visual stream and project efferents to the midbrain superior colliculi for relay to the brainstem burst circuits that drive eye muscle motor neurons.

In an early study Bruce and Goldberg (1985) discovered different types of *presaccadic* neurons in primate FEF. These fire 50–200 milliseconds before saccade initiation. Each presaccadic visuomovement and movement neuron (two of the types they distinguished) had a preferred *movement field*, firing optimally (greatest number of action potentials per second) before saccades of a single amplitude and direction, near-optimally before saccades with similar parameters, less optimally before less similar saccades, and only at baseline rate before significantly different saccades. Many presaccadic FEF neurons are thus active before a single saccade. The dimensions of the particular saccade for which these cells code is apparently a *weighted vector average* of all active cells: the average of cells' preferred movement field times activity rate. Previously, Lee *et al.* (1988) had developed and experimentally verified this population coding concept for eye movement in superior colliculus, and Georgapoulos *et al.* (1986) had done so for arm movement in primate primary motor cortex. It has since become a key resource in computational neuroscience. Outputs of these FEF circuits bound for superior colliculi are coded not in retinotopic (visual) coordinates, but rather in oculomotor (eye movement) coordinates. Bruce and Goldberg (1985) elegantly demonstrated this using a *double-step saccade paradigm.* (See Bickle *et al.* in press for a nontechnical discussion.) Hence, the presaccadic activity in primate FEFs perform ocular-motor coordinate transformations, changing retinotopically-coded visual input into sequences of eye movement (motor)-coded commands.

In a later study, Goldberg and Bruce (1990) discovered *postsaccadic* FEF neurons. These began firing soon after a saccade was initiated. These neurons also had preferred movement fields. About one-third of the presaccadic neurons studied also had postsaccadic activity. The relationship between the preferred pre- and postsaccadic movement fields in these dually active cells was intriguing. A cell's optimal postsaccadic activity followed saccades *of exactly the opposite amplitude and direction* as its optimal presaccadic activity. Taken together, these pre- and postsaccadic messages provided evidence that the FEF uses *vector subtraction* to compute sequences of saccades (Goldberg and Bruce 1990).

Based on the computational power of vector subtraction and the details of

its neural implementation in primate FEFs, we have constructed a neurocomputational model of sequential processing with the following components (Bickle *et al.* in press). The *Vector Subtraction Core* exactly mimics the pre- and postsaccadic activity in FEF neurons (see Figure 5). *Next Target Nodes* represent the dimensions of the next target in the sequence from the initial origin. *Previous Step Nodes* represent the opposite dimensions of the step just executed (in keeping with the postsaccadic activity in FEFs). *Next Step Nodes* sum these values to compute the dimensions of the next step through the vector space. In a computer simulation we are developing, the dimensions represented in these nodes are weighted vector averages of the preferred movement dimensions (through the particular vector space at issue) of all active cells.

To enable the system to compute sequences greater than two steps, we supplement the Vector Subtraction Core with a *Working Memory Store* (see Figure 6). Dimensions of Previous Step Nodes move directly to the first 'level' of the Working Memory Store. They continue to move up the 'levels' after each subsequent step. The number of levels in the Store reflects the system's duration and limits of working memory. (We assume that different neural systems subserving different cognitive and conscious processes will have Working Memory Stores with different durations and limits.) The dimensions of each layer in the Working Memory Store project to the Next Step Nodes. Along with values from Next Target Nodes and Previous Step Nodes, they are used to compute the dimensions of the next step. We designed specific details of the Working Memory Store to mimic cell properties and connectivities in primate dorsolateral prefrontal cortex (DLPC), primarily neurons with 'working memory fields' in area 46 (Funahashi *et al.* 1993; Goldman-Rakic 1996). In a computer simulation we are developing, dimensions in each layer are weighted vector averages of the "working memory field" (movements through the vector space at issue) of all active cells.

Although our model can compute multiple steps through a vector space before sending commands to initiate the first one, it must also be able to interrupt sequences mid-stream and return to the initial origin. We possess this capacity for many cognitive and conscious processes. While driving down a boring stretch, you are distracted by an attractive roadside billboard. Quickly you regain your focus on the road ahead, although it contains no salient stimuli. Based on some cell properties and connectivities in FEF 'suppression sites' (Burman and Bruce 1997), we introduce a *Return to Origin* mechanism into our model (see Figure 7). When activated, this mechanism inhibits input to the Next Step Nodes from the Next Target Nodes. This amounts to adding a $\langle 0, 0, \ldots, 0 \rangle$ message to the values from Previous Step Nodes and Working Memory Store layers. Summing this information yields a vector with exactly the dimensions necessary to return the system to its initial origin from the last location it occupied in the

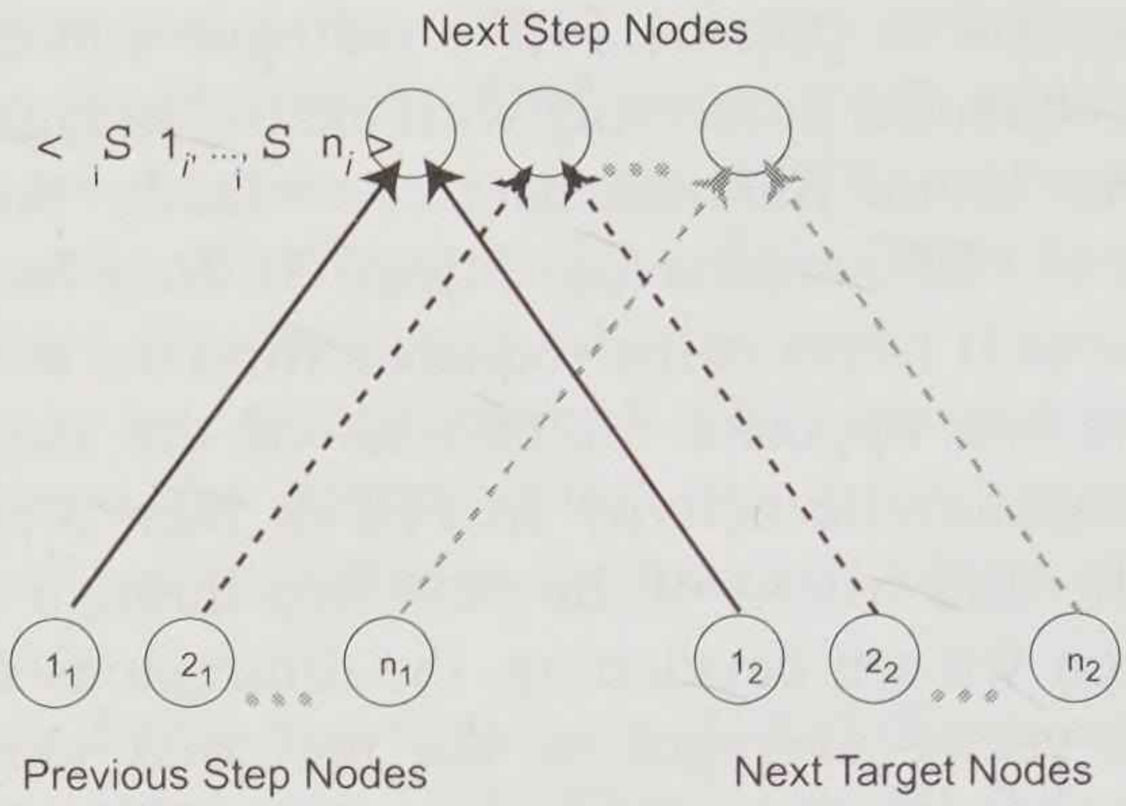


Figure 5. *Vector subtraction core*
Processing in the vector subtraction core generalizes from the neural implementation of vector subtraction in primate FEFs. (Adapted from Bickle *et al.* in press.)

vector space. The Return to Origin mechanism is under *Variable Threshold* control, tied to some level of the Working Memory Store. The biological model for, and neurocomputational details of, the Variable Threshold are our topic in the next section.

The activity of our full model is illustrated in Figure 8. The Next Step Nodes first compute the dimensions of **O** → **A** from the Next Target Nodes only:

$$\mathbf{O} \rightarrow \mathbf{A} = \langle 3, 10 \rangle$$

(No other nodes are active at this stage.) The dimensions of **A** → **B** are then computed from the dimensions of the vector exactly opposite the one just executed, now in the Previous Step Nodes, and the location of the next target (in Next Target Nodes):

$$\begin{aligned} \mathbf{A} \rightarrow \mathbf{B} &= \langle -3, -10 \rangle + \langle 12, 8 \rangle \\ &= \langle 9, -2 \rangle \end{aligned}$$

Dimensions that most recently occupied the Previous Step Nodes in the Vector Subtraction Core now occupy the first layer of the Working Memory Store. Appropriate dimensions about the step just executed occupy the Previous Step Nodes. Location (relative to origin) of the next target occupy the Next Target Nodes. The Next Step Nodes compute the dimensions of the next step in the sequence:

$$\begin{aligned} \mathbf{B} \rightarrow \mathbf{C} &= \langle -3, -10 \rangle + \langle -9, 2 \rangle + \langle 5, -4 \rangle \\ &= \langle -7, -12 \rangle \end{aligned}$$

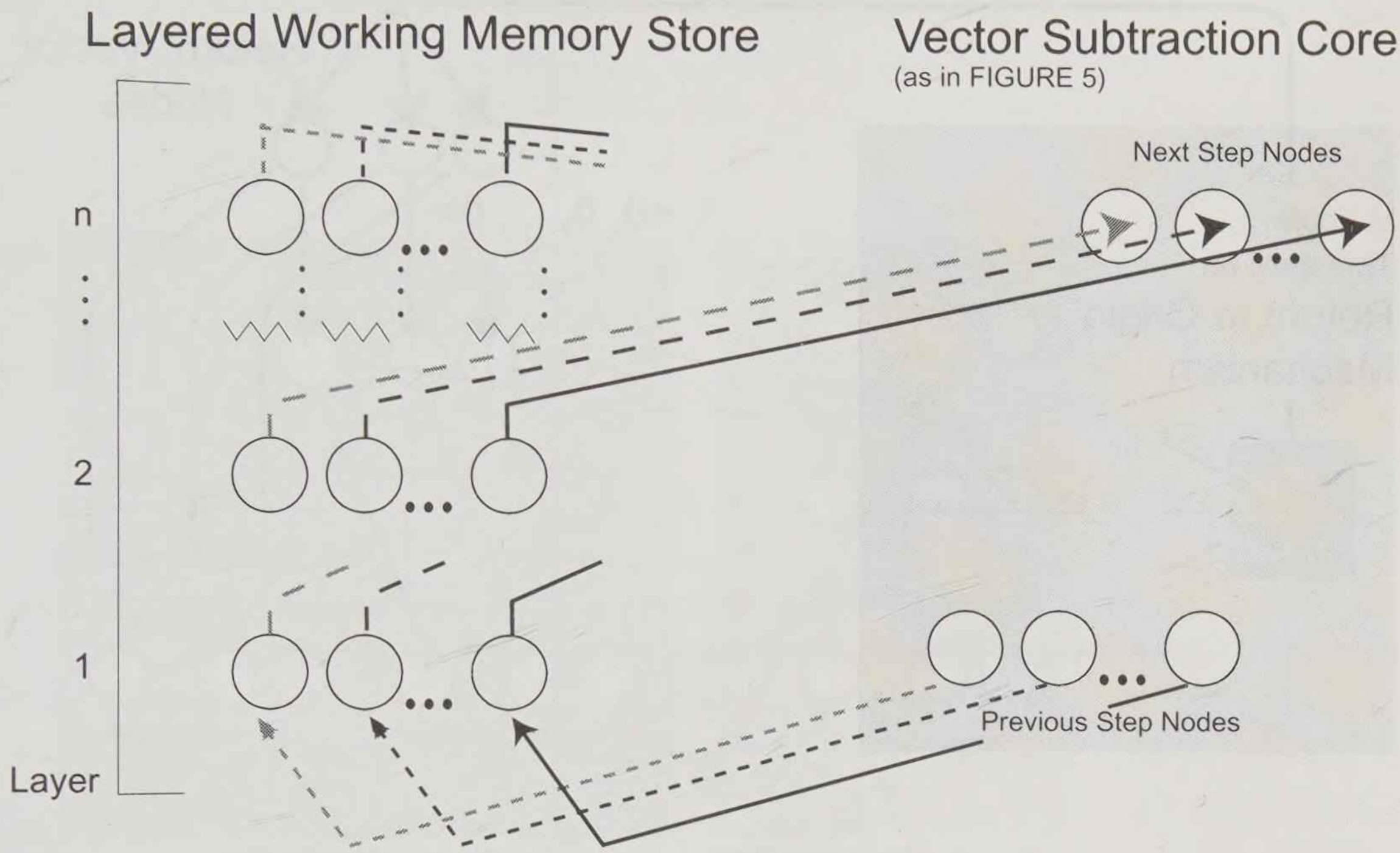


Figure 6. *Working memory store*
Layers hold dimensions of successive previous steps (as they are projected from Previous Step Nodes of Vector Subtraction Core) and reflect temporal limits and storage capacity for the cognitive task and neural network. Each layer continues to project the dimensions of previous steps to Next Step Nodes. Vector Subtraction Core remains as in Figure 5; only components relevant to Working Memory Store are shown here. (Adapted from Bickle *et al.* in press.)

Assuming that the Variable Threshold on the Return to Origin mechanism is set higher than the second layer in the Working Memory Store, the system will then compute the dimensions of $\mathbf{C} \rightarrow \mathbf{D}$:

$$\begin{aligned}\mathbf{C} \rightarrow \mathbf{D} &= \langle -3, -10\rangle + \langle -9, 2\rangle + \langle 7, 12\rangle + \langle -4, -8\rangle \\ &= \langle -9, -4\rangle\end{aligned}$$

But if the Variable Threshold is set at the second layer of the Working Memory Store, its activation inhibits the Next Target input to the Next Step Nodes, in essence adding a $\langle 0, 0, \ldots, 0\rangle$ vector to the inputs from the Working Memory Store and the Previous Step Nodes. The result computed in the Next Step Nodes is the dimensions that return the system to its initial origin from **C**:

$$\begin{aligned}\mathbf{C} \rightarrow \mathbf{O} &= \langle -3, -10\rangle + \langle -9, 2\rangle + \langle 7, 12\rangle + \langle 0, 0\rangle \\ &= \langle -5, 4\rangle\end{aligned}$$

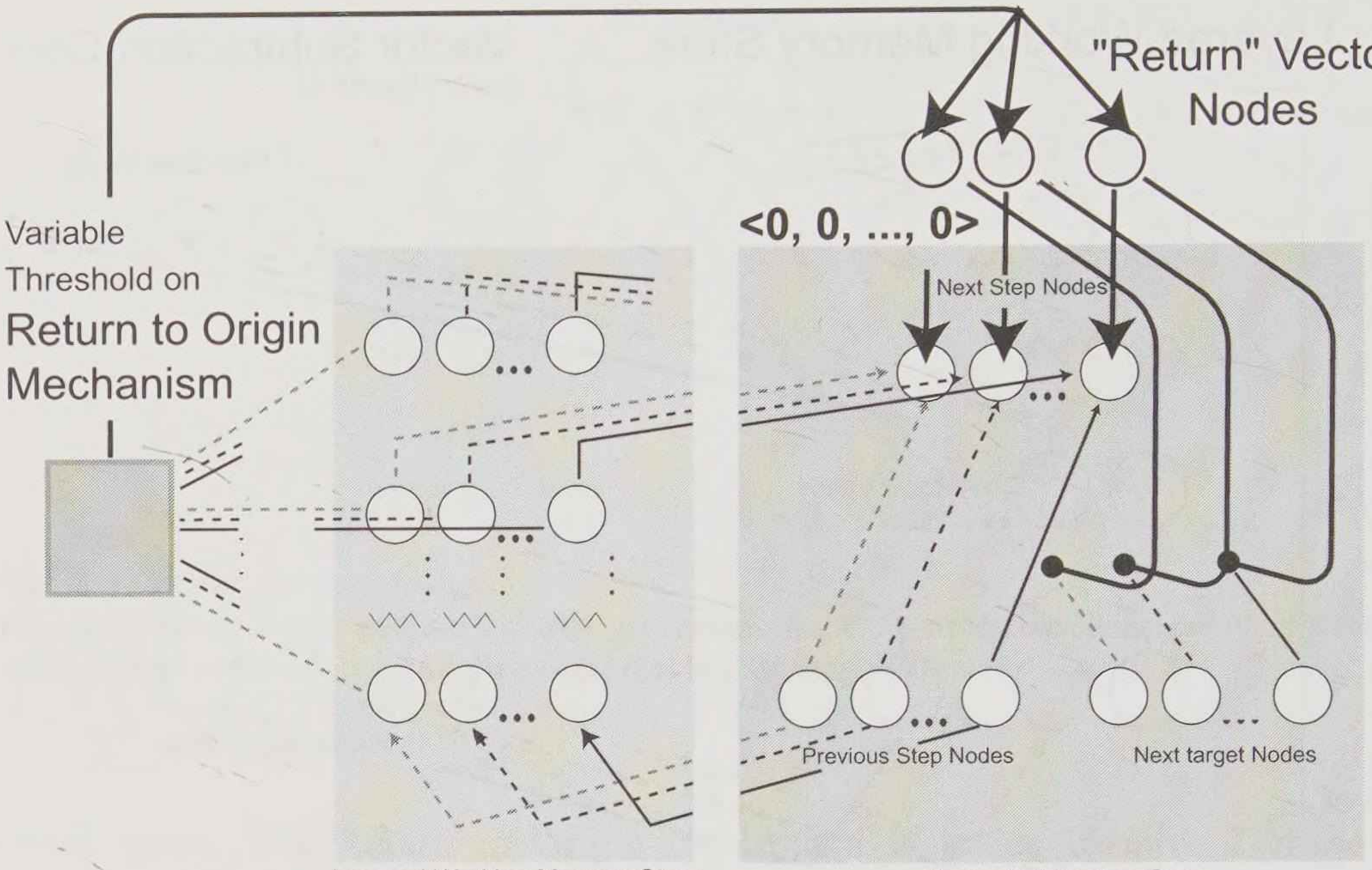


Figure 7. *Return to Origin Mechanism*
The Return to Origin Mechanism is tied to a Variable Threshold set at some level of Working Memory Store. When the return mechanism is engaged, input from Next Target Nodes is inhibited and replaced by ⟨0, 0, …, 0⟩. The resulting computation returns the network to an initial origin. Vector Subtraction Core and Working Memory Store remain as in Figures 5 and 6, respectively; only components relevant to Return to Origin Mechanism are shown here. (Adapted from Bickle *et al.* in press.)

As we have acknowledged, the primate saccade command system is not itself a 'cognitive' system. Most of the time we are not consciously aware of its outputs. But its FEF component computes sequences that possess the characteristic sequential features of cognition and the stream of consciousness (Bickle *et al.* in press). Furthermore, the neural tissue in these regions — Brodmann's areas 8a (FEF) and 46 (DLPC) — is standard 'frontal-type' cortex, in its basic cytoarchitectural structure, distribution of neuron types and number, and intra- and inter-column connectivities (Parent 1996). Many other regions in frontal cortex contain cells with properties and connectivities sufficient to implement vector subtraction in this same fashion. Our neurocomputational model thus generalizes directly to other regions known through physiological and neuropsychological studies to be involved in higher cognition and consciousness. Finally, our neurocomputational model yields testable physiological and anatomical predictions about these

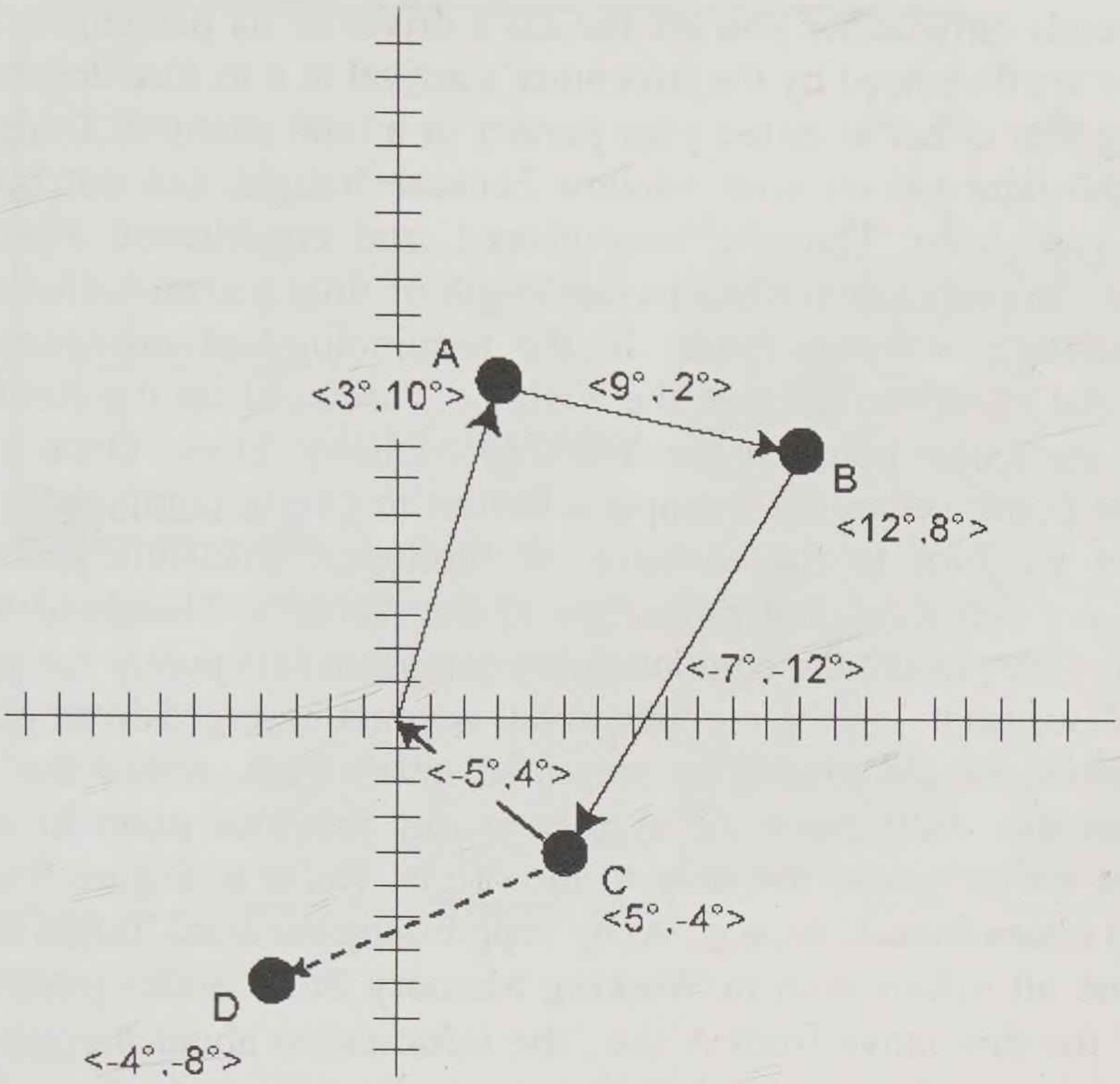


Figure 8. *Computing the Return to Origin Mechanism*
When the Return to Origin Mechanism is activated, the network computes the dimensions of **C** → **O** (solid line) instead of the dimensions of **C** → **D** (dashed line).

regions' cell-biological mechanisms for cognitive and conscious activity. It even yields some initially surprising but ultimately defensible implications about the phenomenology of our conscious streams (Bickle *et al.* in press).

4. Modeling the variable threshold

Up to this point, our model's Variable Threshold on the Return to Origin mechanism had been justified only on computational grounds. We needed it to account for a fact about cognitive and conscious processing. Sequential processes, running off in semi-ballistic fashion, can return to an initial point of origin. The factors controlling the 'Return' mechanism have to do with 'affect' and motivation. Motivational influences can sometimes overcome stimulus salience. Continued attention to the billboard on your left, as opposed to the road directly

ahead, depends on whether you are the car's driver or its passenger. The length of time you are distracted by the latecomer's arrival at a lecture depends on your recognizing him or her as either your partner or a total stranger. Tonight you are aware of the faint tap on your window because tonight, but not last, you are expecting your lover. Learned, remembered, and conditioned attachments to points in the sequence contribute to the length of time a semi-ballistic sequence runs off through a vector space. In the terminology of our general neurocomputational model, affect sets the Variable Threshold on the Return mechanism at a particular point in the Working Memory Store. Once a sequence reaches that point, activation prompts a Return to Origin command.

Before we turn to the question of biological plausibility, we need to enumerate the computational properties of our Variable Threshold and Return mechanism. (We present the computational extension first purely for pedagogical purposes. Constraints on the computational account emerged from our study of the neurobiological details, to be presented next). First, notice that there is a computation that will return the system to *any* previous point in a sequence through the vector space, not only to the origin. Refer to Figure 8 above. The system can return from **C** to, e.g., **A**, by suppressing the Next Target information about **D** and all information in Working Memory Store nodes prior to the one coding for the first move from **A** (i.e., the information about the move from **O** $\rightarrow$ **A**). The result will be that only information about the paths from **A** $\rightarrow$ **B** and **B** $\rightarrow$ **C** (plus the $\langle 0, 0, \ldots, 0 \rangle$ dimension which replaces the Next Target location) reaches the Next Move nodes. The resulting computation is

$$\begin{aligned} \mathbf{C} \rightarrow \mathbf{A} &= \langle -9, 2 \rangle + \langle 7, 12 \rangle + \langle 0, 0 \rangle \\ &= \langle -2, 14 \rangle \end{aligned}$$

these are exactly the necessary dimensions to return to **A** from **C**. The original Return to Origin mechanism of our general model is, of course, just a special case. It simply includes information about every move since the origin.

We call the component that sets the Variable Threshold on this extended Return mechanism "SAM": the *Significance Activation Mechanism* (see Figure 9). The level of the Working Memory Store at which SAM sets the threshold depends on the motivational significance of the contents of earlier states in the sequence (i.e., of points in the vector space representing those concepts or ideas). Thus for SAM to be a component of a *neuro*computational model of affective effects on sequential cognitive and conscious processes, we need to describe a neural system fulfilling the following demands.

(1) The model neural system must be involved in affective behavior and motivation.

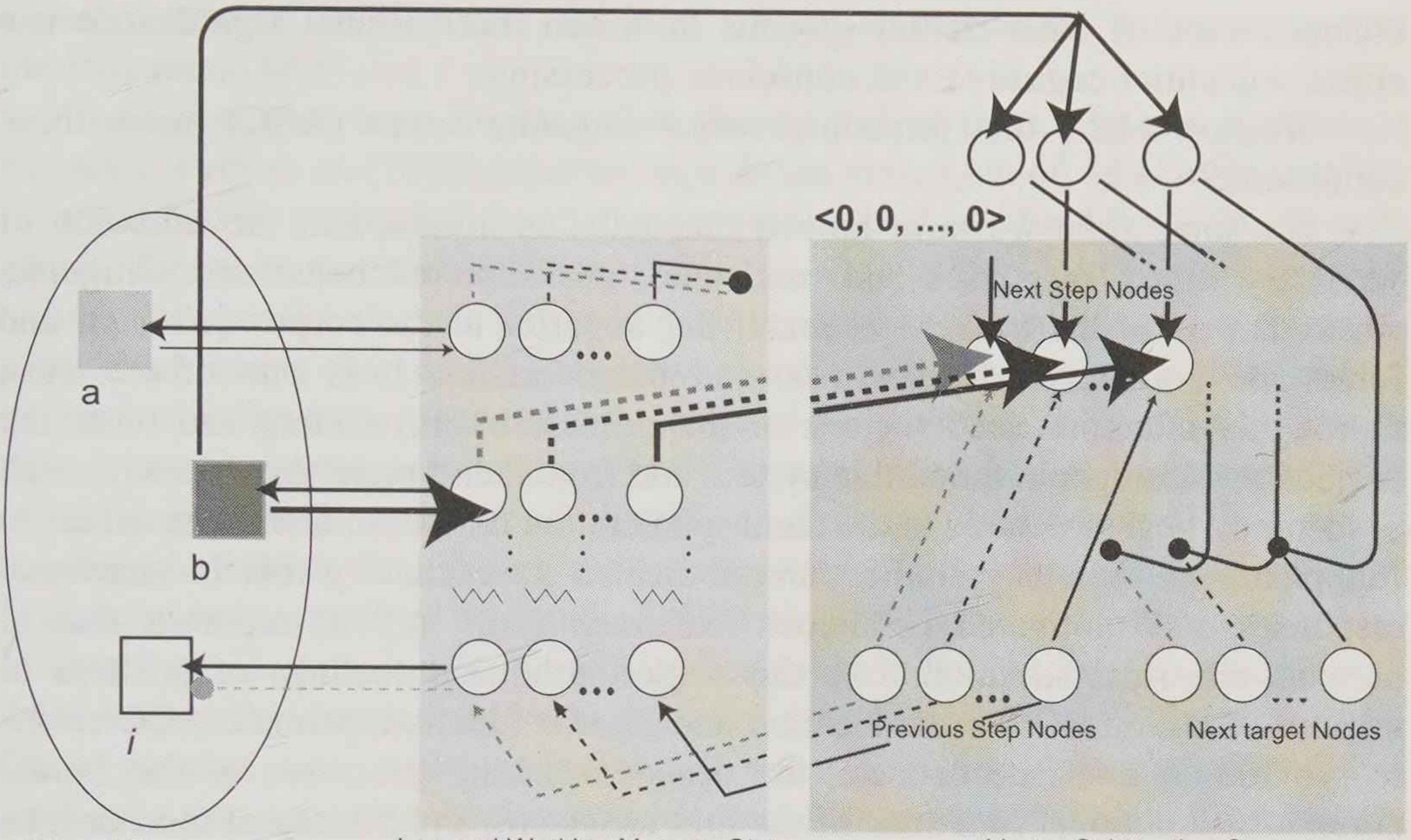


Figure 9. *The Significance Activation Mechanism (SAM)*
SAM returns the system to a significant point in the sequence. Outside input to SAM nodes (a, b, *i*) predetermines the significance of individual points in the sequence (shown here by degree of darkness). Nodes in Working Memory Store project to SAM nodes, but feedback projections remain inhibited (*i*) unless outside input exceeds threshold as in b. When this occurs, additional input to Working Memory Store Nodes effects the strength of those nodes' input to Next Step Nodes in the Vector Subtraction Core. This suppresses input of dimensions for all steps *prior to* the significant target. The Return Mechanism is engaged as in Figure 7.

(2) It must have appropriate anatomical connections with another neural system known to compute sequential outcomes. (All the better if it is connected with a system known to implement vector subtraction.)
(3) It must have demonstrable effects on sequential processing in the other system.
(4) Its cell properties and connectivities must be consistent with the computational features of SAM. Ideally, its cell properties and connectivities will yield insight into how SAM works neurocomputationally.
(5) The resulting account of SAM's structure must generalize fruitfully beyond the model biological system to yield testable predictions about cell properties and

connectivities of other neural systems in which motivational significance can effect sequential cognitive and conscious processing.

We now claim that primate *anterior cingulate cortex* (ACC) meets these conditions.

For some time we have known that ACC is involved in the initiation of voluntary motor responses and motivated, goal-directed behaviors. Cingulate cortex is part of the limbic system. It lies superior to the corpus callosum and forms the boundary between neo- and paleocortices. It is one of the most thoroughly interconnected regions of the primate brain, sending and receiving projections throughout the limbic system and frontal and parietal cortices (as well as to subcortical structures in the diencephalon, the midbrain, and the brain stem) (Shepherd 1994). Many studies have indicated its excitatory role in emotional responses and motivated behaviors. Talairach *et al.* (1973) reported that in humans electrical stimulation to this region produced feelings of positive or negative emotions. Given its location and known function, primate ACC seems to provide an interface between the decision-making processes of the frontal cortex, the emotional functions of the limbic system, and the neural mechanisms controlling movement (Marshall *et al.* 1997; Carlson, 1994; Vogt *et al.* 1992). Posner *et al.* (1994) included cingulate cortex in their 'attentional network' and attributed blood flow changes in its anterior region to 'attention to motion,' especially response selection. More recently, Mesulam (1998) described ACC as one of three transmodel areas in a distributed neural network for spatial awareness and directed spatial attention. In that network, it provides a map 'of expectancy and relevance.' (Interestingly, the frontal component of Mesulam's model includes the FEFs and DLPC. We report anatomical details about connections between these regions and ACC below.)

ACC activity is not correlated with simple detection tasks, but rather with motivated response selection when one or more competing behaviors must be suppressed. PET studies indicate strongest activity (compared with baseline rates) in ACC during tasks that require attentional processing of multiple inputs. Pardo *et al.* (1990) found high activity in ACC during the Stroop test, which requires subjects to report the color of a color word display (for example, to report "blue" when the word "red" is displayed in blue). Subjects must suppress reporting the word to generate the correct response. Corbetta *et al.* (1991) showed significant activity in ACC only when a subject had to detect a change in one of three changing features of a stimulus (speed, color, shape), but not when detecting a change in only one parameter. Aitken (1981) and Kirzinger and Jurgens (1982) report that ACC lesions in rhesus monkeys reduce condition-specific vocalizations but not spontaneous vocalizations. Thus ACC is active maximally when tasks require action selection, processing of competing inputs, or suppression of one or more responses. All of these features concern motivation and affect.

Clearly, primate ACC meets the first of our demands on a neural system implementing SAM.

ACC has also been implicated in the control of purposive and intentional saccades. Petit *et al.* (1993) reported significant activation of ACC over baseline (as measured by changes in normalized regional cerebral blood flow) during self-paced, voluntary saccades in humans. Paus *et al.* (1993) trained humans to make a direction-specific saccade depending on the type of visual stimulus presented. After establishing the stimulus-response association, subjects were given one of two tasks. In the over-practiced task, subjects had to make the same response to stimuli as during training. In the reversal task, subjects had to make exactly the opposite saccade (in terms of amplitude and direction) to the stimuli than during training. Activation foci during the reversal task but not the over-practiced task (measured by normalized regional cerebral blood flow) were in anterior portions of cingulate. Gaymard *et al.* (1998) studied saccade deficits in two patients with ACC lesions that overlapped in the posterior portion of the rostral cingulate zone. These patients were deficient in saccade tasks involving active fixation disengagement and intentional, purposive saccades (memory-guided and previously conditioned, without a visual stimulus). In conjunction with the known anatomical connections between ACC, FEF, and DLPC, their results provide clues about how SAM is neurally implemented in this system. We discuss Gaymard *et al.*'s results in more detail below. For now, however, notice that the ACC meets the third of our demands on a model neural system implementing SAM.

Much has been learned over the past two decades about ACC's connectivities with other regions. Using anterograde and retrograde tracing techniques, Vogt and Pandya (1987) have shown that DLPC efferents (primarily from area 46) terminate in ACC (specifically in the portion occupying Brodmann's area 24c). ACC afferents also include FEFs and areas in the posterior parietal cortex (areas 7, 7a, and LIP) known to code for object location and eye movement coordinates. These projections terminate primarily in caudal ACC (Brodmann's areas 24a and 24b). More fine-grained neuroanatomical studies reveal that these frontal and parietal connections are reciprocal and monsynaptic (Morecraft *et al.* 1993). Thus our model neural system for SAM has the appropriate connections to receive information about previous eye movements in the sequence (as they are held in the Working Memory Store modeled on DLPC, and Previous Step Nodes modeled on FEF postsaccadic activity).

ACC also receives information about the *significance*, and not just the surface properties, of all targets in the saccade sequence. It is heavily connected with other structures in the limbic system known to subserve emotional reactions. It receives afferents from lateral basal and accessory basal nuclei of the amygdala, structures involved in avoidance and reward behavior (Shepherd 1994). Hippocampal neurons also project directly onto ACC, providing input about

learned or remembered associations with particular targets. ACC does not receive visual input directly, but visual information from the inferotemporal terminal points of the ventral ('what?') stream and parietal terminal points of the dorsal ('where?') stream are relayed to ACC via the parahippocampal formation, as are inputs from thalamic nuclei (Gaymard *et al.* 1998). In this way, ACC receives information about object location and identity, as well as attached memories and associations.

Neuroanatomists are still exploring the complex intra-cingulate circuitries. But given the nature of the information ACC receives from these various regions, it is plausible to think that the internal circuitry integrates and filters this polymodal information, projecting it forward to frontal cortex. Anatomically, ACC is ideally situated to integrate *expectancy and motivational factors with visual identity and location information*, and with information about previous and targeted eye movements. There is even the beginning of an intra-ACC circuit explanation for how it codes the *relative* significance of targets, given the particular task at hand. ACC contains a large population of GABAergic (hence inhibitory) interneurons. These can modulate ACC output by setting a firing threshold on neurons projecting to frontal regions (Kalus and Senitz 1996). ACC reciprocal efferents to DLPC and FEFs appear to be under tonic inhibition by these interneurons. This inhibition is released for selective neurons as excitatory input from hippocampus, amygdala, and the ventral visual stream (via the parahippocampal formation) reaches ACC. At present, not enough is known about intracingulate circuitry to confirm this account. We are building a computer simulated neural network of this hypothesized circuitry to understand the types of cell properties and connectivities necessary to generate this kind of integration and output. Despite limitations in our current microanatomical knowledge, clearly ACC meets the second demand on a neural implementation of SAM.

We mentioned above that in conjunction with these known anatomical details, specific results from Gaymard *et al.*'s (1998) lesion study provide a clue about how SAM is neurally implemented in the ACC. Saccade latency and gain in the patients with ACC lesions were similar to normals in a gap task, where the central fixation point was extinguished 200 milliseconds before the peripheral target flashed. However, their saccade latency was significantly longer in overlap, memory-guided, and antisaccade tasks. In the overlap task, the central fixation point remains illuminated while the peripheral target appears. In the memory task, subjects must maintain fixation on the central target for some time after the peripheral target has been illuminated and extinguished, and then saccade to the target's remembered location. In the antisaccade task, subjects must saccade to a location exactly opposite (amplitude and direction) the location of the peripheral target from the central fixation. Interestingly, all tasks on which ACC-lesioned patients were deficient are known to require FEF involvement. Apparently, ACC

lesions lead to FEF deactivation. Saccade gain (amplitude) in these patients was also slightly hypometric (short of target) in the overlap task, and even more so in the memory-guided task. In the latter task, patients displayed little improvement in accuracy between eye position at the end of the initial saccade and final eye position. These amplitude results suggested to Gaymard *et al.* (1998) that the deficit produced by ACC lesions was neither sensory nor motor (the gap task results), but rather resulted from the summing of the saccade amplitude with an abnormally persistent fixation signal. In keeping with previous work on the role of ACC in rhesus monkeys on delayed response tasks and recent PET studies in humans, they hypothesize that ACC enhances the level of neural activity in the DLPC. Cingulate lesions thus produce a reduced level of DLPC activation, with an ensuing hypometric gain dimension emerging from the vector averaging taking place there. Finally, they attribute patients' inability to correctly execute chronological sequences of saccades to a similar hypothesis of frontal deactivation. They note explicitly that all of these explanations are consistent with the known circuitry between ACC, FEFs, and DLPC.

The connections between these findings and the role we ascribe to SAM are remarkable. Excitatory input from ACC to frontal regions provides a neural implementation of SAM's selective excitation of the appropriate nodes in the Working Memory Store and Vector Subtraction Core. SAM's selectivity is based on ACC's integrating memorial and affective factors received from its limbic afferents. SAM's excitatory effects result from ACC's excitatory efferents to DLPC and FEFs. The 'abnormally persistent fixation signal' Gaymard and his colleagues refer to is none other than activity in FEF suppression sites. Recall from the previous section that these neurons provide the biological model for the $\langle 0, 0, \ldots, 0 \rangle$ activation message substituted for the dimensions of the Next Target in the vector space by the Return to Origin mechanism. (See Bickle *et al.* for a detailed discussion of suppression sites.)

Gaymard *et al.*'s (1998) explanations even suggest how SAM's effects are implemented in the ACC-DLPC-FEF circuitry. We mentioned in the previous section that values in the nodes of our neurocomputational model are implemented neurally as *weighted vector averages* across a neuronal population. Each neuron has a preferred movement field through the vector space, and contributes the dimensions of that movement relative to its firing rate. DLPC and FEF neurons receiving these selective excitatory bursts from ACC afferents thus 'speak louder' in the vector averaging computation. Their 'votes' for their preferred movement dimensions through the vector space are stronger than are those of neurons not receiving selective ACC excitation. The result is that their preferred dimensions are more prominent in driving activity in the neuronal population implementing the Next Step Nodes of our model. Neurobiologically speaking, the computation of the Next Step Nodes is also the weighted vector

average of the neurons in the population projecting commands beyond the system. Owing to their increased activation rate (action potentials per second), neurons activated by the ACC excitatory projections make a much greater contribution to the dimensions of the next step through the vector space (the weighted vector average) than do neurons not receiving this additional excitation. This ACC-DLPC-FEF circuitry thus meets the fourth demand on a model neural system implementing SAM.

Finally, this model neural system for SAM generalizes far beyond an account of motivational influence on saccades. ACC-frontal circuits provide a plausible hypothesis about the neural mechanisms of affective influence on cognitive and conscious sequential processing generally. ACC has a somatotopic organization, with specific "cingulate motor areas" (Picard *et al.* 1995). Its efferents project extensively throughout primate frontal cortex, including to areas that neuropsychological and physiological studies have revealed subserve higher cognition and consciousness. For example, orbital frontal areas 11, 12, 13, and 14 in rhesus monkeys all receive reciprocal and monosynaptic projections from various ACC regions (Morecraft *et al.* 1993; Vogt and Pandya 1987). Damage to these regions is now clearly correlated with combined emotional and cognitive deficits involved in primate social interactions (Kolb and Whishaw 1996). Orbitofrontal lesioned monkeys show significantly reduced social interaction. They quickly lose their status in the troop's dominance hierarchy, and display inappropriate social behavior to troop members. Large frontal lesions that include orbital cortex cause a drastic reduction of facial expressions, posturing, and gesturing in social situations, without a loss of the motor capacities necessary for these behaviors. Large frontal lesions combined with anterior limbic lesions (including ACC) drastically reduce social vocalizations. The sequential aspects of social cognition, so prominent in primate behavior and so susceptible to emotional and affective influence, seem explicable within our neurocomputational model extended to include SAM. At the very least, our model predicts cell properties and connectivities in these regions that would be sufficient to implement sequential cognitive/conscious processing and affective influences upon them. These are the cell properties and connectivities that implement the Vector Subtraction Core, the Working Memory Store, the Return mechanism, and SAM in other frontal cortical regions and the ACC. If these cell properties and connectivities are present in other regions — as the common ACC-frontal circuitries and cytoarchitectural similarity across frontal cortical regions suggest that they should — then this known neural implementation of vector subtraction and SAM is a plausible account of the mechanisms of sequential cognitive processing, the stream of consciousness, and affective influences on both. Further computational exploration of these circuits, including computer simulations of the biological details, can provide anatomists and physiologists with

more detailed accounts of the connectivities and cell properties for which to search. ACC-frontal circuits thus meet our fifth demand on a neural implementation of SAM.

5. Conclusions

Our model provides a biologically plausible neurocomputational mechanism for a motivated sequence of eye movements. The properties and connectivities of each component in the model generalize to frontal and cingulate regions subserving other cognitive processes and the stream of consciousness. Our model thus provides a biologically plausible neurocomputational account of features of cognitive and conscious processing, including the effects of motivation and affect. But does our model reveal anything about the *phenomenology* of motivated sequences? Does it tell us anything about the *nature* of affect? Does it shed any light on what it *means* (in neurobiological terms) to recognize an object, a spoken word, a "faint forgotten tap" in a stretch of preprogrammed cognitive/conscious activity as being motivationally significant? We believe that the answer to these questions is yes.

Our model shows that affect is a *component* of the mechanism that produces sequential cognitive and conscious processes possessing the characteristic features we noted in the Introduction. The age-old notion that cognition is separate from 'lower-level' emotional centers is no part of our account. The cell properties and circuitries on which we base our model involve excitatory input from motivational (i.e., limbic) centers necessary to drive activity in 'cognitive' and 'conscious' neural regions. Our model provides a neurocomputational mechanism for the fusing of affect with sequential cognitive and conscious processes. Given its biological plausibility it also points toward a possible neurophysiological explanation for this *integration*.

In everyday language we explain our reason for continuing to do *this*, not *that*, by saying that we were 'not at all motivated to do *that*! It had no significance'. We may explain our reason for *not doing that* sooner (and instead *continuing to do this*) by saying that we were 'only slightly motivated to do *that*. It wasn't very significant.' We may explain why, during the course of a sequence of doing (attending, speaking, listening, thinking) we returned to a previous deed (focus, word, thought) by saying that we were '*motivated* to return! It was more interesting, meant more to us.... It was significant!' We submit that these explanations map onto our neurocomputational model and given the biological plausibility of our model, map onto a neurobiological account of phenomenal motivation. A point in the sequence to which no learned, memorialized or conditioned associations have been made, nor which has salient surface features,

results in no input (from areas containing this information) to the Significance Activation Mechanism. SAM feedback to Working Memory Store remains inhibited. That point has no motivational significance. A target which elicits *some* associations, i.e., has *moderate* significance, may result in weaker input to Working Memory Store because the combined (summed) strength of inputs releases inhibition. That weaker input may or may not affect vector averaging in Working Memory nodes. Not all inputs to SAM need occur at a discrete instant in time. Points can conjure up old memories and trigger inactive circuits. Activating these circuits takes time and results in delayed input to SAM. Correspondingly, the temporal summation of these inputs delays release of SAM inhibition. During that delay, the Next Move Nodes *continue to compute the dimensions of successive steps through the vector space*. When the message is finally sent forward to Working Memory Store it may result in the computation of a return message. In retrospect, the target has motivational significance, it just took longer to 'realize' it!

Remember too that our model assumes a limit on the number of levels in Working Memory Store. It is possible that by the time SAM inhibition is released, the dimensions of the appropriate step will have exhausted their time in Working Memory. We are left with the vague feeling that something was important after all, but 'what was it?' On the other hand, some targets can trigger such immediate and strong input to SAM that inhibition is released virtually immediately. Input from SAM to Working Memory quickly dominates output of those layers and a return message is signaled before more successive steps can be computed. In neurobiological terms, varying degrees of (phenomenal) motivation *may just be* varying degrees of temporal delay.

Acknowledgments

We gratefully acknowledge Cristian Skinner's contribution to this project. As in the past, we thank Robert Graham for his continuing tutelage in neuroanatomy and human behavior, and remember our friend, Gary Peterson.

References

Aitken, P. 1981. "Cortical Control of Conditioned and Spontaneous Vocal Behavior in Rhesus Monkeys". *Brain Language* 13: 171–184.

Baars, B. 1997. *In the Theatre of Consciousness*. Oxford: Oxford University Press.

Bernstein, M., S. Stiehl, R. Graham, and J. Bickle. 1999. "A Neurocomputational Model of Cingulo-parieto-frontal Circuits Controlling Saccadic Eye Movements". *Society for Neuroscience Abstracts* 25(1): 153.

Bickle, J., C. Worley, and M. Bernstein. In press . "Vector Subtraction Implemented Neurally: A Neurocomputatinal Model of Some Sequential Cognitive and Conscious Processing". *Consciousness and Cognition.*

Bickle, J., M. Bernstein and C. Worley. 1998. "A Neural Network Model of Vector Subtraction and Variable 'Redirect' Mechanisms for Selective Visual Attention in Area LIP and FEF. *Society for Neuroscience Abstracts* 24(2): 1677.

Bickle, J., R. Graham, and M. Bernstein. 1997. "Modeling a Voluntary Selective Visual Attention Mechanism Using Parietofrontal Cell Properties and Connectivities". *Society for Neuroscience Abstracts* 23(1): 1589.

Bower, J., and D. Beeman. 1995. *The Book of Genesis: Exploring Realistic Neural Models with the GENERAL NEural SImulation System*. Santa Clara, CA: Telos

Bruce, C., and M. Golberg. 1985. "Primate Frontal Eye Fields: I. Single Neurons Discharging Before Saccades". *Journal of Neurophysiology* 53: 603–635.

Burman, D., and C. Bruce. 1997. Suppression of Task-related Saccades by Electrical Stimulation in the Primate's Frontal Eye Field. *Journal of Neurophysiology* 77: 2252–2267.

Churchland, P. M. 1995. *The Engine of Reason, the Seat of the Soul*. Cambridge, MA: MIT Press.

Churchland, P. M. 1989. *A Neurocomputational Perspective*. Cambridge, MA: MIT Press.

Churchland, P. S., and T. Sejnowski. 1992. *The Computational Brain*. Cambridge, MA: MIT Press.

Corbetta, M., F. Miezin, S. Dobmeyer, G. Shulman, and S. Petersen. 1991. "Selective and Divided Attention during Visual Discrimination of Shape, Color, and Speed: Functional Anatomy by Positron Emission Tomography". *Journal of Neuroscience* 11: 2383–2402.

Crick, F., and C. Koch. 1998. "Consciousness and neuroscience". *Cerebral Cortex* 8: 97–107.

Crick, F., and C. Koch. 1990. "Towards a Neurobiological Theory of Consciousness". *Seminars in Neuroscience:* 263–275.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.

Funahashi, S., M. Chafee, and P. Goldman-Rakic. 1993. "Prefrontal Neuronal Activity in Rhesus Monkeys Performing an Anti-saccade Task". *Nature* 365: 753–756.

Gaymard, B., S. Rivaud, J. Cassarini, T. Dubard, G. Rancurel, Y. Agid, and C. Pierrot-Deseilligny. 1998. "Effects of Anterior Cingulate Cortex Lesions on Ocular Saccades in Humans". *Experimental Brain Research* 120: 173–183.

Gazaniga, M. (ed). 1997. *Conversations in the Cognitive Neurosciences*. Cambridge, MA: MIT Press.

Geogopoulos, A., A.Schwartz, and R. Kettner. 1986. "Neural Population Coding of Movement Direction. *Science* 233: 1416–1419.

Goldberg, M., and C. Bruce. 1990. "Primate Frontal Eye Fields. III. Maintenance of Spatially Accurate Saccade Signal". *Journal of Neurophysiology*: 64, 489–508.

Goldberg, M., H. Eggers, and P. Gouras. 1992. "The Ocular Motor System". In E. Kandel, J. Swartz, and T. Jessell (eds.). *Principles of Neuroscience* 3rd edition. New York: Appleton and Lange.

Goldman-Rakic, P. 1996. "Regional and Cellular Fractionation of Working Memory". *Proceedings of the American Academy of Sciences* 93: 13473–13480.

James, W. 1890. *Principles of Psychology*. New York: Henry Holt.

Kalus, P., and D. Senitz. 1996. "Parvalbumin in the Human Anterior Cingulate Cortex: Morphological Heterogeneity of Inhibitory Interneurons". *Brain Research* 729: 43–54.

Kirzinger, A., and U. Jurgens. 1982. "Cortical Lesion Effects and Vocalization in the Squirrel Monkey". *Brain Research* 233: 299–315.

Kolb, B, and I. Wishaw. 1996. *Fundamentals of Human Neuropsychology*. New York: W. H. Freeman.

Lee, C., W. Rohrer, and D. Sparks. 1988. "Population Coding of Saccadic Eye Movements by Neurons in Superior Colliculus". *Nature* 332: 357–360.

Marshall, J., P. Halligan, G. Fink, D. Wade, and R. Frackowiak. 1997. "The Functional Anatomy of a Hysterical Paralysis". *Cognition* 64: B1-B8.

Mesulam, — M. 1998. "From Sensation to Cognition". *Brain* 121: 1013–1052.

Morecraft, R., C. Geula, and M. Mesulam. 1993. "Architecture of Connectivity Within a Cingulo-frontal-Parietal Neurocognitive Network for Directed Attention". *Arch Neurology* 50: 279–284.

Parent, A. 1996. *Carpenter's Human Neuroanatomy*, 9th edition. Baltimore: Williams and Wilkins.

Paus, T., M. Petrides, A. Evans, and E. Meyer. 1993. "Role of the Human Anterior Cingulate Cortex in the Control of Oculomotor, Manual, and Speech Responses: A Positron Emission Tomography Study". *Journal of Neurophysiology* 70(2): 453–469.

Petit, L., C. Orssaud, N. Tzourio, G. Salamon, B. Mazoyer, and A. Berthoz. 1993. "PET Study of Voluntary Saccadic Eye Movement in Humans: Basal Ganglia-Thalamocortical System and Cingulate Cortex Involvement". *Journal of Neurophysiology* 69(4): 1009–1017.

Picard, N., and P. Strick. 1995. "Differential Involvement of M1 and a Premotor Area in a Motor Task". *Society for Neuroscience Abstracts* 21(2) 2077.

Posner, M., and S. Deheane. 1994. "Attentional Networks". *Trends in Neuroscience* 17: 75–79.

Shepherd, G. 1994. *Neurobiology* 3rd edition. New York: Oxford University Press.

Talairach, J., J. Bancaud, S. Geier, M. Bordas-Ferrer, A. Bonis, G. Szikalz, and M. Rusu. 1973. "The Cingulate Gyrus and Human Behavior". EEG *Clinical Neurophysiology* 34: 45–52.

Vogt, B., and D. Pandya. 1987. "Cingulate Cortex of the Rhesus Monkey: II. Cortical Afferents". *Journal of Comparative Neurobiology* 262: 271–289.

Vogt, B., D. Finch, and C. Olson. 1992. "Functional Heterogeneity in Cingulate Cortex: The Anterior Executive and Posterior Evaluative Regions". *Cerebral Cortex* 2: 435–443.

CHAPTER 8

Motivation and Emotion

An Interactive Process Model

Mark H. Bickhard
Lehigh University

This chapter outlines dynamic models of motivation and emotion. These turn out not to be autonomous subsystems, but instead are deeply integrated in the basic interactive dynamic character of living systems. Motivation is a crucial aspect of particular kinds of interactive systems — systems for which representation is a sister aspect. Emotion is a special kind of partially reflective interaction process, and yields its own emergent motivational aspects. In addition, the overall model accounts for some of the crucial properties of consciousness.

1. Representation

I begin with representation, and outline a model of representation as a fundamental solution to the biological problem of action selection.

1.1 *Interaction selection in a complex interactive system*

Any complex organism must solve the problem of action selection — what to do next. In sufficiently simple systems, a triggering relationship may suffice, in which environmental inputs directly trigger particular actions. In some bacteria, for example, if they find themselves swimming up a sugar gradient, they continue swimming, but if the inputs correspond to their swimming down a sugar gradient, they stop swimming and tumble for a moment (D. Campbell 1974, 1990).

In more complex circumstances, however, simple triggering cannot suffice. The action and interaction potentialities for the organism are too numerous, and the reliability of those actions and interactions is too weak. A frog, for example, may see a fly, and, therefore, have the potentiality of flicking its tongue in a

certain way followed by eating. But it may simultaneously see the shadow of a hawk overhead, in which case it also has the selection option of jumping into the water. Both potentialities must be somehow indicated to or for the frog so that a selection between them can occur.

Furthermore, if the hawk shadow is not present and the frog misses the fly, it may be advantageous to detect that failure of the tongue flicking action and, on the basis of that detection, to make a further selection of interaction. That further selection might be to try again, or might be to move to a different location where flies are perhaps more numerous or slower. It can be advantageous, in other words, to be able to detect failures of actions, as well as to be able to select among potential actions.

A slight addition to the ability to indicate potential interactions suffices to allow such error detection. In particular, if interaction potentialities can be indicated, then so also might the internal outcomes of those interactions be indicated in association with them. That is, it is not only the interactions per se that are indicated as potentialities to select among, but also the internal outcomes that can generally be expected if they are in fact selected.

Furthermore, such an indication of outcomes provides the basis for making such interaction selections in the first place: if the outcomes are related to current goals, then select the associated interactions. Then, if the indicated outcomes are not attained, that constitutes the detection of error, and can influence further processing, including further selections of interactive processes.

A simple digital architecture that would permit such indications is that of pointers, as in a standard computer. A more biologically realistic process would involve a more continuous process of preparation for further interactive processes together with the ability to detect when those preparations fail to be prepared for the actual course of interactive flow. The preparations themselves constitute the indications of potentiality, while the failure of preparation to be in fact prepared constitutes the failure of the interactions to yield the outcomes, the interactive flow, for which they were selected. Elsewhere I discuss details of such a continuous preparation process, called microgenesis (Bickhard and Campbell 1996). The possibility of such continuity is important for some later issues in this discussion, but I will not elaborate the architectural and dynamic specifics here.

An important question at this point is: how are indications of interactive potentiality set up? What determines what is potential at any particular point in time? The answer is relatively simple: the outcomes of prior interactions serve as the basis for indicating what will be the next interactive potentialities. Conversely, the indication of an interactive potentiality will in general be conditional on the outcomes of particular prior interactions. The logic of such indications is based on the fact that interactions with an environment can serve to differentiate that environment. The internal course of an interaction will depend both on the

organization of the subsystem engaged in the interaction and on the environment being interacted with. If a subsystem is capable of, say, two possible final internal outcome states (two for simplicity of discussion), A and B, then actually arriving at A will differentiate the current environment as of the type that yields outcome A, and as different from the environmental type that yields outcome B. Outcomes of interactions, then, differentiate the environments with which the interactions have taken place. Environments of type A, in turn, may also be environments in which further interactions Q, R, and S, are possible, each with its own associated set of indicated outcome states. Each of those outcome states, for one further step, may indicate — if arrived at — some further set of interactive potentialities.

That is, indications of interactive potentialities may branch, with multiple possibilities being indicated, and can iterate, with each potential outcome serving to indicate still further potentialities. These branching and iterated indications (not to mention the possibilities if continuous outcome spaces are taken into account) can link into vast and complex webs of conditionalized indications of interactive potentiality.

In general, then, an interactive system will be continuously interacting, and continuously preparing for further interaction on the basis of prior interactive flow. Those preparations constitute indications of potentiality, among which further selections of the course of interactive processes are made in accordance with any relevant goals.

1.2 *Representation*

The discussion of interactive systems and the selection of the course of interaction has made no mention of representation. Nevertheless, I claim that an outline of the emergence of representation has already been given. That is, representation emerged naturally in the evolution of interactive systems as a solution to the problem of interaction selection.

In particular, the indication of potential interactions is the point of emergence of the crucial properties of aboutness, truth value, and content. First, the indication of the potentiality of particular interactive processes in an environment is an indication *about* that environment — an indication that it is appropriate for those interactions. It is an implicit predication that this environment is appropriate for these interactions. Similarly, conditionalized indications constitute general predications — type A environments are subsystem Q type environments.

Second, that indication might be false. The environment might not in fact support reaching one of the indicated internal outcomes. Furthermore, if none of the indicated outcomes is reached, that indication is thereby falsified for the

system itself. There is system detectable error — system detectable *representational* error.

Third, there is the emergence of content. Some patterns of environmental properties will support an interactive indication and some will not. Such an indication, then, predicates some one of those sufficient patterns of properties to the environment, and those properties constitute the content of the representation. Content in this form is implicit, not explicit as in most models of representation, a difference that I argue elsewhere has powerful consequences, such as resolving the frame problems (Bickhard and Terveen 1995).

1.3 *Challenges*

This is a very primitive form of representation — appropriate perhaps to flatworms and maybe frogs — and it is subject to its own challenges. Such challenges have been addressed in detail previously, but there are two that I will respond to here.

The first is a potential circularity: representation has been modeled making use of a notion of goal, and if goals, in turn, are themselves necessarily representational, then representation will have been modeled in terms of representation. The goals needed here, however, are not necessarily representational. They need only have the character of internal set points that regulate the internal flow of control in an interactive system. Such set points may, or may not, correspond to something — blood sugar level, for example — but need not represent it. Once representation is emergently available, of course, then goals might themselves make use of them.

The second challenge is to the adequacy of this interactive model of representation: can it account for more familiar forms of representation in addition to these primitive action potentiality indications. One such familiar kind of representation is that of small physically manipulable objects, such as a child's toy block. The complex webs of interactive indications can form representations of such objects. A toy block, for example, offers the potentialities of multiple visual scans, multiple manual manipulations, chewing, dropping, and so on. Furthermore, every one of these potentialities indicates the potentiality of all the rest, perhaps with intermediate interactions along the way, as if a visual scan indicates the potentiality of another visual scan so long as the appropriate turn of the block has occurred. Such a subweb, then, is internally completely reachable.

It has one additional critical property. The entire web of potentialities will remain invariant under a large class of additional interactions. The toy block will continue to offer its interactive possibilities — will remain invariant — under putting the block away in the toy box, moving to another room, hiding it, and so on, though it will not remain invariant under crushing or burning. Such reachable

invariances among interactive webs constitute the representation of small objects. Clearly, this is basically a Piagetian model of object representation.[1]

The interactive model of representation captures several characteristics of phenomenological awareness that should be mentioned. The model is of a continuous flow of interactive process that is inherently contentful — that exhibits aboutness and intentionality. It is necessarily from a point of view, and is correspondingly deictic and indexical. It is inherently embodied: disembodiment renders interaction impossible. It is inherently temporal: successful interaction is as much a matter of coordinated timing of interactions as it is of sequence of actions. Even this relatively simple elaboration of the model, then, captures important properties of consciousness (Bickhard 1998a, in press-a).[2]

1.4 *Encoding models of representation*

Standard models of representation do not look much like the interactive model. Standard models, in fact, do not require any interaction at all. Most focus only on one aspect of the overall interactive process, the differentiations that, in the interactive model, ground the representational indications.

In particular, a simple form of interaction is one with no outputs — a passive processing of inputs. Such a passive process will differentiate environments according to which internal states are produced, just as will full interactions, though in general with less overall differentiating power. Furthermore, it is clear that the sensory systems of complex organisms engage, at least in part, in precisely such passive input processing.

But, whereas the interactive model gives such processes the function of differentiating environments, of providing ongoing sensitivity to the environment, so that appropriate indications of interactive potentialities can be set up, standard models ignore that output aspect of interaction and construe the differentiations themselves as being representational. The differentiating internal outcomes are deemed to represent, to encode, whatever it is that they have differentiated (Bickhard 1993; Bickhard and Terveen 1995).

In the interactive model, differentiations are not assumed to have any content, are not assumed to be representational themselves at all. A differentiating outcome of an interaction does not announce what it is that it has differentiated, nor, for that matter, that it is a differentiation at all. All that the interactive model requires is that it has *in fact* differentiated environments in a way that is *in fact* useful for the indications of further interactive potentialities. There is no need that what has been so differentiated be known or represented.

But these factual differentiations also constitute, in any particular case, factual correspondences between the internal states and whatever has been differentiated, and these correspondences are typically offered as models of

representation. Such correspondences may be postulated in differing forms — as causal, as lawful, as informational, and so on — but some such type of correspondence is supposed to constitute representation.

There are myriad multifarious failures of logic and of naturalism in such models. I will briefly mention only two: emergence and error. The central characteristic of representation is content. Content is what determines what a representation is *supposed* to represent, and, therefore, it is what determines whether a particular application of a representation to a particular situation or target (Bickhard 1993; Cummins 1996) is true or false. Content is the normative aspect of representation. Accounting for the nature and emergence of content is, thus, the central problem.

Unfortunately, current correspondence or encodingist models make little progress in accounting for content in any naturalistic way. They attempt to capture the specifications of content in a strictly externalist manner, with little or no attention to how content, especially its normative character, could be dynamically realized. If some element is in a favored kind of correspondence — causal, informational, lawful, etc. — with something else, then that something else is proposed as the content. But there is no model of how content could exist, could emerge — of how the crucial information about the correspondence could be available — in the processes of the supposed epistemic system itself.

But representation did not exist at the moment of the Big Bang, and it does exist now, therefore it has to have emerged. Therefore, any model that cannot account for such emergence is falsified.

It is often acknowledged that we have no model for content, for mental representation, e.g., "we haven't got a ghost of a Naturalistic theory about [encoding]" (Fodor 1987: 81). Instead of taking this as a refutation of current models, however, the failure to account for representational emergence is taken as a premise in arguments for the necessary innatism of all content. If content can't emerge in learning in development, then it must be innate (Fodor 1981; Bickhard 1991). But if it can't emerge, then it can't emerge in evolution either, and Fodor's argument begs the question — "What I think it [the *Language of Thought* argument] shows is really not so much an a priori argument for nativism as that *there must be some notion of learning that is so incredibly different from the one we have imagined* that we don't even know what it would be like as things now stand." (Fodor in Piattelli-Palmarini 1980: 269).

The general failure to account for content has many manifestations. One of them is a failure to account for the normativity of representation, in particular, to account for the possibility of representational error. In encoding models, there are only two possibilities: either the favored correspondence exists or it does not exist. But, if it exists, then the representation (supposedly) exists, and it is correct, while if it does not exist, then the representation does not exist, and *it*

cannot be incorrect. There are three representational possibilities that must be accounted for — exists and correct, exists and incorrect, and does not exist — but there are only the resources of two kinds of cases to do the job. It's impossible.

Much effort has been devoted to finding a way out of this dilemma, but they all fail to naturalize content and error. To determine what a representation is supposed to represent requires, in current models, the assessment of complex evolutionary or learning histories (Dretske 1988; Millikan 1984, 1993) or the equally complex assessment of complex relations among counterfactuals (Fodor 1990; Loewer and Rey 1991). None of these are remotely reasonable as a model of content in a simple epistemic system. Then to assess whether the representational instance is true or false requires comparing those inaccessible contents to what is actually currently being represented. But representing what is currently being represented is the original problem all over again.

These models are realizable, if at all, only by an external observer to the epistemic system at issue, an observer who can, at least in principle, make the complex assessments of history and counterfactuals to determine the 'content' and who has, again at least in principle, independent representational access to the environment so that he or she can compare the deployed content with what is actually out there in the world — who can determine that the COW representation is being used for what is in fact a horse on a dark night, and, therefore, is false. Such a dependence on an external observer fails to naturalize representation. Among other problems, it fails to account for the representations of the observer, except by initiating a vicious regress.

Some models attempt to make a virtue out of this necessity for an observer by construing the problem of representation as one of accounting for how it is useful to use the language of representation. That is, they construe representation as a manner of speaking, having no further ontological nature, and address issues of when it is explanatorily useful to make use of such a manner of speaking or writing (Bogdan 1988; Clark 1997; Dretske 1988). Clearly there are some phenomena, including normative phenomena, that are emergent only in the realm of social practice: marriage and money come to mind. But the relationship of the individual to the realm of social practice is already a normative, a representational, relationship, so representation cannot be subsumed *into* social practice without committing to a full social idealism. That is not only a failure of naturalism, it is internally incoherent.

There are many more failures of such models (Bickhard 1993; Bickhard and Terveen 1995), but, although they are frequently acknowledged, the usual assumption is that some form of encodingism is the only possibility and that the problems will be overcome eventually. I argue that the failures are inevitable so long as representation is not understood as a dynamic phenomena of pragmatic

action and interaction, not just a spectator phenomena of input processing (Bickhard 1993, 1996, 1998a, 1998b, 1999; Bickhard and Terveen 1995).

2. Motivation

Representation has been modeled above as an aspect of an underlying interactive system ontology. Representation is the aspect of indicating further interactive processing potentialities; the aspect of anticipating the flow of interaction. My claim is that, just as representation is an aspect of such a system ontology, so also is motivation a different aspect of the same ontology.

Before elaborating on that claim, I first need to address what is taken as the problem of motivation. A classical construal of motivation has been as that which induces a system to do something rather than nothing. The organism is assumed to be inert unless motivated to do something, thus motivational metaphors such as various kinds of pushes and pulls, drives and 'motivations' (such as competency motivation). That is, the organism is assumed to be inert unless some sort of 'energy' is provided to make it move.

But organisms are alive, and living beings cannot stop, cannot be inert, without simply ceasing to exist as living beings. Living beings *cannot* do nothing. So the problem of motivation cannot be that of what makes an organism do something rather than nothing. The problem of motivation must be what makes an organism do one thing rather than another — what are the processes of the selection of the course of further activity, of further interactive activity (Mook 1996).

Rather clearly, that is precisely the interactive system function that representation was proposed to subserve. That is, anticipation of what's possible — representation — serves the function of selecting what among those possibilities to select next — motivation. Motivation is the aspect of selection of processes, and representation is the aspect of anticipation in the service of such selection (Bickhard 1997).

This is a minimal model of motivation, as is the initial model of representation, and requires similar attention to more complex and more familiar kinds of motivation. Not all motivation is simple selection or goal directed selection. As for representation, this minimal model holds perhaps for flatworms and maybe frogs. Some more subtle versions of process selection — of motivation — will be outlined later as emergents of more complex processes.

3. Learning

I will not focus on learning in this chapter, but I do need one property of

learning for the model of emotion to follow. Learning requires a monitoring of ongoing interactive processes. Learning introduces variation when things are not going well, and stability when they are proceeding according to plan. In this case, 'plan' is the anticipations of the microgenesis process. If microgenesis, the set up for the next interactive processing, is destabilized when failure to anticipate occurs, and is stabilized so long as the anticipations are successful, then we have a minimal model of learning: such a system will tend to stabilize on interactive processes that proceed successfully according to the anticipations and goals of the system.

Note that even with this minimal model, we can account for several phenomena. If an input path into the central nervous system is neurally wired so that inhibitory interactions with the inputs are possible, and if an actual input stream is restricted to such a pathway, then it is possible for the system to learn to interact with such an input stream strictly via such neural inhibitory anticipations of the flow of that input stream. This is classical habituation (Staddon 1983). A well habituated simple tone doesn't progress higher than the first cochlear nucleus — the anticipatory interactive processes can be completed at that level. A more complex tone, however, may require a small participation of the temporal lobe in order for the interactive anticipations to succeed. That these are anticipations rather than crude pathway inhibitions is evidenced by the fact that reducing the volume of the tone, for example, produces arousal — the volume anticipations fail.

Suppose now that the input flow does not remain in one modality. Suppose, in fact, that it crosses from sound, a tone of some sort, into pain — a foot shock, say — where pain is, among other things, a form of input for which no successful interactions are possible (to a first approximation: Douglas 1998; Eccleston and Crombez 1999). Now to successfully anticipatorily interact with this flow, something must be done about the shock. The only way to successfully interact with the shock is to avoid it, so the proper response to the tone is to remove oneself from the grid at the bottom of the cage. The full interaction now involves skeletal muscles. Classical conditioning is a direct result of the ongoing stabilization only on successful anticipatory interaction.

For one further elaboration, consider an input that originates from low blood sugar, perhaps in the hypothalamus.[3] Again, to a first approximation, there is no direct inhibitory interaction possible, but, nevertheless, some form of successful interaction is possible. In particular, interaction that results in raising blood sugar will successfully interact with this input. What will succeed in raising blood sugar will, in general, depend on multiple additional differentiations and representations about the environment. Refrigerators usually work fine, if available. Hunting may be involved if in the wild. In any case, we have a model of instrumental conditioning.

Most learning is more complex than these examples, at least in mammals. Most learning is heuristic. Accounting for heuristic learning requires a more complex model than has been outlined here (Bickhard and Campbell 1996), but these points suffice for my current purposes.

4. Emotion

A creature that had available only interaction and learning would, if in a sufficiently variable environment, suffer a potentially serious limitation. In an encounter with a novel situation, the only possible responses would be direct interaction or learning trials. Microgenesis would, by assumption, be not fully defined — would not set up clear and dynamically well organized anticipations of interactive potentiality. That is the dynamic side of the assumption of novelty. But the only monitoring of such uncertainty of microgenesis is by the learning process. Such interactive uncertainty is what learning is supposed to correct.

But learning, even heuristic learning, is at best a trail and error process, a process engaged in evolutionary epistemology. A first encounter with a tiger on a jungle trail might evoke interactions of foot wiggling, or an attempt at a handshake, or various other learning and interactive trials, but there is, in an organism limited to interaction and learning, no other possibility. In particular, there is no way for such an organism to develop general modes of interactive response to situations of interactive (microgenesis) uncertainty — it is only the learning process that has access to any information or signal of such uncertainty.

Nevertheless, the learning process does involve the generation of some version of such a signal, and if (a copy of) that signal could be fed back into the interactive system as an input, then the interactive system would be in a position to potentially be able to interact with its own conditions of uncertainty similarly to interacting with environmental conditions. The interactive system would learn, would stabilize, on forms of interaction that tended to be successful in interacting with internal uncertainty in the same sense in which it would learn to interact with tones and shocks and hunger. With such a capability, the organism could develop general ways of dealing with kinds of uncertainty situations, such as running whenever strange and large animals are encountered.

The modeling proposal is that emotions *are* such interactions with internal dynamic uncertainty. As is by now familiar, this is a minimalist model, appropriate perhaps to reptiles, and elaboration is required to account for familiar cases.

4.1 *Negative and positive*

First, I address the distinction between negative and positive emotions. A simple

mode of successful interaction with uncertainty would be an interaction that succeeded in eliminating the uncertainty, perhaps by leaving or by altering the situation. Notice, however, that the situation that produces the uncertainty is not identical to the situation that the organism interacts with — the organism is interacting with its own uncertainty in addition to the external environment per se. If the response to that uncertainty is more uncertainty — uncertainty about how to deal with the uncertain situation — then the overall uncertainty increases. Uncertainty can create anticipations of more uncertainty. A runaway feedback of uncertainty creating more uncertainty is a kind of panic attack, and is a paradigm for a negative emotion.

On the other hand, suppose that the situation is uncertain in the sense that no particular interactions are already known to succeed in this kind of situation, but that the general *kind* of uncertain situation is well known in the sense that procedures are known that tend to reduce or eliminate *this* kind of uncertainty. I don't know how to solve this math problem, but I do know how to go about figuring out how to solve it. If successful interactions tend to be stabilized, and if resolution of uncertainty is a successful interaction (which it is by the model as developed so far), then uncertainty situations in which there is anticipation of resolution of that uncertainty should be stabilized in learning. Uncertainty for which there is strong anticipation of resolution is the model for positive emotions.

The distinction between negative and positive emotion, then, turns on the anticipations involved about the potentialities for resolving the uncertainty. Situations of interactive uncertainty are of strong adaptive importance, and anticipations of success or failure in resolving such uncertainty are constitutive of the positive or negative character of that importance. Further differentiations of kinds of emotions will occur depending on what sorts of categorizations of uncertain situations are learned and what kinds of interactive styles come to be associated with them.

4.2 *Biological, developmental, and social aspects of emotions*

It would make adaptive sense, in this view, for evolution to have created innate supports for some basic uncertainty response styles, for some basic emotions (Ekman and Davidson 1994), but it does not follow in this view that all emotions would be blends of such basic emotions. Learning has full power to develop further differentiations of emotion situations and emotion interactive processes associated with them, including some that will be largely culturally specific (Harré 1986). It would also make adaptive sense, in this view, for emotional expression and emotion recognition, at least in complex social species, to be strongly involved in social interaction and social cognition (Ekman 1984; Ekman and Davidson 1994), though, again, it does not follow that these functions would

constitute the most fundamental ontology of, or adaptive reason for, emotions.

Modeling the typical developmental differentiations of emotions should, in this view, capture the development of more and more refined forms of uncertainty situation categorization, response styles, and regulation skills (Gross 1998), beginning with a relatively undifferentiated arousal (Scherer 1984; Thayer 1989). After an initial differentiation of positive and negative, negative arousal seems to differentiate into fear and anger, and so on (Harlow and Mears 1983; Saarni, Mumme, and Campos 1998; Sroufe 1984, 1995). Later emotional possibilities emerge with the capability for reflexive consciousness at about age four. Reflexivity is possibly involved in such emotions as guilt (Taylor 1987).

4.3 *The space of affectivity*

Emotions are interactive processes with anticipations of uncertainty about successful interaction with regard to some particular situation. That is, there is generally a cognitive focus for emotions (Nissenbaum 1985). There is no constraint in the model, however, to prevent uncertainty about successful interaction, and anticipations or lack thereof concerning the resolution of uncertainty, to occur more globally, without any particular focus. Such unfocused 'emotional' processes provide a potential model for moods (Rosenberg 1998).

Emotions as designated in English are occurrent phenomena. A readiness or propensity to experience some particular emotion might be characterized as a personality style if it is generic to multiple kinds of situations, and a mood if it is relatively continuously ongoing, but, if it is tied to specific cognitive foci, we tend to describe it as an attitude — a propensity to have particular emotional reactions to particular kinds of objects or situations. The emotions model, then, yields rather readily candidate models of moods and attitudes.

The space of processes and dispositions that is differentiated by the occurrent and non-occurrent distinction and by the focus and unfocused distinction is a relatively continuous space, not a pair of dichotomies. Depression, for example, is relatively ambiguous between mood and emotion, while we at times refer to emotional dispositions — non-occurrent — as personality characteristics or styles: an angry person, for example, or an angry mood, even if not at all angry at this moment.

4.4 *Some emergent motivations*

Successful forms of interaction will be learned and will be sought. This includes successful forms of emotional interaction. Positive emotions, then — interactions with forms of uncertainty situations for which there is strong anticipation of

resolution — will be sought, and, therefore, situations that are expected to yield positive emotions will be sought.

The expectations of resolution of uncertainty, or the lack thereof, are learned just as much as the uncertainty categorizations and response styles per se. Positive and negative emotional stances toward particular objects, then, are not necessarily responses to intrinsic characteristics of phenomena. One person, for example, may learn that mathematics problems pose an interesting challenge that is fun to address, while another may learn that the same problems offer only further frustration and failure. Similarly, new learning may allow bringing new forms of exploration to an object — new forms that offer new resolvable uncertainty: Toddlers sometimes like to play with grass, picking it and tossing it, and so on, but the novelty soon wears off. Later, however, that same toddler might become a botanist and discover many new ways in which grass can be fascinating.

Intrinsic characteristics of an object or phenomenon, however, can limit the novelty that it can offer. Nursery rhymes relatively quickly lose their interest to an adult. But others can offer essentially unlimited novelty — there is always something new to hear and experience in Beethoven's Ninth or avant-garde jazz.

Learning to seek such experiences constitutes learning a kind of process selection, and, thus, a kind of motivation. We name these variously as competence motivation, mastery motivation, or esthetic motivation. These are emergent kinds of motivation, emergent from the inherent dynamics among interaction, learning, and emotions.

Some other motivational phenomena are also emergent in these dynamics. For example, as mentioned above, the living system is always active, always doing something. If sufficiently driven by inputs that require full resources for successful interaction, such as pain or hunger, those forms of interaction will dominate. If such 'external' driving of the central nervous system processes is minimal or absent, the processes do not simply cease. They continue, and continue to seek forms of successful interaction, including uncertainty interaction. The individual will seek situations and objects that offer resolvable uncertainty. Exploration, curiosity, and esthetics are examples of the kind of motivational phenomena that emerge if not displaced by more demanding forms of process (Maslow 1962).

Furthermore, such explorations of what is most satisfying will tend to discover and emphasize not only what provides the greatest opportunity externally, but also what fits best with prior kinds of talents and experience in the individual. That is, such explorations will *tend* to develop the potentialities of the person, so long as they are not precluded or blocked by more demanding forms of process. Such a tendency to actualize the potentialities of the person is sometimes referred to as a motivational process itself (Csikszentmihalyi 1990; Holdstock and Rogers 1977; Maddi 1996; Mook 1996), but it is not so much a

direct matter of selection of further activity as it is an emergent tendency of consequences of such selections.

4.5 *A few comparisons*

The model outlined here is a dynamic model, based on a recognition of the necessarily open dynamics of any living system. Emotions are a particular kind of dynamics — forms of interaction with the system's own internal dynamical uncertainty about how to proceed and how to anticipate the interactive flow. Emotions are, in this view, an adaptation to a basic informational property of the organism-environment relationship — uncertainty — and, as such, manifest their own adaptive rationality (de Sousa 1987; Lazarus 1991). The effects of emotional processes are, of course, not always beneficial, but representation and motivation too can be in maladaptive error.

The uncertainty that gives rise to emotion processes is a kind of evaluation (Frijda 1986; Oatley 1992), but it is not an evaluative process that is independent of, or follows on, the interactive representational processes. Instead, it is an aspect of the flow of representational and motivational interaction. The differences between this model of evaluation and notions of evaluation in alternative models turn largely on the difference between interactive and encodingist models of the nature of cognition and representation. If representation is constituted as encoding elements, then setting up or activating such elements in perception and cognition will necessarily be distinct from evaluating and judging the situation thus represented. In particular, this model is in stark contrast to models of emotion as particular kinds of propositional attitudes (see Griffiths 1997, for a discussion).

The model is consistent with strong biological supports for some basic kinds of emotions — evolution is likely to have scaffolded the development some of the most important general forms of uncertainty interaction — but it is also consistent with a ubiquitous involvement of social and cultural learning in emotions, and even the social and cultural ontology of some of them in which the basic categorizations of situations are themselves inherently socially constituted. In this, the model is closer to the dynamic and developmental framework endorsed by Griffiths (1999), for example, than the emotional programs notion in Griffiths (1997). Emotional expressivity in social species should, in this view, be expected to be of basic importance to the character and regulation of social interaction, but, again, constitutes neither the basic ontology of emotions nor their most basic adaptive function.

5. Conclusions

The model of emotions outlined here makes sense only on the foundation of the interactive model of representation and motivation. It is not possible to develop the intrinsic notions of evaluation of uncertainty with the same properties in an encodingist cognitive framework. A primary moral of the model, then, is that such phenomena cannot be approached independently of one another: in this case, assumptions about cognition have major implications and impose major constraints on models of emotion.

Within the model, representation, motivation, and emotion are all aspects or kinds of interaction. They are integrated in an intimate way that is necessarily fragmented in encodingist models. In this integration, the model makes contact with multiple facets of emotions research and theory, such as biological bases for 'basic' emotions, developmental aspects of emotions, the social construction of emotions, and the importance of emotional expressivity and recognition, without reifying any particular such facets into the ontology of emotion.

The larger framework of the model is a dynamic systems model of living beings as far from thermodynamic equilibrium systems (Bickhard 1993, 1998b). As such, the model makes contact with other dynamic systems approaches (Port and van Gelder 1995; Thelen and Smith 1996), but without ignoring representation (Bickhard, in press-b). Persons are complex dynamic open systems with multiple emergent properties, such as representation, motivation, learning, emotions, consciousness, language, and so on, and will not be understood without honoring that fundamental dynamic nature.

Notes

1. Another representational challenge concerns representations of abstractions, such as of numbers. A similarly Piagetian model accounts for such kinds of representations, but requires additional elaborations of the model that I will not pursue in this chapter (Campbell and Bickhard 1986).
2. A number of additional properties, such as those of qualia, require (I argue) a model of reflexive consciousness, in addition to simple conscious awareness. I will not address those issues here (Bickhard 1980, 1998a; Campbell and Bickhard 1986).
3. Hunger signals are a form of vicariant or surrogate for the maintenance of the biological integrity of the organism (Brown 1990; Campbell 1974; Christensen 1996; Christensen, Collier, Hooker, in preparation). Such vicariants — e.g., hunger, thirst, pain, and so on — are fundamental to successful interacting: no organism can calculate, even heuristically, back to the basic criterion of biological integrity, and must, therefore, depend on such surrogates. I will not focus on these points in this chapter, though the general nature of the functioning of a few of them are indicated in passing.

References

Bickhard, M. H. 1980. "A Model of Developmental and Psychological Processes". *Genetic Psychology Monographs* 102: 61–116.

Bickhard, M. H. 1991. "The Import of Fodor's Anti-Constructivist Argument". In Les Steffe (ed), *Epistemological Foundations of Mathematical Experience.* New York: Springer- Verlag, 14–25.

Bickhard, M. H. 1993. "Representational Content in Humans and Machines". *Journal of Experimental and Theoretical Artificial Intelligence 5*: 285–333.

Bickhard, M. H. 1996. "Troubles with Computationalism". In W. O'Donohue, R. F. Kitchener (eds), *The Philosophy of Psychology*. 173–183. London: Sage.

Bickhard, M. H. 1997. "Is Cognition an Autonomous Subsystem?" In S. O'Nuallain, P. McKevitt, E. MacAogain (eds), *Two Sciences of Mind*. 115–131. Amsterdam: John Benjamins.

Bickhard, M. H. 1998a. "Levels of Representationality". *Journal of Experimental and Theoretical Artificial Intelligence 10*(2): 179–215.

Bickhard, M. H. 1998b. "A Process Model of the Emergence of Representation". In G. L. Farre, T. Oksala (eds) *Emergence, Complexity, Hierarchy, Organization*, Selected and Edited Papers from the *ECHO III Conference*. *Acta Polytechnica Scandinavica*, Mathematics, Computing and Management in Engineering Series No. 91, Espoo, Finland, August 3–7, 1998, 263–270.

Bickhard, M. H. 1999. "Interaction and Representation". *Theory & Psychology 9*(4): 435–459.

Bickhard, M. H. In press-a. "The Emergence of Contentful Experience". In T. Kitamura (ed), *What Should be Computed to Model Brain Functioning?* Singapore: World Scientific.

Bickhard, M. H. In press-b. "Dynamic Representing and Representational Dynamics". In E. Dietrich, A. Markman (eds), *Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines*. Mahwah, NJ: Erlbaum.

Bickhard, M. H., Campbell, R. L. 1996. "Topologies of Learning and Development". *New Ideas in Psychology 14*(2): 111–156.

Bickhard, M. H., Terveen, L. 1995. *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution.* Amsterdam: Elsevier Scientific.

Bogdan, R. 1988. "Information and Semantic Cognition: An Ontological Account". *Mind and Language 3*(2): 81–122.

Brown, T. 1990. "The Biological Significance of Affectivity". In N. L. Stein, B. Leventhal, T. Trabasso (eds), *Psychological and Biological Approaches to Emotion*. 405–434. Hillsdale, NJ: Erlbaum.

Campbell, D. T. 1974. "Evolutionary Epistemology". In P. A. Schilpp (ed), *The Philosophy of Karl Popper*. 413–463. LaSalle, IL: Open Court.

Campbell, D. T. 1990. "Levels of Organization, Downward Causation, and the Selection-Theory Approach to Evolutionary Epistemology". In Greenberg, G., and Tobach, E. (eds), *Theories of the Evolution of Knowing*. 1–17. Hillsdale, NJ: Erlbaum.

Campbell, R. L., Bickhard, M. H. 1986. *Knowing Levels and Developmental Stages.* Contributions to Human Development. Basel, Switzerland: Karger.

Christensen, W.D. 1996. "A Complex Systems Theory of Teleology". *Biology and Philosophy* 11: 301–320.
Christensen, W.D., Collier, J.D., Hooker, C.A. In preparation. "Autonomy, Adaptiveness, Anticipation: Towards Autonomy-theoretic Foundations for Life and Intelligence in Complex Adaptive Self-organising Systems".
Clark, A. 1997. *Being There*. Cambridge, MA: MIT/Bradford.
Csikszentmihalyi, M. 1990. *Flow: The Psychology of Optimal Experience*. New York: Harper.
Cummins, R. 1996. *Representations, Targets, and Attitudes*. Cambridge, MA: MIT.
de Sousa, R. 1987. *The Rationality of Emotion*. Cambridge, MA: MIT.
Douglas, G. 1998. "Why Pains Are Not Mental Objects". *Philosophical Studies 91*: 127–148.
Dretske, F.I. 1988. *Explaining Behavior*. Cambridge, MA: MIT Press.
Eccleston, C., Crombez, G. 1999. "Pain Demands Attention: A Cognitive-Affective Model of the Interruptive Function of Pain". *Psychological Bulletin 125*(3): 356–366.
Ekman, P. 1984. "Expression and the Nature of Emotion". In K.R. Scherer, P. Ekman (eds), *Approaches to Emotion*. 319–343. Hillsdale, NJ: Erlbaum.
Ekman, P., Davidson, R.J. 1994. *The Nature of Emotion*. Oxford: Oxford University Press.
Fodor, J.A. 1981. "The Present Status of the Innateness Controversy". In J. Fodor *RePresentations* 257–316. Cambridge: MIT Press.
Fodor, J.A. 1987. "A Situated Grandmother?" *Mind and Language 2*: 64–81.
Fodor, J.A. 1990. *A Theory of Content*. Cambridge, MA: MIT Press.
Frijda, N.H. 1986. *The Emotions*. Cambridge: Cambridge University Press.
Griffiths, P. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: U. of Chicago.
Griffiths, P. 1999. Author's Response. *Metascience 8*(1): 49–62.
Gross, J.J. 1998. "The Emerging Field of Emotion Regulation". *Review of General Psychology 2*(3): 271–299.
Harlow, H.F., Mears, C.E. 1983. "Emotional Sequences and Consequences". In R. Plutchik, H. Kellerman (eds) *Emotion: Theory, Research, and Experience*. 171–197. New York: Academic.
Harré, R. 1986. *The Social Construction of Emotions*. Oxford: Basil Blackwell.
Holdstock, T.L., Rogers, C.R. 1977. "Person-Centered Theory". In R.J. Corsini (ed). *Current Personality Theories*. 125–151. Itasca, IL: Peacock.
Lazarus, R.S. 1991. *Emotion & Adaptation*. Oxford: Oxford University Press.
Loewer, B., Rey, G. 1991. *Meaning in Mind: Fodor and his critics*. Oxford: Blackwell.
Maddi, S.R. 1996. *Personality Theories: A Comparative Analysis. 6th Ed.* Pacific Grove, CA: Brooks/Cole Pub. Co.
Maslow, A.H. 1962. *Toward a Psychology of Being*. Princeton, NJ: D. Van Nostrand.
Millikan, R.G. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
Millikan, R.G. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Mook, D. G. 1996. *Motivation: the organization of action*. 2nd ed. New York: W. W. Norton.

Nissenbaum, H. F. 1985. *Emotion and Focus*. Stanford: CSLI.

Oatley, K. 1992. *Best Laid Schemes: The Psychology of Emotions*. Cambridge: Cambridge University Press.

Piattelli-Palmarini, M. 1980. *Language and Learning*. Cambridge, MA: Harvard.

Port, R., van Gelder, T. J. 1995. *Mind as Motion: Dynamics, Behavior, and Cognition*. Cambridge, MA: MIT Press.

Rosenberg, E. L. 1998. "Levels of Analysis and the Organization of Affect". *Review of General Psychology* 2(3): 247–270.

Saarni, C., Mumme, D. L., Campos, J. J. 1998. "Emotional Development: Action, Communication, and Understanding". In W. Damon, N. Eisenberg (eds), *Handbook of Child Psychology. 5th ed. Vol. 3: Social, Emotional, and Personality Development*. 237–309. New York: Wiley.

Scherer, K. R. 1984. "On the Nature and Function of Emotion". In K. R. Scherer, P. Ekman (eds), *Approaches to Emotion*. 293–317. Hillsdale, NJ: Erlbaum.

Sroufe, L. A. 1984. "The Organization of Emotional Development". In K. R. Scherer, P. Ekman (eds), *Approaches to Emotion*. 109–128. Hillsdale, NJ: Erlbaum.

Sroufe, L. A. 1995. *Emotional Development*. Cambridge: Cambridge University Press.

Staddon, J. E. R. 1983. *Adaptive Behavior and Learning*. Cambridge: Cambridge University Press.

Taylor, G. 1987. *Pride, Shame, and Guilt*. Oxford: Oxford University Press.

Thayer, R. E. 1989. *The Biopsychology of Mood and Arousal*. Oxford: Oxford University Press.

Thelen, E., Smith, L. B. 1996. *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT.

CHAPTER 9

Mind, Brain, and Chaos

Nicholas Georgalis
East Carolina University

Most work in cognitive science, whether computationally or biologically based, relies on the idea of 'unconscious representations.' I argue that there are no such things. In contrast, conscious states do represent objects outside of the agent. My second main thesis concerns the 'explanatory gap' between brain states and mental states. I propose an analogy to enhance our understanding of just what is required to close the explanatory gap; I argue that a new science is needed. There is third main thesis, related to the second. Of all the currently available approaches to the study of brain activity, only nonlinear dynamics, chaos theory in particular, has a unique characteristic sufficient to close the explanatory gap. Every other approach to the study of the brain appears doomed to failure in this regard. Identification of an additional feature of chaotic systems provides the framework for the explanation of another conundrum, mental causation. Furthermore, given the chaotic model of the dynamics of perception arising from Walter J. Freeman's work, it appears that emotion plays a central role in perception. The application of chaos theory to brain activity dovetails with my first thesis, as nonlinear dynamics does not require state transitions from representations to representations, yet it has the resources to account for the representational nature of conscious states.

1. Unconscious states are nonrepresentational

There is an innocent and straightforward sense of 'unconscious beliefs' such that no one, I take it, would deny that at any given moment, each of us has a vast number of them. I have in mind quite ordinary beliefs that each of us has but is not currently entertaining. All non-occurrent beliefs that one has are unconscious — beliefs that could be expressed by sentences such as 'Gold is a metal,' and 'Two is the only even prime number.' Just as clearly, any such belief is one that

we could entertain and therefore, at times, be conscious of. Such beliefs easily pass from being unconscious to conscious, and back. Being unconscious in this sense does not make them deeply hidden; they are not Freudian.

Thus understood, there clearly are unconscious beliefs, and each *in some sense* is identical to its conscious manifestation. What identity there is between a conscious and an unconscious belief is secured by the *sentence* expressing the belief, which just means that the *sentence* expressing the belief is the same whether the belief is conscious or unconscious. But not everything we say of conscious beliefs is true of unconscious beliefs because their realizations are different.

Part of what is at issue here is just what it is to *have* a belief, and whether one *has* conscious and unconscious beliefs in the same way. I take it that when one "has" a belief, be it conscious or unconscious, it is realized in *some* way, be it neurobiological, computational, individuated somehow by its causal connections to the environment, or something else. We have every reason to expect that, however a belief is realized, its conscious and unconscious realizations are different. (I will here assume the realizations are neurobiological.)

An objection one might raise to my claim that conscious and unconscious realizations of the same belief are different might go as follows: although the realizations are unlikely to be identical, there is a "common component" to both, one that constitutes or corresponds to the belief itself. They are not quite identical, however, because a conscious belief obviously requires something additional. For example, if one's unconscious belief, **p,** is realized by some brain state, then one's conscious belief **p** is realized by that very same brain state plus others, or that brain state enters into different relations with other brain states when there is a conscious state; these other states or relations do not obtain when the belief is unconscious.

The main thing wrong with this kind of move is that it is committed to a "storehouse" view of unconscious beliefs, as if each and every belief one has, including the unconscious beliefs, is somehow stored intact in one's head. But at any given moment, one *has* indefinitely many, perhaps infinitely many,[1] unconscious beliefs and, according to this view, distinct states for each one must somehow be realized. As is well known, this is at best highly implausible. Better to reject the idea of a common realization, even partial, and think more along dispositional lines, in parallel with various dispositional physical properties. For example, that a glass *is fragile* is realized in its molecular structure. Importantly, the very same molecular realization of its fragility is *simultaneously* a realization of indefinitely many other dispositional properties. If the molecular structure that realizes a glass's fragility were exposed to extreme heat or an acid, rather than sharply struck, the molecular structure that would result, the realization of the new property, would be different from that which would result from its being

struck. So the molecular structure that realizes the glass's fragility also realizes all its other dispositional properties. Countless dispositional properties are realized in the original molecular structure; which of these is actualized depends on what conditions that self-same molecular structure is subjected to.

The model for unconscious belief advanced here, more generally, unconscious mental states, is analogous: One and the same, say, brain state is *simultaneously* a realization of any number of unconscious mental states. Given that brain state, various conscious states may subsequently be realized; which is realized depends on how the agent is prompted. The parallel to dispositions is not exact but only a first step. Ultimately, the model for unconscious mental states is based on chaos theory and, therefore, unlike dispositions, is not stimulus driven. (See Section 3.)

The neuronal realizations of unconscious mental states should be further distinguished from the broader class of neuronal states that have nothing to do with cognition, those involved in, say, digestion or respiration. The value of categorizing some non-conscious states as 'unconscious *mental* states' is to indicate the existence of relations between them and various conscious mental states, relations that other non-conscious states do not share. In spite of these relations, my position is that *all* unconscious mental states are *non*-intentional, *non*-representational.

I contrast my view with John Searle's, since he holds that unconscious beliefs are representational.[2] Searle's argument is straightforward; it relies on Leibniz's Law.[3] One of John's beliefs is that asparagus is a vegetable. It is the same belief that John has, whether John is consciously entertaining it or not, say, he is asleep, otherwise unconscious, or conscious but thinking about other things entirely. When in any of these latter states, it is correct to say he still has the self-same belief that asparagus is a vegetable. His conscious belief is intentional. Therefore, *this* belief when unconscious is intentional.

The appeal of this argument turns on the reification of beliefs. While it is correct to hold that one *has* the same belief, whether it is consciously or unconsciously had, it is not some identical *thing* that one has on either occasion. It is not some little 'nugget' tucked away somewhere in the brain. The dispositional model of beliefs advanced above undercuts treating belief as a thing. Thus applying Leibniz's Law as a basis for an objection to my view fails, since the realization of conscious and unconscious beliefs is different. For convenience we may speak of the 'same belief' consciously or unconsciously had on different occasions, but it is not the self-same *thing* in the agent. To say, as above, that John has the belief that asparagus is a vegetable when not consciously entertaining it is to make a pragmatic claim about what John would do or say when prompted in various ways. How the belief is "had" is explicated as how it is realized, and conscious and unconscious realizations may well have different properties.

Now the representational nature of conscious beliefs plays a clear role in the analysis of opacity of conscious beliefs. But it is not at all clear how representation could be involved in talk of *unconscious* beliefs realized in the brain.[4] Recall, on my view, a single sentence expresses both the conscious and the unconscious belief, and it is the sentence that bears the burden of the identity. Any opacity questions that may arise regarding unconscious mental states are reducible to the *sentences* that express those unconscious states; they just do not apply to the realizations of the unconscious mental states.

Even if Searle's argument were sound, however, it would provide no help as to *how* unconscious beliefs or mental states *could be* intentional. There appear to be enormous conceptual and methodological problems in answering this "how" question. Searle himself denies that the physiological facts are themselves sufficient to infer aspectual shape, even though on his view belief states are caused by and realized in brain states. (The aspectual shape of a mental state is very closely related to its representational nature. One might hold a belief under one representation, but disavow it under an extensionally equivalent representation.) He argues that no matter how complete the behavioral or even neurophysiological evidence, there would still be an inference from these facts to the aspectual facts. Behavioral or neurological facts cannot, he says, constitute aspectual nor intentional facts.[5] Given this, it would seem preferable to avoid this how question altogether. Obviously, the question is avoided on my account, since unconscious beliefs do not have aspectual shape, are not intentional, are not representational.

Why should we avoid this question? One reason relies on what Searle himself has argued, and just cited: behavioral or physiological facts cannot themselves constitute intentional facts. Unfortunately, this point wipes out too much, as it would seem to pose an equally formidable problem of explaining how *conscious* states could be intentional or representational. Though nothing I have yet said suggests a solution to this, I will address it later (Sections 2 and 3). It turns out that this is another version of the explanatory gap problem between brain and phenomenal states, discussed below. We will see that my application of a certain model of brain activity, supported by empirical evidence, provides the framework for answers to these questions. But even at this stage, there are grounds for treating conscious and unconscious states differently with regard to their representational nature.

There is strong evidence that some conscious states are representational, and there is no like evidence that any unconscious states are representational. Hardly anyone would dispute that at least some conscious mental states are representational. I submit that the reason this is so compelling is that it is evident from the first-person perspective that we each enjoy regarding some of our own states.[6] There is absolutely no corresponding datum, from neither the first- nor third-

person perspective, regarding the representational nature of unconscious states. (Computational and biological models take for granted 'unconscious representations.' They do not offer direct evidence for such. To the extent that they are argued for at all, it is based on whatever success the models that employ them enjoy.) So representational unconscious states are at least initially problematic.

It is generally assumed in cognitive theories that a system's successful adaptive behavior is dependent on its having inner representations of its environment. Just how these inner representations are explicated varies considerably from theory to theory, but it is almost universally held that adaptive systems have representations and that they need not be conscious of them. Suppose, however, that the production of adaptive behavior can be explained by, say, nonrepresentational brain activity. This would undercut these appeals to a system's successful environmental behavior as support for unconscious representations. It would also be a neurophysiological basis supportive of my claim that unconscious beliefs are not intentional, not representational.

Such a neurophysiological model exists in the work of Walter Freeman.[7] Discussing that work, Christine Skarda argues that "... the patterns of neural activity responsible for behavior do not 'represent' anything, that brains do not 'read' them, and that 'neural representations' need not play a role in the production of behavior in animals." (1987: 189) She maintains that Walter Freeman's research on the olfactory bulb disconfirms the assumption that inner representations of the environment are required for adaptive behavior.[8] Surely, these are controversial claims; I will return to them in Section 3. My purpose in citing them is to remind us that there is an active empirical research program that denies the behavior of a system is dependent on the system's neuronal states *representing* the environment. So, though my contention that there are no unconscious representations is unorthodox, it is not without both philosophical and empirical support.

Consider the fact that we humans clearly do represent items in our environment. Do we not frequently do so unconsciously? For example, when we negotiate our environment to avoid obstacles without being conscious of them, does not this require *unconscious representations* of those obstacles and their relative positions to oneself?

Again, I can only suggest an answer here,[9] and I do so by appeal to the work of Kathleen Akins (1996). She distinguishes the *ontological* from the *sensory-motor project*. She argues that the former does involve representations of stable objects in the environment but the latter does not. After a careful examination of the neurophysiology involved in the sensory-motor project, she concludes that there is a rather large "... gap between the needs of the sensory-motor project and the demands of the ontological project ...". (1996: 370) The gap, she rightly insists, requires explanation. "... [H]ow exactly does the information

provided by our sensory systems co-exist with, form a whole with, the ontology imposed by a representational system?" (1996: 370–371) She concludes: "Trace out the causal path between the object of perception, the stimulation of the receptors, and whatever neural events that thereafter eventuate [sensory-motor project] and this alone will not *explain*, in the required sense how *genuine representation* arises [ontological project]." (1996: 372, last emphasis added.)

The idea is that whatever occurs in the sensory-motor system, there is no straightforward interpretation of it such that it represents the stable objects in the environment, the objects of which we are conscious. There are no reliable correlations between types of physiological states and kinds of stimuli originating in the agent's environment. One might object that this does not show that the states of the sensory-motor system *themselves* are not representational. It could be maintained that these states *do* represent; it is just that *what* is represented is different from what is represented in the ontological project and at the conscious level. Of course, if Freeman is right, this objection is a non-starter. (As we will see in section 3, Freeman holds that the stimulus is "washed out".) But as I have yet to give sufficient details of this position, let us examine other ways of meeting this objection.

The objection itself is more powerful against me than it is against Akins, since all she argues is that certain traditional approaches to naturalizing aboutness that take the sensory-motor system as the basic aboutness relation will fail; whereas I maintain the stronger thesis that unconscious states of such a system simply do not represent. Period. Nevertheless, if one cannot get to the ontological project from the sensory-motor project, as Akins' work suggests, then at the very least, if "representations" are involved at the sensory-motor level, *they are of a very different sort than that at the conscious level.* And that may be all I need. Since what one obtains in pursuit of the sensory-motor project is not sufficient for the ontological project, it is not sufficient for genuine representation.

Now talk of 'genuine representation,' as Akins and I have, always raises suspicions as to what is so "genuine" about one's favorite analysis. But the point can be made innocently enough. What is important is not the word used, but the fact that "representations" in the sensory-motor project appear, at best, to be significantly different from those in the ontological project, and there seems to be no way of going from the former to the latter, for they are not reliably correlated. If the gap and the significant differences between the two senses are real, then the common practice of using the same term to cover both is a dubious one; so, the term 'representation' is better reserved for just one of these senses.

I have argued for an analysis of unconscious beliefs that allows them no intentionality. This stands apart from my more recent claim as to the nonexistence of unconscious representations. I have only offered considerations on behalf of this bolder claim here. But If I am right on all this, it would go far in

explaining why there has been so much success in studies of cognition that ignore consciousness, even if consciousness is necessary for representation. These studies succeed when they focus on unconscious processes. It is a grave error, however, and one that is frequently made, to infer from such successes that results pertaining to unconscious "beliefs" or "representations" are transferable to conscious beliefs and representations, or that consciousness itself is epiphenomenal and not central to representation and cognition. Such errors stem from the deeply flawed, but widely held, assumption that the sense of *belief* and *representation* employed in discussions of conscious and unconscious states is the same. If I am right, this assumption is false.

2. Physicalism and the explanatory gap

Physicalism comes in a variety of forms. Common to all such forms is the view that everything is physical, and there is a physical explanation of every fact. Just what 'physical' means in such formulations varies. I will recommend below one account that draws its inspiration from a well-known episode in the history of physics.

It is widely agreed that our phenomenal states pose a particularly "hard problem" for physicalism, however construed, since there seems to be an "explanatory gap"[10] between, say, brain states and our phenomenal experiences. I will discuss an important solved analog to the explanatory gap problem. It provides us with a concrete understanding of just what it is we *ought to seek*, when we seek to close the explanatory gap. This case also demonstrates that an entailment relation is not necessary and so, contrary to some, need not be sought.

My account of physicalism and how the gap problem should be closed preserves the explanatory gap as the serious problem that it is, but shows that it poses no serious threat to physicalism, properly understood. Still, I will show that physicalism must avoid certain evasive, "ostrich" strategies, and that any appeal to fundamental mysteries also should be avoided. I also generalize the problem in that there is a gap not only between brain and phenomenal states, but also between brain and contentful mental states such as beliefs and desires.

What is the close analogy to the explanatory gap problem that has been solved? Consider an investigator at a relatively early stage of chemistry, perhaps a fictitious stage: Suppose it is known that H_2O molecules constitute water but little is known about intermolecular bonds. Molecules are (roughly) *discrete entities*, yet water presents itself as a *continuous and flowing stuff*, a liquid. When I use the term 'liquid' here, I do not mean it in its technical, physio-chemical sense; rather, it should be understood as a straightforward observation term. It is what a speaker of English would understand by the term, even if entirely

ignorant of chemistry and physics. It is this sense of the term that is relevant, if the analogy is to work.[11] That is, the gap in the water case is between its discrete molecular constitution and its ordinarily observed continuous flow. Being so disparate in nature, how could the former give rise to the latter? We can understand this once we know that H_2O molecules can form weak intermolecular bonds, such that the bonds between large clusters of them can easily break and reform with other large clusters, the clusters themselves changing over time, thereby "sliding," as it were, across one another.[12] Furthermore, it is a familiar fact that small discrete entities in near proximity to one another may give the appearance of continuity, especially when viewed from a distance. Thus is dispelled the "mystery" that justifiably exists concerning the gap between the liquidity of water, its observed continuity and ability to flow, on the one hand, and its discrete molecular constitution, on the other.

Note that the actual theory of molecular bonding is quite complicated, the success of the explanation, however, does not turn on any arcane knowledge of that theory. *Only some of the simplest general features of the theory,* coupled with continuity considerations, *are needed to provide an adequate account of the liquidity of water.* With that slim knowledge, what is mysterious becomes *comprehensible to us.*

We should also note that this account is *not* tantamount to simply saying that liquidity *supervenes* on a certain type of molecular bonding, though it may be doing that too. Mere talk of supervenience without appeal to, say, intermolecular bonds and continuity considerations would be *just* to name the mystery; it would not provide a plausible "mechanism" for the supervening property.[13] While many different properties may 'supervene' on various sets of base properties, mere statement of a supervenience relation between the sets of properties is nearly vacuous. One must also state how the base properties *determine* the supervenient ones. The importance of specifying the dependence relation becomes all the more evident once it is realized these dependence relations will certainly vary from case to case. Thus, though weak intermolecular bonds may constitute the dependence relation in the molecule/liquid case, no one would expect that same dependence relation operative in the brain/mental state case.[14]

The gap problem between brain and mental states is so difficult not only because no one has identified any plausible mechanism connecting them that enables us to understand how the one *could* give rise to the other but, relatedly, because the features of the former are so radically different from the latter. A solution to the gap problem requires the *specification* of some "mechanism," something that could play a similar role to the intermolecular bonds in the molecule/liquid case. Merely saying that mental states supervene on brain states is almost a vacuous claim; it certainly advances no understanding.

Hypothesis: some unknown entities, states, processes, or features, **X**, of the brain are to phenomenal or mental states as intermolecular bonds are to liquidity. There may be no such **X**, but any physicalist research program, one of whose goals is to solve the explanatory gap problem, must seek such an **X**.

For definiteness, I will assume throughout that a solution lies in neuroscience. Whether current neurophysiological theories and approaches are sufficient to uncover such an **X**, if it exists, is an empirical question. With this caveat in mind, there seem to be three options:

1. Current neurophysiological resources are generally sufficient, though certain relevant details or connections among them have yet to be determined.
2. The current resources are not adequate — a scientific revolution in neurophysiology, of the kind Kuhn has described in other fields, is required.
3. There is no such neurophysiological **X**.

The current intractableness of the problem of the explanatory gap, while not conclusively establishing that option 1 is a dead end, does make it implausible. The analogy provides the reason for saying this: the solution to the molecule/liquidity gap did not rely on any technical or detailed knowledge of the character of intermolecular bonds, *but only on some of their simple most general aspects*. But while we have a rather good understanding of the firing rates of neurons and the various chemical neurotransmitters, we still *do not have a clue how such features could give rise to phenomenal qualities*. If specialized knowledge is not required and we already have the essentials of what we need, the gap problem should have been solved already, or at least the outline of its solution should be at hand. But it is not, and we do not even have a plausible line on it. Some exploit this recalcitrant fact to the extent of claiming that it shows physicalism is false, that is, they endorse option 3, or some generalized version of it, maintaining that it is an unsolvable mystery, beyond our capabilities. We shall see that such a conclusion is premature. For although option 1 currently seems unlikely, that is not the only way physicalism could be saved and the explanatory gap problem solved. There is still option 2–we need a new science.[15] For the reasons cited above, I reject option 1. (Later, I will give a more conclusive reason why option 1 is not tenable.) I further hold that, given a plausible understanding of physicalism, we must reject option 3. To see this consider first a simple-minded view of Physicalism, one no one would take seriously today. It holds that all things, events, states, or processes are explainable within a Newtonian framework. Clearly, physicalism thus understood is false. No contemporary would be tempted by such a narrow physicalism, even so, such a view would have been, indeed was, held by many in the late eighteenth and first half of the nineteenth century.

The history of this case is well known. A relevant problem then was to

account for electromagnetic phenomena in Newtonian terms. (I note that I am not advancing this as a gap problem, though one might; I exploit it for its bearing on the concept of physicalism.) All such efforts failed. The Newtonian laws required supplementation with those of Maxwell. Was this a refutation of physicalism? No one, I think, would say so, not even then. Though it was a defeat for a physicalism narrowly understood in Newtonian terms, the concept was enlarged. To my mind, this was a rational reaction. Newtonian mechanics enjoyed huge successes and held out tremendous promise in its early years, but subsequent sustained scientific investigation determined that more had been promised than could be fulfilled; furthermore, the Maxwellian enhancement, which solved the target problems, was good science by all relevant standards. Subsequent scientific developments, the special and general theories of relativity and quantum electrodynamics, expanded our conception of the physical even more radically.

What bearing does all this have on ruling out option 3? The point is that any attempt to limit the concept of physicalism to our *current* successful scientific theories is a mistake. For among other reasons, it unjustifiably assumes that our current science has "got it essentially right." Not only is such a belief contrary to the inherent tentative nature of science, it is inductively unsound, given the various dramatic changes and upheavals in scientific theories over the years that has resulted in our expanded view of the physical.

Physicalism understood as restricted to current science is most probably false. If one adopts this restriction, then given the apparent hopelessness of closing the gap with the resources of current science, one is pushed toward option 3. I have attempted to show that such a restriction, however, is unjustified. If the exact character of physicalism is understood as characterized by our best science, not taken as an essentially finished product, but as ongoing and open-ended, then physicalism is very probably true. For thus understood, it is simply the stricture that we will not count any thing, process, state, or feature as explained until we have a scientific account of it, *whatever that is*. This makes physicalism a rather uncontroversial, indeed, trivial thesis, since all that is ruled out is non-scientific accounts, say, supernaturalism.

Given this conception of physicalism, endorsement of option 3, in its general interpretation, would be tantamount to giving up and declaring the explanatory gap to be an unsolve*able* mystery, as some have done. Such a stance bears a trace of arrogance in that it unjustifiably presupposes foreknowledge of what future science will bring, more accurately, what it will *not* bring. We *may* never have the requisite science to close the gap (though I think it is at hand, as I will explain in the next section); still, it is difficult to see why, at this stage of infancy of brain science we should draw the pessimistic conclusion that closing the gap is forever beyond our reach.[16]

A conception of physicalism constrained to some particular stage of scientific development, while more dramatic, is also far more problematic. Certainly, the restricted version to current science is quite dubious, since there is not even a single general and basic paradigm in neuroscience. Neuroscience has yet to produce its Newton or Einstein. What could decide in favor of such a narrow conception of Physicalism? Nothing short of its actual fulfillment and this, certainly, is not on the horizon.

Physicalism conceived as broadly as I have recommended, though so uncontroversial as to be trivial, has the virtue of shifting the attention to more productive discussions of the explanatory gap. Frequently physicalists and anti-physicalists alike view the explanatory gap as a threat to physicalism. The gap is only a threat if one adopts an unjustifiably narrow view of physicalism. Taking this narrower view, coupled with the current (apparent) hopelessness of the situation, results in a pernicious reaction to the explanatory gap by both sides. Physicalists tend to argue that the gap is not that big, that serious or even that it is an illusion, while antiphysicalists declare it is not solvable. If I am right, these defensive strategies on the part of physicalists, one might call them "ostrich strategies," are as misguided as the "mysterian" strategies of the antiphysicalist, for the narrow view of physicalism presupposed by both sides is unwarranted.

The explanatory gap is serious and — *at this time* — quite mysterious. That said, there is no reason to hide our heads in the sand or to be unduly and prematurely pessimistic. Brain science is, after all, still in its infancy. We have no reason to think that there can be no *new* understanding of the way the brain works that will close the gap between brain activity and mental states in a way analogous to how our understanding of intermolecular bonds closed the gap between discrete molecules and the fluidity of liquids. In short, to avoid 'ostrich' or "mysterian" strategies, we should endorse the wider conception of physicalism.

Some have rejected this line. For example, David Chalmers, has argued against the idea advanced here that we need a new science. (See his 1996: 118. See also Churchland 1996.) Chalmers' argument is that all physical theories come down to two basic elements: structure and dynamics of physical processes. "But from structure and dynamics, we can only get more structure and dynamics." (1996: 118) He concludes that new structure and dynamics will not help in closing the explanatory gap, since it is just more of the same.

Chalmers supplements this argument with two other requirements that he thinks advocates of a new science cannot fulfill. He says that it is difficult to evaluate the claim that it will take a new science in the abstract, "[o]ne would at least like to see an example of how such a new physics [science] might *possibly* go" (1996: 118). Furthermore, he holds that "[n]o set of facts about physical structure and dynamics can add up to a fact about phenomenology" (1996: 118).

A fair interpretation of the locution 'add up' is that of entailment, since on the very same page he raises the rhetorical question: "How could a theory that is recognizably a physical theory entail the existence of consciousness?" So the two further requirements are: A. a concrete proposal, or an indication of how the new science might go. B. The new science entails consciousness.

As to requirement A, I will present a specific proposal later. For now, I note that I have already indicated how a new science *could possibly go*. I did so by reminding us how science has gone with respect to a similar problem; namely, it uncovered an informative connection between the diverse elements of discrete molecules and continuous liquidity. But if one accepts requirement B, then one might protest that the analogy did not even accomplish that. For it does not close its own gap, since the structure and dynamics of molecules does not *entail* liquidity.

I am inclined to agree that the entailment relation is lacking in the analogy, though it may well be that there are a number of plausible missing premises that could be uncovered that would turn it into an entailment. Still, I am not tempted to mount such a defense, as I think the entailment requirement is too strong. Rather, I offer the analogy as *evidence* that the entailment relation is simply not necessary: we *do* see how understanding the loose intermolecular bonds render the gap between the disparate items *understandable to us without an entailment relation.* If an entailment relation is not necessary to close this particular gap, why should we require it *a priori* in the brain/mental states gap? Where is the argument for the claim that the relation that closes the gap between diverse entities must be entailment?

Let us suppose, as seems true, that the molecule/liquid gap is closed without an entailment relationship between the structure and dynamics of molecules and liquidity. What seems to be operative is some "plausibility relation" between the disparate items. Given these disparate items and the weak intermolecular bonds, it is *comprehensible to us* that such structure and dynamics could give rise to the observed liquidity.

The problem of the explanatory gap is just such a "comprehensible to us" problem. How can we understand how brain states, apparently so different from mental states, nevertheless, give rise to them? Requiring that the relevant scientific theories entail that there are mental states is, to my mind, to set too high a standard. I, for one, would think that the explanatory gap between these disparate items closed, if one could give a comparable account of their relations to that given in the case of molecules and liquidity. *If we could only just see how something new and different could arise out of the brain states*. Of course, none of this is to eschew entailment relations. All the better if they are produced. My point is simply that they are not necessary for closing the explanatory gap.[17]

There remains Chalmers' main objection. According to him, what is essential to a physical theory is structure and dynamics, and a new physical

theory will just give us more of the same; so, a new physical theory will not help in closing the explanatory gap. But Chalmers casts his argument at far too general a level. The *particular character* of the structure and dynamics is what makes a difference. Just as I earlier argued that talk of supervenience is empty without specification of the relevant determination relation, so too is general talk of structure and dynamics. In the molecule/liquid analogy, it is the *particular character* (weak intermolecular bonds) of the *structure and dynamics of molecules* that enables us to understand how liquidity can arise out of such qualitatively different items. So, too, we may expect that the *particular* character of the new dynamics is what will enable us to close the explanatory gap between brain and mental states. Additionally, the analogy leads us to the hopeful expectation that detailed or highly sophisticated knowledge of the new dynamics may not be necessary to close the gap.

3. Chaos

It is the fact that mental states are so qualitatively different from brain states that gives rise to the conundrum of the explanatory gap. In short, we seem to get something totally new and different out of the activity of neurons and neurotransmitters. A significant contributing factor to this mystifying situation is the implicit assumption that the relevant mathematics is linear. For within such models, it is true that 'the whole is the sum of its parts.' You do not get more, and you certainly do not get anything different from the linear combination of the parts.[18] Mental states being so evidently different from brain states, showing how they could arise out of the latter seems to be an impossible, even logically impossible, undertaking, so long as inquiries are restricted to linear models. Until recently, nonlinear models were eschewed in scientific investigations because of their complexity, while linear models applied to the activity of neurons and neurotransmitters, for the reasons given, appear *logically inadequate* to close the gap. The situation, thus, seems desperate and is a compelling reason to reject option 1, listed earlier. In consequence, if one's sights are limited to linear models, option 3 appears inescapable.

Prospects are brighter when *self-organized* systems governed by nonlinear dynamics are considered.[19] These are open systems, some of whose properties are not completely explainable by properties of parts of the system. Consider a simple example. When oil is heated uniformly from below, the heat is randomly dissipated, and there is no qualitative change in the properties of the liquid. However, a qualitative change occurs once a certain temperature is reached: the liquid rapidly assumes a lattice structure of hexagonal convection cells (Bénard cells). This new behavior is not merely driven by input and initial conditions; the

molecules collectively organize themselves. This collective behavior of large populations of elements of the system constrain the behavior of individual elements so that there is a kind of "top-down causality," contrary to the usual patterns of explanations. Such self-organized behavior governs the dappling patterns of animals' coats, hurricanes, cellular slime molds, and even in the formation of the visual cortex (Kelso 1997: 6–15).

From such cases, we may extract a unique and peculiar attribute of nonlinear interactions of elements that provides for at least the possibility of the closing the explanatory gap. It is the **X** of Section 2:

(i) The dynamics of self-organized systems described by nonlinear equations can yield properties of the system that are qualitatively different from any linear combination of its variables.

Nonlinear dynamics is the only known candidate possessing this feature, and this is exactly the kind of feature that can resolve the basic conundrum over the explanatory gap. For the closing of the gap seems to require getting something more and different out of interacting elements.[20] Freed from the constraint of linear models and given (i), what appeared to be a recalcitrant mystery is transformed into a manageable, if still formidable, problem. The enormous task remains to characterize the nonlinear equations of brain state variables so that something more and different, mental states, can arise out of their interactions. Viewing the brain as a self-organized system governed by nonlinear dynamics is the general core of a new paradigm of brain behavior that at least provides the framework for a solution, a solution that appears impossible against a background of linear models.

Noticing that (i) is exactly what is *needed* to do the job falls far short of actually producing the relevant nonlinear mathematics. Still, it is gratifying to realize the abstract point that there is this unusual and powerful feature of nonlinear dynamics that provides a *basis for our comprehension* of how brain states can give rise to mental states (even though they are so radically different from one another). This is rather like the situation with our earlier analogy. Knowledge of the general character of weak intermolecular bonds is sufficient for us to comprehend how liquidity can arise from a multitude of discrete entities; no sophisticated knowledge of the theory of molecular bonds is required. It is no small point that, in a similar way, a most serious logical obstacle to closing the explanatory gap between the disparate items of mental and brain states is swept away. Furthermore, we will see that there is some cause for optimism with regard to the existence of nonlinear models of brain activity.

In the discussion of the heated oil, we observed that the emergent property of the system, hexagonal convection cells, exerted a kind of top-down causality, properties of large collections of molecules constrain the behavior of individual

molecules. This is a second powerful and unique feature of self-organized systems that provides the basis for removal of another major logical conundrum between brain and mind, mental causation. Here, certain high-level mental states, such as beliefs and desires, appear to activate certain microstates, the activation of motor neurons, which results in the agent moving to fulfill her desire. We may express the feature at issue as follows:

(ii) The dynamics of self-organized systems described by nonlinear equations yield global properties of the system that causally constrain microelements of the system.

Assume, for the moment, that conscious thoughts and perceptions are just these sorts of emergent properties of self-organized brains governed by nonlinear dynamics. (Two prominent neuroscientists, Earl MacCormac (1996: 151) and Walter J. Freeman (1991 and 1995, e.g., have so hypothesized, respectively). The idea that conscious thoughts or perceptions actually bring about certain behaviors by means of certain sensory motor responses, is *demystified* on this assumption, for we have simple analogical illustrations (Bénard cells discussed above, and the other examples cited) of how such emergent and qualitatively different states can constrain microstates.

So, a second apparently logically impossible realization of a relation between brain and mind becomes *comprehensible to us, given the new paradigm*. Again, this only holds out the abstract possibility, one whose promise could only be fulfilled after completion of enormous scientific work, work demonstrating the relevant nonlinear equations that support such speculations. But, once again, it is a most gratifying abstract possibility, as it transforms a seemingly logically impossible situation into one that is within our grasp.

Thus, I suggest that the conundrums of the explanatory gap and mental causation may well be simply artifacts of methodologies based in linear micro-reduction models.

Walter J. Freeman has successfully applied chaos theory to neurophysiological systems.[21] He did extensive work on the olfactory system of rabbits, which I will focus on, but he has found similar results in rats and cats, and in the visual system of monkeys (Freeman and Dijk 1987), cats (Eckhorn *et al.* 1988; Gray *et al.* 1989), and humans (Schippers 1990). Somewhat more tentatively, similar results were found in the somatosensory system of humans (Freeman and Mauer 1989). I must add that there is no way I can do justice to the extensive important work that Freeman has done in this area. I will only touch on a few highlights of his varied experimental and mathematical support for his model and strongly encourage the reader to "follow his nose" in pursuing Freeman's extensive research.[22]

There is a significant split among brain researchers as to the importance of

relatively local behavior of neurons (feature neurons or "grandmother cells"), on the one hand, or the global activity of large populations of neurons, on the other, for understanding perception and cognition. Freeman apparently started in the former camp, but twelve years of experimentation using traditional methods on the olfactory system of the rabbit transformed him into a leading exponent of the importance of global activity and the application of chaos theory. His attempts to find significant relations between odorants and relatively local neural populations failed.[23]

The initial detection of an odor is by a spatial sheet of receptors in the nose. These receptors signal through the olfactory nerve another sheet of cells in the olfactory bulb. Although single neurons do fire in response to specific odors, the same cell fires in response to different odors and the same odor can excite many different cells. Freeman used electroencephalograms (EEGs) to record the activity. What he found was that the patterns of two successive sniffs of the same odor were as different from one another as were the patterns between two sniffs of different odors. The spatial patterns of the EEGs did not reliably correlate with the input stimuli. They changed not only when stimuli changed but when anything else changed. Change even occurred on reinforcement of an odor not previously reinforced. This change in reinforcement pattern for a particular odor had a ripple effect in that the spatial patterns for all the other odors also changed, indicating associative memory.

In brief, the receptors in the nose activate certain neurons in the olfactory bulb. The bulb then constructs a global spatial pattern that is dependent on past experience in that which neurons fire synchronously is a function of the specific past activations of neurons. (We will see below that the entorhinal cortex, thought to be involved with both memory and emotion, also plays a significant role in the construction of this pattern.) It is the global pattern of cooperative activity of neurons that is transmitted to the olfactory cortex and other areas. Freeman allows that the sheets of receptors in the nose and at the initial input to the olfactory bulb are feature detectors that refine sensory input, but they do not constitute perception. Importantly, the raw sensory data recorded in these sheets is "washed out," not transmitted as such (though it can be "recovered" if the need arises). Stimuli destabilize neuronal populations, causing them to construct patterns that express the *significance* of the stimuli to the animal based on its experience; it does *not* express the stimulus *per se*. (Compare Akins' (1996) discussion of heat, where no single correlation is found between the heat stimulus and the various resultant changes in physiological states.)

The EEGs for a particular odorant indicate an attractor. Although different sniffs of the same odorant produce differences in EEG as great as do different odors, the spatial patterns of EEG for a given odorant are similar, as remarked above. Each inhalation causes a burst of bulbar activity. Although each EEG tracing varies, a common waveform or carrier wave is embedded in the tracings,

and the identity of the odorant is reliably identified in the bulbwide spatial pattern of the carrier wave amplitude. The aperiodic common carrier wave is everywhere in the bulb. It occurs not only during bursts but also between bursts, when there is no extractable stimulus. As Freeman says: "The carrier wave is not imposed from outside the olfactory system by the receptors or by other parts of the brain. It emerges as a cooperative activity that is self-organized by the neural masses." (1992: 468) This lack of external driving of the activity is what indicates that it is self-organized which, in turn, is characteristic of chaotic systems.

The varying set of different receptor cells in the nose that cause the global activity in the olfactory bulb to exhibit this spatial pattern is its basin of attraction. Thus, one attractor can be entered from a variety of starting points. I emphasize that the particular array of receptor cells excited varies considerably from one sniff to another, even when of the same odorant, but each odorant has its own attractor and basin. Input to the basal chaotic attractor causes a burst to the appropriate chaotic attractor for the input. To Freeman, chaos is essential, as it is the only way to account for the observed rapid shifts from one brain state to another.[24]

According to Freeman, the general dynamics of perception is as follows. The brain seeks information by directing an individual to sense. Self-organizing activity in the limbic system (a part of the brain that includes the entorhinal cortex and is thought to be involved in emotion and memory) directs the search by transmitting a search command to the motor systems and simultaneously sends what is called a reafference message to the sensory systems. The reafference message directs the sensory systems to prepare to respond to new information. The sensory systems strongly respond with a burst, every neuron in a given region participating in a collective activity. Resulting synchronous activity from these systems is then transmitted back to the limbic system, where they are combined and form what Freeman calls a 'gestalt'. In a fraction of a second, the process repeats. Freeman concludes: "… an act of perception is not the copying of an incoming stimulus. It is a step in a trajectory by which brains grows, reorganize themselves and reach into their environment to change it to their own advantage" (Freeman 1991: 85).

Generalizing on some earlier work of others (Helmholtz, von Holst, Mittlestadt and Sperry), Freeman sees all goal-directed movement as initiated by the limbic system, with the entorhinal cortex playing a central role. All sensory systems converge in the entorhinal cortex, and the entorhinal cortex is involved in memory and emotion. It was already noted that the pattern of the burst activity associated with a sniff was a function of experience and did not correlate with the external stimulus itself. What all this suggests is that the experience that shapes the burst activity is not simply a recording of what has happened, but includes its emotional value or significance to the animal.

One would expect that the significance of the smell of a fox would be different for a rabbit and a dog; for the one, memories of chase and fear, for the other, food and expectation of a meal. According to Freeman, the bursts correlated with such smells are unique to the individual. It is a commonplace that one's emotions can affect one's perceptions and thoughts. The model of perception provided by Freeman would seem to provide some physiological basis for such a commonplace, as the reafferent messages transmitted to the sensory cortex from the entorhinal cortex apparently reflect the significance of the stimulus to the particular animal based on its experience.

In concluding, let me remind the reader that in Section 2. I argued that a new paradigm of brain science is needed to close the explanatory gap. I argued that what was important in the new science is its *particular* structure and dynamics. I illustrated this in my recommendation of an analogy that provides a model of what we need to close the explanatory gap. We saw that the *particular* structure and dynamics must be such that they can make *comprehensible to us* how the interaction of certain elements and their properties can give rise to qualitatively different properties.

The new science that appears to provide a basis for closing the explanatory gap between qualitatively different mental and brain states is that of chaos theory applied to brain activity. The new *particular* feature that this theory provides, and which is unique to it, is presented above as (i). We found yet another special feature of this theory, (ii), which may well provide the basis to make comprehensible to us how there can be mental causation. Based on how the molecule/liquid analogy turned out, we expected and hoped that detailed knowledge of the new theory would not be required to make comprehensible to us how it can resolve the conundrums we face. Apparently, our hopes and expectations in this regard have been fulfilled. We may take further gratification in the fact that not only does chaos theory do all this, which is at a rather abstract level, but also it already has been successfully applied to concrete questions of brain activity.

Some have criticized the application of chaos theory to brain activity because unlike, say, computational models, there is no clear or significant place for representations in chaotic dynamics, and it is a widely held assumption that representations (both conscious and unconscious) play a pervasive role in cognition. The point of my first section is to diffuse this concern by arguing that there simply are no unconscious representations, though many take them for granted. Representations are central to *conscious* thoughts — but only there.[25] Where they are operative, they may well be manifested as bursts to attractor states. Such bursts constrain the firing of individual neurons in a top-down fashion; this is a manifestation of feature (ii). Furthermore, we know that these bursts are self-organized and are not driven by external stimulation. By feature

(i), we may expect that the resulting burst states are significantly and qualitatively different from their antecedent states. This is exactly what we should expect if only conscious states are representational.

Appendix

A general and brief explanation of some of central concepts of chaos theory are here provided for the convenience of the reader. The relevant form of chaos is deterministic as opposed to entropic. Entropic chaos simply moves to increasing disorder without emerging patterns of activity in the system. Where there is deterministic chaos, similar patterns appear and disappear periodically in the dynamical activity of a system, but they do so with a complexity that makes it difficult to detect and, therefore, appear to be entropic. The complexity is due to the iterative nature of the equations governing the behavior. In iterative equations, the outcome of each computation is the input for the next computation. A characteristic feature of chaotic systems is that very slight differences in the starting point or weak input can lead to huge differences in eventual outcome. This extreme sensitivity to initial conditions is referred to as the 'butterfly effect.' This results in the extraordinary feature that the outcomes are not predictable, even though the equations are often quite simple and deterministic. Superficially, at least, this feature would seem to go far in explaining why, although we can frequently give an ad hoc explanation of someone's behavior, we are not as successful in predicting what someone will do, despite similar initial conditions.

A key concept required for our purposes is that of an *attractor*. When a dynamical system settles into a certain pattern, that pattern is known as an attractor; there are several kinds of attractors. The simplest attractor is the *single-point attractor*. The point at which a pendulum subject to friction eventually stops typically illustrates it. When a pendulum is not subject to friction, it will continue in exactly the same pattern of motion forever. Such repeating patterns are known as *limit-cycle attractors*. There are a number of other kinds of attractors, which like the one just defined, continue to repeat the same pattern of motion. Earlier I stated that similar patterns are repeated in chaotic systems. Characteristic of these systems, unlike the ones just discussed, is that they *never* traverse the same path, but the paths are similar and do trace recognizable patterns. Because of the uniqueness of each path, such patterns are known as *strange attractors*.

The set of initial values that terminate in an attractor is known as its *basin of attraction*. In chaotic systems, there are rapid *bursts* between strange attractors as well as back to the basal state. These bursts are called *bifurcations*; they

constitute significant phase transitions in the system. The transition from rest to walking, or from walking to a trot, or in fluid dynamics, the shift from laminar flow to turbulence, are all common illustrations of bifurcations.

Notes

1. E.g., for each particular number, I unconsciously believe that it is a natural number.
2. Let me make it clear that I think Searle is, for the most part, profoundly right about intentionality, and though I think he is wrong in holding that unconscious beliefs are intentional, abandoning this does not necessitate wholesale changes in his view.
3. In arguing for the connection principle (1992: 156–62), Searle advances his claim that unconscious beliefs have aspectual shape in some detail. He has confirmed (in conversation) that something like the short argument here presented captures the thrust of the point contained in the larger argument of the broader conclusion.
4. The above would seem to undermine Searle's *connection principle:* "all unconscious intentional states are in principle accessible to consciousness ..." (1992: 156); there are no unconscious intentional states on my view. Still, a modified form of the connection principle survives; one that I believe preserves the spirit of Searle's principle. Space prohibits elaboration.
5. He thinks we would also need some law-like connections between the levels, as he says, "... we would still have to have some law-like connection that would enable us to infer from our observations of the neural architecture and neuron firings that they were realizations [for example] of the desire for water not of the desire for H2O," (1992: 158)
6. See my (1990) where I argue that certain third-person accounts fail in providing for this first-person access, and (1999) where I argue that Burge's original and influential thought experiment in favor of anti-individualism fails. My positive account is in my (1994) and (1996).
7. I discuss Freeman's work in some detail in section 3.
8. More fully, she says:"The patterns of neural activity that Freeman records in the EEG is the result of a self-organized, acquired tendency to behave in a certain way given the goals of the system considered as a whole ... The deciding factor in brain function is not neural activity patterns per se, ... not the input pattern and its transformation within the animal, but the internally generated neural dynamics created by the system itself." (1987: 198)
9. I argue for this at length in "Represent and Information Bearers", unpublished.
10. It is well known that David Chalmers (1996) is credited with coining the former expression and Joe Levine (1983) the latter.
11. It is important to stress this ordinary observational sense of the term, as David Chalmers objected to this analogy, used in a distant ancestor of this paper, based on a different interpretation. He says, "... the water case is not a useful analogy for you, because in that case it's clear that what needs to be explained is structure and function — we explain the liquidity, transparency, etc. and it's obvious that those are the sort of structural/functional matters that are wide-open to physical explanation." (12/13/98) For it to be "obvious" that these properties are open to a structural/functional explanation, one must presume, at the least, that these properties are understood in a physio-chemical way, rather than in the way I recommend here, a way that

corresponds to our ordinary observation of liquids. I will have more to say about the role of structure and function in these matters latter in this section.

12. Similar remarks apply to the solidity of ice and water vapor, where the strength of the intermolecular bonds are much stronger or weaker, respectively.

13. I have discussed this in my (1995).

14. I raised these points in criticism of John Heil's treatment of supervenience in my (1995) review of his (1992). Apparently, he now agrees, for in the conclusion of his (1998: 153–154) he states, "I have argued that the concept of supervenience as standardly formulated provides little in the way of ontological illumination … [S]upervenience is not explanatory … Supervenience claims hold, when they do, because the world is a particular way. What we need to be clear about is what that way is. Different cases yield different results."

15. As to the competition between options 1 and 2, one is reminded of Kuhn's remarks pertaining to the differing reactions of scientists to a paradigm under stress.

16. Colin McGinn (1989) has argued for such a pessimistic conclusion. Space limitations do not permit discussion here of his subtle arguments.

17. One reason someone might think they are necessary is if the explanatory gap problem is confused with a reduction of phenomenal states to brain states and one accepts the classical model of reduction of one kind of entity to another. I think it is a mistake to view the explanatory gap problem as a reduction problem, regardless of the model of reduction.

18. True, the whole may function differently than the parts individually, but the whole is not qualitatively different from its parts. Most functionalists (not without exception) admit their inability to reduce qualia, phenomenal states, to functional states. In my 1994 and 1996, I argue that they are also unsuccessful in reducing propositional contents to functional states.

19. Self-organization is a characteristic of chaotic systems. See Appendix.

20. J. A. Scott Kelso (1997) lists among the conditions for a self-organized system that there must be a large number of elements with nonlinear interactions. He points out that this requirement constitutes a major break with Sir Isaac Newton, whom he quotes as holding that "[t]he motion of the whole is the sum of the motion of all the parts." (Definition II, Principia) In contrast, Kelso holds "[f]or us, the motion of the whole is not only greater than, but different than the sum of the motions of the parts, due to nonlinear interactions among the parts or between the parts and the environment." (Kelso 1997: 16)

21. Detailed discussion of Freeman's work supporting the role of chaos in brain processing of information is in the papers by Chris King, Carl M. Anderson and Arnold J. Mandell, Earl R. MacCormac, and David M. Alexander and Gordon G. Globus in Stamenov and MacCormac (1996). Others also successfully apply chaos theory to the brain. For a fascinating and clear account of other successful applications of chaos theory to the brain, see Kelso (1997).

22. See, for example, his 1996, 1991, 1992 1995, a sequence that goes, roughly, from simpler to more technical works.

23. It is worthwhile to note that Freeman rejects the "binding problem" as a psuedo-problem, for it is based on the idea of feature detector neurons in each of the sensory cortexes, which he rejects because it is observer relative and what is significant for the brain is the global activity of neurons. (The binding problem is how the outputs from the different feature detectors are connected.) A proposed solution, one which Freeman himself had a hand in, is that feature neurons fire synchronously (40 Hz.). Aside from the difficulties with feature detector neurons,

this particular solution is now rejected by him since the frequency distribution in the visual and olfactory systems is not synchronized but broad spectrum, as Freeman has demonstrated.

24. A further fact that supports the contention that the system is chaotic is that the bulb and the olfactory cortex excite each other so that neither settles down, nor do they agree on a common oscillation frequency. (If the connection is severed between the two, they become stable and quiet.) This competition increases sensitivity and instability and contributes to chaos.

25. I regret that space limitations do not allow for further discussion of the implications of my thesis that there are no unconscious representations for Searle's Connection Principle.

References

Akins, Kathleen. 1996. "Of Sensory Systems and the 'Aboutness' of Mental States". *The Journal of Philosophy XCIII*: 337–372.

Alexander, David M., and Gordon G. Globus. 1996. "Edge-of-Chaos Dynamics in Recursively Organized Neural Systems". In MacCormac, Earl, and Maxim I. Stamenov, (eds), *Fractals of Brain, Fractals of Mind.* Amsterdam and Philadelphia: John Benjamins, 31–74.

Anderson, Carl M., and Arnold J. Mandell. 1996. "Fractal Time and the Foundations of Consciousness: Vertical Convergence of 1/f Phenomena From Ion Channels to Behavioral States". In MacCormac, Earl, and Maxim I. Stamenov, (eds.), *Fractals of Brain, Fractals of Mind*. Amsterdam and Philadelphia: John Benjamins, 75–126.

Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.

Churchland, Patricia Smith. 1996. "The Hornswoggle Problem". http://www.merlin.com.au/brain_proj/psch_2.htm.

Eckhorn, R., R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk, and H. J. Reitboeck. 1988. "Coherent Oscillations: A Mechanism of Feature Linking in Visual Cortex?" *Biological Cybernetics 60*: 121–130.

Freeman, Walter. 1996. "Interview with Jean Burns: Societies of Brains". Journal of Consciousness Studies 3 (2):, 172–180.

Freeman, Walter J. 1995. *Societies of Brains: A Study in the Neuroscience of Love and Hate*. New Jersey: Lawrence Erlbaum Associates.

Freeman, Walter J. 1992. "Tutorial on Neurobiology: From Single Neurons to Brain Chaos". *International Journal of Bifurcation and Chaos* 2(3): 451–482.

Freeman, Walter J. 1991. "The Physiology of Perception". *Scientific American 264 (2)*: 78–85.

Freeman, W. J., and K. Maurer. 1989. "Advances in Brain Theory Give New Directions to the Use of the Technologies of Brain Mapping in Behavioral Studies". In Maurer, K (ed.), *Proceedings, Conference on Topographic Brain Mapping*. Berlin: Springer-Verlag, 118–126.

Freeman, W. J., and B.Van Dijk. 1987. "Spatial Patterns of Visual Cortical Fast EEG During Conditioned Reflex in a Rhesus Monkey. *Brain Research 422*: 267–276.

Georgalis, Nicholas. 1999. "Rethinking Burge's Thought Experiment". *Synthese* 18: 145–64.

Georgalis, Nicholas. 1996. "Awareness, Understanding, and Functionalism". *Erkenntnis 44*: 225–256.

Georgalis, Nicholas. 1995. "Review of The Nature of True Minds". *Philosophical Psychology 8 (2)*: 189–193.

Georgalis, Nicholas. 1990. "No Access for the Externalist: Discussion of Heil's 'Privileged Access'". *Mind 99*: 101–108.

Georgalis, Nicholas. 1994. "Asymmetry of Access to Intentional States". *Erkenntnis 40*: 185- 211.

Gray, C. M., P. Koenig, K. A. Engel, and W. Singer. 1989. "Oscillatory Responses in Cat Visual Cortex Exhibit Intercolumnar Synchronization Which Reflects Global Stimulus Properties". *Nature 338*: 334–337.

Heil, John. 1998. "Supervenience Deconstructed". *European Journal of Philosophy* 6 (2): 146- 155.

Heil, John. 1992. *The Nature of True Minds*. Cambridge: Cambridge University Press.

Kelso, J. A. Scott. 1997. *Dynamical Patterns*. Cambridge: MIT Press.

King, Chris. 1996. "Fractal Neurodynamics and Quantum Chaos: Resolving the Mind-Brain Paradox Through Novel Biophysics". In MacCormac, Earl, and Maxim I. Stamenov (eds), *Fractals of Brain, Fractals of Mind,* 179–234. Amsterdam and Philadelphia: John Benjamins.

Levine, Joseph. 1983. "Materialism and Qualia: The Explanatory Gap". *Pacific Philosophical Quarterly 64*: 354–61.

MacCormac, Earl R. 1996. "Fractal Thinking: Self-Organizing Brain Processing". In MacCormac, Earl, and Maxim I. Stamenov (Amsterdam and Philadelphia: John Benjamins, 127–154.

McGinn, Colin. 1989. "Can We Solve the Mind-Body Problem?" Mind 98:, 349–66.

Schippers, B. 1990. "Spatial Patterns of High-Frequency Visual Cortical Activity During Conditioned Reflex in Man". Masters Thesis, Laboratory for Medical Physics, University of Amsterdam, Netherlands.

Searle, John. 1992. *Rediscovery of the Mind*. Cambridge: MIT Press.

Skarda, Christine. 1987. "Explaining Behavior: Bringing the Brain Back". *Inquiry 29*: 187–202.

Part III

Emotional Learning and Development

CHAPTER 10

Child Development and the Regulation of Affect and Cognition in Consciousness

A View from Object Relations Theory

Peter Zachar
Auburn University, Montgomery

1. Introduction

The positive role played by the emotions in 'rational' adaptation has only recently become a topic of discussion among scientifically-minded psychologists and philosophers. The scientific community's reluctance to study emotions has ironically depended, in part, on a bias derived from the Cartesian idea of emotions as subjective passions, as things that take us over against our will and reason (Averill 1980). Even though ideas about the base nature of emotions can be traced back to Plato, Charles Darwin's placement of emotional functions in the prehuman epoch of our species' evolutionary history has probably exerted the greatest influence on physiologically-oriented moderns. Darwin even thought of emotional reactions as evolutionary artifacts having limited adaptive value for *Homo sapiens*. J. Hughlings-Jackson's model of the evolutionary levels of organization in the brain, with the cortex and reasoning at the 'top' followed by the emotions and reflexes at 'lower' levels reflects this Darwinian view. In this model, emotional reactions are primitive, and rationality requires their minimization.

This model has been criticized by Damasio (1994) who claims that the cerebral structures to which scientists have traditionally attributed 'rationality' are not built on top of limbic system structures, but are built out of and integrated with them. From the standpoint of contemporary neuroscience, 'rationality' requires the functional integration of emotion and cognition. As an example, Damasio points out that the important deficits in cases of frontal lobe brain injuries such as those of Phineas Gage involve an inability to experience the emotional reactions normally felt in response to imagined consequences of intended courses of action. Without emotional understanding, especially of

negative consequences, our ability to engage in practical decision making is decimated. The passions are not as passive as Descartes thought.

Freud's relation to this folk model of emotion is a complicated one. From one standpoint he agreed with Hughlings Jackson's view that cerebral mechanisms are built on top of emotional brain centers, and clearly thought that reality testing and adaptation were taken care of by cerebral mechanisms — or ego mechanisms. He also saw instincts as animalistic. On the other hand, he was sympathetic with certain 19th century Romantic ideas about the pervasive role played by emotions in the psyche, and saw ego mechanisms and emotions as intimately related. This dual view was a common one in Germany. The founder of scientific psychology, Wilhelm Wundt, also believed that emotional states constitute the primordial form of consciousness. In fact, a focus on the role played by emotions in the construction of consciousness is common to many thinkers who adopt an active rather than a passive model of the mind. For them, the entire mind including the emotions is actively involved in understanding.

What made the psychoanalytic perspective that Freud introduced unique was its focus on motivation. Freud and his followers have always attempted to explain behavior in terms of its roots in emotion and motivation. Not all analysts adopted Freud's reductionism regarding the physiological or sexual nature of motivation, but all have agreed that basic motivational forces and their transformations permeate consciousness. Their theories include the various *object relations* theories of child development (Klein 1964; Fairbairn 1952; Winnicott 1965; Jacobson 1964, Mahler 1979). These theorists take the need for interpersonal relationships rather than for tension reduction to be the primary motivational force. Because emotions connect us with and separate us from others, they take center stage in the object relations paradigm.

In this chapter, I review the child development literature from an object relations perspective, showing how the development of a more integrated consciousness and sense of self requires the development of a relatively cohesive emotional life. My description is normative, and does not address individual differences. I will also discuss implications that this model of psychological development has for the relationship between emotion and cognition.

2. The object relations model of development: An overview

Although developmental psychologists are more familiar with John Bowlby's (1969) work on *attachment*, most clinicians in the psychodynamic tradition prefer the research of Margaret Mahler and her colleagues on the development of *psychological autonomy* (Mahler, Pine and Bergmann 1975). In contrast to Bowlby's colleague Mary Ainsworth (1979), who described behaviors related to

attachment, Mahler attempted to understand how a child becomes a psychologically autonomous and distinct individual. Like all psycho-dynamically-informed theorists, she paid careful attention to developments in the child's emotional life. According to Mahler, the development of autonomy depends on the achievement of emotional integration.

Symbiosis. Mahler's research-based descriptions of development begin at the two month point, when the infant has not yet grasped what clinicians refer to as the internal-external distinction, i.e., infants are unable to distinguish sensation coming from inside the body from sensation coming from outside. Mahler refers to the time in an infant's life before the establishment of a stable internal-external distinction as the *symbiotic* stage. By symbiosis, Mahler means that the mother and infant are so coordinated that the infant experiences their interaction as a 'dual-unity' rather than as an interaction between two distinct individuals.

With respect to development during the symbiotic stage, Mahler believes that intuiting the internal-external distinction is the first step in learning to consciously distinguish self from other. This view is shared by those developmental psychologists who believe that an infant's ability to distinguish self from other begins with the ability to distinguish self-propelled movement (agency) from the external movement of objects (Golinkoff 1983; Johnson 1988; Poulin-Dubois and Shultz 1988; Premack 1990). Piaget (1954) was also concerned with the internal-external distinction, except that his research focused on the child's understanding of the external world of objects. In contrast, Mahler was interested in understanding the development of a coherent internal world as reflected in the child's interactions with primary caretakers.

A critic of Mahler, Daniel Stern (1985), argues that symbiosis should not be seen as a stage of *passive* self-other undifferentation, but as an infant's subjective reaction to *active* self-other coordination. For example, successful cuddling requires mother *and* infant to mold their bodies to each other, although Mahler herself believed that much of the adaptation was done by the infant. In response to Stern, Mahler and her colleagues claim to describe what, in theory, the mother-child interaction would seem like from the perspective of the vulnerable infant who is intensely dependent on someone outside itself for adequate *regulation* — i.e., it would have to seem like symbiosis.

In psychodynamic psychology, the most important kind of regulation is emotional. Many developmental psychologists believe that at birth, infants exist in a state of relative emotional diffuseness. At birth emotion is arousal, but infants soon gain the ability to express *positive* and *negative* arousal. Discrete positive and negative emotions are cognitive elaborations of positive and negative arousal. As the infant gains the cognitive abilities to understand distinctions between self, other, internal, and external, discrete emotions emerge. For example, using a cognitive model, Lewis (1993) has described the sequence in

which specific emotions emerge during development. The emotions that appear during the period of symbiotic relatedness include interest, joy, sadness, disgust, and anger. In Lewis's model, experiencing anger assumes an implicit sense of self and an awareness that desires and expectations are being thwarted by an obstacle.

The issue of undifferentiated versus discrete emotions has been debated by developmental psychologists since the 1920s (e.g., Bridges 1932; Lewis 1993). Discrete emotions researchers such as Izard (1991) and Ekman (1995) have identified specific facial expressions made by infants, and have shown that these expressions conform to the facial expressions made by adults when in specific kinds of emotional states. Izard and Ekman both infer that if the infant has the same expression as an angry or happy adult, they must also be experiencing the same subjective state as the adult.

Although we are prone to interpret discrete emotional expressions on the faces of infants, research by Camras (1992) indicates that these expressions are not reliably elicited by situations, so the infants may not be experiencing discrete emotions. For example, infants may express fear when not wanting to be fed or sadness in reaction to a food they do not like. Infants also display a composite negative reaction pattern best described as distress-pain-anger-sadness before they display discrete emotions, even though these composite reactions are ignored by discrete emotions researchers who only count hits with respect to their discrete-emotion coding schemes.

As with most developmental achievements, it is a judgement call as to when behavioral and affective precursors to anger become anger proper. If emotions are primordial forms of consciousness, infants probably experience something like joy, sadness, and anger, but distinctions between these states are looser than they are for children and adults. In other words, infant emotional states have fuzzy boundaries.

Emotion involves more than just discrete affect. Stern (1985) claims that *vitality affects* appear before discreet affects. By vitality affects he means the pacing of experience in terms of surging, fading, rushing, etc. Emotional states rise and fall over time, and the experience of this change as it occurs is important. As a matter of fact, coordination of these vitality affects is an important component of symbiotic relatedness. The mother's ability to match the infant's 'wows' and 'huhs?' is the kind of affective attunement that leads to the experience of symbiotic relatedness. It is an important part of her ability to regulate infant emotional states.

Separation-Individuation. Mahler's primary concern is *separation-individuation*, a process which gains speed at about six months and is completed by the third year. It covers the time period in which an infant becomes a child, i.e., separates from the early dependency and symbiotic attunement with the

primary caretaker to become a more psychologically distinct individual.

Differentiation. According to Mahler, Pine and Bergman (1975) visible proof of the internal-external distinction is seen when the six-month-old baby, in its mother's arms, pulls away to get a better look at her. At this time, infants adopt what Mahler and colleagues call a permanently alert sensorium: a level of ongoing alertness usually associated with mature consciousness. They are more fully 'awake.' Infants also begin to prefer the primary caretaker to other persons in their lives. Mahler calls this the *differentiation* subphase. The child does not just differentiate itself from the mother; it differentiates the mother from others in terms of emotional preference.

Highlighting this aspect of development, Bowlby (1969) called it *attachment*. Obviously, differentiation and attachment are interrelated pro-cesses. Attaching to the mother requires being able to differentiate her from self and others. In contemporary psychodynamic theory, attachment, autonomy, trust, and dependency are considered 'developmental lines' rather than 'bounded stages.' A developmental line is what Stern (1985) calls an ongoing life issue. Differentiation, for example, begins at birth, but its effect on the child's developing sense of self and other becomes more pronounced at certain points. One can therefore talk about a differentiation phase, but differentiation proper is not bounded by a specific time period.

Physiologically, differentiation depends on an infant's vision becoming more adult-like. Until they are about six months old, infants cannot scan a room full of objects as an adult can. They are more attuned to edges than to objects. As Kalat (1998) notes, they cannot voluntarily shift attention from one object to another, and visual acuity is limited to peripheral vision: they see better from the side than they do looking straight on. At six months this changes, and they can focus on whole objects.

By six or seven months children also have enough of an internal-external distinction to realize that they are dependent on an object outside themselves. This leads to the beginning of separation anxiety, and the need for soothing. Their growing understanding of basic separateness enables them to have more complex reactions, such as abandonment anxiety. In fact, separations from the mother are not traumatic until about the sixth or seventh month, which is where fear first appears in Michael Lewis's (1993) model.

Both symbiosis and differentiation overlap with the *contagion* phase of infant emotional development. Contagion refers to 'contagious emotion,' such as smiling in response to another person's laughter. Magai and McFadden (1995) note that infants in the five to seven month range are extraordinarily vulnerable to being influenced by another's positive and negative emotional states. During the contagion phase, the mother's ability to regulate her own mood contributes to the mood regulation of the infant. By the eighth month, contagion reactions

are reduced in frequency, supporting Mahler's idea that differentiation is in progress.

At this time, infants shift from passively adopting another person's reaction to actively seeking information about others' reactions in order to know how they themselves should react. For example, at one year, children look at their mother's face to determine which emotional reactions they should have in ambiguous situations, a phenomenon known as *social referencing*. (Campos and Sternberg 1980; Sorce, Emde, Campos, and Klinnert 1985). The distinction between contagion and social referencing is that in referencing, children are using parental reactions to an event outside the parent-child dyad to figure out how they should react to that event.

The differentiation phase also occurs at the same age that the child psychiatrist Winnicott (1958, 1965) thought marked the appearance of *transitional* objects: external objects to which the child develops a strong emotional attachment. Common examples of transitional objects are blankets or Teddy Bears. It is not the object itself that is transitional; rather the child is making a transition from absolute dependence on the mother for emotional regulation to relative dependance in which she or he develops the ability for self-regulation (i.e., autonomy). The transitional object must be an external object that is the sole possession of the child to cuddle or crush as she or he will. Its presence and absence must also be under the child's control. It is an external object that provides emotional regulation, but it is *the child's own* — not Mom's or Dad's or brother's or sister's.

Practicing. Psychological development is constantly propelled forward by physical maturation. The centrality of differentiation is replaced by the *practicing* subphase, which is contingent on the young child's physical ability to crawl and then walk. Practicing begins as early as nine months. Following Greenacre (1957), psychoanalytic thinkers label this time in a child's life 'the love affair with the world,' which becomes a source of wonder and adventure. This is potentially a time of emotional expansiveness. Parental enthusiasm about the now-walking toddler re-enforces his or her growing sense of mastery. As 'theory-of-mind' researchers point out, children's attempts to get their parents to participate in their activities and 'see what they see' also signals the beginning of their understanding of the subjectivity (or internal world) of the other (Wellman 1988; Gopnik 1993).

Although the toddler can now physically separate itself from the primary caretaker on its own, it is not yet an autonomous individual. Separation anxiety is also on the rise, so expansiveness is context dependent, seen in the toddler's tendency to explore the world, but also to glance back at the primary caretaker (enjoying her from a distance) and to return for 'emotional refueling.' Bowlby referred to this as proximity seeking. According to Eagle (1984), research on

humans and animals shows that exploration is positively related to the mother's immediate availability as a safe home base. In Mahler's study, those toddlers with the best *distance contact* with the mother ventured the farthest away from her in the infant playroom.

Rapprochement. The crucial subphase in the separation-individuation process is called *rapprochement.* At the height of the practicing phase, toddlers are literally walking away from the mother, but during *rapprochement,* they alternate between actively moving away and actively approaching her. Their growing awareness that she is a separate person creates a need to establish contact, even though the need for individuation is also strong. This conflict about closeness represents the beginning of the psychodynamic mind. The child's emotional ambivalence, seen in rapid mood swings, is disharmonious for mother and infant, so *rapprochement* also refers to a phase that ends with the achievement of a new level of internal harmony. The development of the psychological resources that restore/establish harmony also creates a distinct psychological individual. Mahler's theories of a psychological individual and of how autonomy develops puts emotions on center stage in the theater of consciousness.

The onset of disharmony is seen in children's growing demandingness around fifteen to eighteen months of age. They expect the primary caretaker to fully participate in their explorations and have limited tolerance for not being attended to (when they want attention). They also learn that there are things they cannot accomplish, but may expect that their caretaker can accomplish anything. Toddlers are too concrete in their thinking to understand that parents can't do anything they want.

Rapproachement-age children also have distinct goals that are being prohibited and expectations that are being thwarted. The battle of wills regarding the child's exploration marks the appearance of a child with a separate and self-regulated internal life. Furthermore, children have a more consciously explicit preference for their *own* goals as the mother continues to put new prohibitions on them. Rudimentary fears about one's goals not being shared by others may be partly responsible for a child's demands that the mother does what he or she wants her to do.

The child's growing reliance on verbal as opposed to more direct nonverbal communication also increases the experience of separateness. Mahler and colleagues believe that enhanced competence is a double-edged sword for the child. The thrill of *mastery* is pleasurable, but mastery brings independence, which increases separation. With language comes an increased ability to deal with symbolic representations, and with that achievement come ideas in the head. Once we have ideas in the head, the problem of other minds is born — and with it existential loneliness.

Goodenough's (1931) observation that anger expression reaches it peak

around age two and then decreases conforms to Mahler's time line. As parents know, when children are angry, they are completely angry and the whole world is terrible. Temper tantrums are the most recognizable form of pure anger. This anger leads to a more extensive separation from the safety of symbiotic relatedness than the child can tolerate. Children in the 'terrible twos' aggressively resist their mother's help and push away from her to explore the world on their own, but then come running back to her in an exaggeratedly clingy and dependent way.

Children at this age are also reflexively clever, for example coercing the mother into helping them, only to reject the help when it is offered. Children probably want help, but when it comes they experience it as interference. In part, their increased willfulness leads them not just to want something done for them, but to want it done *their* way. Because the child can't distinguish between 'help' and 'interference,' the mother is in a no-win situation for a few months. This can be very trying on a parent, who is confused about what their child wants. On the child's part, since separation anxiety is resurgent while willfulness remains strong, the mother is experienced as both good and bad.

Importantly, rapproachement-age children lack the ability to remember and integrate mental states that are active under different emotional contexts (positive and negative). They are stuck in the here-and-now. This is a phenomenon familiar to clinical psychologists. For example, adults who are severely depressed cannot remember what it is like to feel happy. Persons with borderline and narcissistic personality organizations experience anger as rage — where the object of the anger is all bad. Psychodynamic clinicians have labeled this latter process *splitting*–a dissociation between positive and negative emotional states. Technically, however, children's normal splitting behavior is more related to the physical immaturity of the brain. Whether defensive or developmental in origin, splitting accentuates the disharmony experienced during the rapprochement subphase. Because negative emotional states are not modulated by or integrated with positive emotional states, *deprivation, anger*, *abandonment fears*, and *dependency* become exaggerated in the child.

The response of the primary caretaker can facilitate or constrain the achievement of harmony. The 'dual unity' or 'intense closeness' of symbiosis is emotionally satisfying for the mother as well as the child, but when the child begins to separate, the mother has to be willing to let the nature of the bond change. Foregoing symbiotic bonding is a relief for some parents and a major loss for others. Unfortunately, for both types, as soon as they work through their own separation from the infant, the child reverts to the exaggerated dependancy and demandingness of the rapprochement phase. This feels to the parent like regression. The parent who had been enjoying their child's new mastery has to tolerate renewed dependency, while the one who had been mourning the loss of symbiotic unity has to *refrain* from using the child's clinginess to regain the

closeness that was lost during the child's love affair with the world. It takes considerable emotional resources to successfully negotiate the child's conflicts about closeness, but doing so is important.

Primary caretakers must allow the child greater independence and willfulness, and also tolerate excessive clinginess. By allowing children to have their reactions and by participating in those reactions, parents help children learn to regulate and organize transient emotional states. Once a child's emotions are legitimized in this way, connections between those states are easier to make. Stern (1985) notes that participation helps children learn that specific emotional states are forms of human experiences that can be shared with others. Participation is also a means by which children learn to hypercognize and hypocognize emotions as elaborated in the folk models used by their culture.

Emotional object constancy and self-constancy. According to Mahler, the development of the autonomous individual depends on the consolidation of *emotional object constancy* between two-and-a-half and three. Emotional object constancy refers to the constancy of the love object. It entails, but is not identical with, Piaget's concept of object permanence. In object permanence, children learn that external objects continue to exist when they are out of sight. In emotional object constancy, children learn that the love object, or the 'good mother,' continues to exist when she is out of sight. This is more complex than simply remembering an object — it has an emotional component: representing (or schematizing) a *good object* across time and space. Emotional object constancy is important not only when the primary caretaker is physically out of sight, but also when the caretaker does not appear to be so good, as when punishing the child.

Emotional object constancy is important because of what it allows the child to do: explore the world with a greater sense of safety and cohesiveness, even in the face of threats. Having internalized the good mother, the child's ability to function without her immediate availability is optimized. Even when separated and upset, children have a representation of the good mother inside them. Since the mother directly regulates a child's esteem via interactions of the 'you are a good girl' type, having an internal representation of this 'good mother' facilitates the development of emotional self-regulation. Greater curiosity, frustration tolerance, and flexibility are associated with these achievements. For example, children in Mahler's study who had developed emotional object constancy chose to spend more time in the toddler room with other children, rather than being in the infant room where they knew their mothers were watching.

Bowlby, whose training analyst Joan Riviere was a committed Kleinian object relations theorist, had a similar concept which he called the *internal working model* — a representation of the relationship with mother. In the Kleinian object relations model, early interactions with the mother have positive

emotional tones and negative emotional tones, and all children develop internal representations of a good and bad mother. Bolwby makes good mother and bad mother into relational configurations. Consistent with its roots in Kleinian theory, in Ainsworth's (1979) Strange Situation studies the returning mother is the *good mother* for securely attached children, the *bad mother* for avoidant children, and the *good-and-bad* mother for ambivalent children.

In Mahler's model, emotional object constancy co-evolves with the ability to recall information available to us in one emotional state, even when we are not in that state. After its consolidation children's emotional life is not so fragmented into components, because they can recall the emotional tone of past situations even when they are not in those situations. Exaggerated 'all good' and 'all bad' experiences are less common. As a result, emotional reactions become more modulated.

According to Mahler, this time of life also marks the appearance of self-constancy (or self-coherence), a precondition for identity development (i.e., I am a boy, I am a girl). The ability to integrate information available in positive and negative emotional states provides a new kind of stability. Even for adults, the relationship between emotional integration and self-constancy is assumed. For example, in the study of adult psychopathology, specifically in the concept of multiple personality disorder, dissociation initially involves emotional states and their cognitive concomitants rather than a dissociation of cognitive states *per se*.

Fundamental emotional integration requires drawing on emotional resources to look at both sides of an issue — for example, to remember good aspects about a person when one is enraged at them. It is much easier to exaggerate something into mostly good or mostly bad. Muting felt anger with an awareness of more positive feelings, and vice versa, requires the ability to both have an immediate reaction and to step back from it at the same time. This cognition-metacognition interplay develops slowly. Although adults continue to struggle with emotional cohesiveness throughout their lives, according to Mahler the fundamental psychological equipment needed for this achievement is in place by the age of three. However, Harris (1993) points out that children are six or seven before they can consciously think about a single person as having conflicting emotional reactions to the same event.

Psychodynamically, *identification* begins to replace *introjection* at this time. An introject is something that has been internalized, but it is a presence in the child rather than part of the child's identity. For example, as a college student, I recall looking in on my little brother's nap as he manipulated a still-wrapped candy bar, saying in my mother's tone of voice "No, not now, it will be a good treat for after your nap." He lay there repeating this phrase again and again. This was an *introject*, a representation of my mother in him, but not him. It was a transitional experience. Two years later we were playing a video game, and I had

a poor start so I moved to turn the machine off and begin again, and my brother stopped me with a "No; that is cheating; you have to play the whole game!" This was also one of my mother's rules, but it was now said in his own voice. He had *identified* with it and made it a part of his character. With respect to playing the game, he regulated his own (and my) behavior in the fullest sense of the word.

With the development of emotional object constancy, what Mahler refers to as a psychological individual is born. The child has developed a nascent ability to regulate self-esteem on his or her own by drawing on internal emotional resources. Those resources are representations of good external relationships. Everyone must regulate self esteem when narcissistic gratification is not externally available, and that requires what clinicians call a *narcissistic endowment*. In non- clinical terms, to regulate self-esteem a person has to be able to call to mind internal resources that allow them to feel 'good' about themselves when a 'you are bad' psychological state is active. In contrast to the individualistic model of William James (1890), who said that self-esteem is what occurs when actual success compares favorably to idealized pretensions, object relations theorists highlight the interpersonal conditions that precede and continue to ground positive self-evaluations.

Emotional object constancy evolves further as peers and then later sexual partners become important object relationships, but the principles remain the same throughout life. It is easier to see the pattern with children because feeling good about yourself is primarily a function of "I'm good because Mom or Dad said so."

These developments in children's understanding of their internal worlds parallels their growing understanding of the internal worlds of others. According to Hoffman's (1984) theory of empathy development, the time period surrounding Mahler's rapprochement phase is also when children learn to empathize with another's emotions. Before this time, beginning at age one (during differentiation), they have only 'egocentric' empathy, where they understand that another person has feelings, but don't have enough of an internal-external distinction to understand that another's feelings may be functionally different from their own feelings. In Mahler's terms, these children understand separation, but not individuation. This stage is represented by the little boy who brings his mother a baseball to soothe her crying. Between the ages of two and three when individuation occurs, the little boy is less likely to assume that his and his mother's sadness can be dealt with in the same way.

Children's growing understanding of self-and-other, internal-and-external is supported in Michael Lewis's (1993) claim that by the age of three, children experience a full range of emotions, from primary emotions such as surprise and sadness to more self-conscious emotions such as pride, guilt, and shame.

To clarify, the object relations concept of autonomy does *not* assume

independence in the sense of radical separation. The resources that allow us to regulate esteem are internalized representations of good external relationships. These resources also need to be refueled throughout life with further good external relationships. Self-regulation, however, does require autonomous internal resources. If those autonomous internal resources are chronically inadequate, emotional self-regulation has to be gotten from the outside in the form of what Kohut (1977) calls *selfobjects*.

3. Implications: Regulation as organization

In the object relations paradigm, psychological states, including memories and cognitive representations, are emotionally-laden. Psychological organization or self-coherence therefore depends on emotional integration. Emotions such as anger or fear not only organize behavior when they are active; the development of a relatively accurate identity requires an ability to integrate all the different emotional states one is prone to adopt. Since positive and negative emotions are naturally divided, and are later defensively kept apart, self-coherence is an ongoing achievement.

Understanding the cognition-emotion relationship requires answering the question: *What regulates the psyche*? The traditional answer in Western philosophy is that integration depends on the cognitive ability (ego-strength) to integrate conflicting affect states. Integration requires either *processing capacity* or the development of a structure that can *contain* disparate elements. Freud's (1923) definition of ego as a set of brain functions responsible for differentiation, synthesis, and integration conformed to this model. Marc Lewis (1995), however, notes that scientific psychologists believe that emotions also regulate cognition; they influence, for example, attention, intention, memory, and interpretation.

Although rejecting an exclusively top-down approach in which cognitive states regulate emotions is one of the themes of this anthology, it isn't correct to say that emotional states regulate cognition either, because experiencing affect always seems to require a cognitive evaluation with respect to the concepts of self-and-other, internal-and-external. If we say that affect and cognition regulate each other, we are left either with the chicken-and-egg problem of which comes first, or with the necessity of adopting some kind of non-linear (circular or recursive) causality as discussed in dynamical systems theory (Lewis 1995; Camras 1992, Fogel *et al.* 1992).

If thinkers such as Damasio are correct, and cerebral structures are best thought of as being built out of limbic structures rather than being built on top of them, emotional and cognitive functioning may not be as neatly localizable to specific anatomical structures as Jacksonian-McLeanian models of the brain

suggest. Although this model does not require a return to mass-action theories of brain functioning, it does favor a Luria-inspired neurophysiological contextualism, where any specific anatomical structure can't be understood apart from its role in the context of a larger system.

In the Hughlings-Jackson-McLeanian 'cortex built on top of limbic system' model, cognition is like a city built on top of a river, with the emotions being the river, as in London on the Thames. Freud's early claim, based on his 19th century understanding of evolution, was that the (unconscious) emotional river continues to flow even though sometimes hidden by the more visible civil structures. Like Freud, Paul McLean-inspired thinkers in neurology also believed in some kind of interaction between 'rational' versus 'animal' functions. In psychology, some dynamical systems approaches that describe recursive relations between emotional and cognitive subsystems tend to be formulated in the context of this model (e.g., Lewis 1995).

In the Luria-like 'cortex built out of limbic system' model, cognition is more like a brain built out of the spinal chord, or better yet, built out of the body, not on top of it. We can identify different structures in the body, but we also have to understand that the body is a single organism, depending not only on organs working together, but on their being designed to work together. Similarly, our folk psychological distinction between cognition-in-general and emotion-in-general ignores the extent to which human cognition and human emotion have been designed to work together. Unconscious processes have to be more complicated than 'primary/limbic' processes that resist integration into 'secondary/cerebral' processes.

Still unanswered, however, is the important question: *What does the regulating in the psyche*? In the psychodynamic model, regulation means *organization:* how the different components of psychological functioning are organized *vis-à-vis* each other. For the classical psychoanalysts, organization referred to organization of component drives. In object relations theory, what gets organized are emotionally-laden representations of self and other. At a micro-level, an individual representation of 'self' is related to an individual representation of 'the other' via some type of interpersonal interaction. Congruent with the claim that cognition is built out of emotion, every representation of self and other is configured with positive or negative affect. *Representations of self and other* are *emotionally laden states.* At a macro-level, positive and negative self-other composites are organized more globally.

Consider self-esteem. Components of a psychological state such as self-esteem are context-specific; for example, someone has low self-esteem about playing sports, academic abilities, or establishing intimate relationships. In any context, low self-esteem means that a negative-emotion self-representation has been activated — always in implicit or explicit relation to another. If an 'all-bad'

self-representation is active, self-esteem is going to be quite low. Conversely, if an 'all-good' self-representation is active, esteem will be unusually high, a state known as grandiosity. In the object relations model, regulating self-esteem requires avoiding all-good and all-bad emotional states. More specifically, it requires maintaining awareness of a 'good' self-representation when a 'bad' self-representation is active, thereby preventing the development of an extreme or isolated 'all-bad' state (and vice versa).

In Mahler's model, regulating self-esteem requires emotional object constancy. Cognition and emotion work together so closely that not just objects, but 'good objects' continue to exist in the head when no longer externally present. Good object representations in the head help regulate a person's emotional life just as good external objects regulated their emotional lives as children. Regulation means being able to enter or modulate a specific kind of state. When emotional object constancy is a transitional phenomenon, it is an introject, an inner presence that says 'you are good.' Later, after internalization is completed, a good self-representation is part of one's identity, which corresponds to the ability to access information available to a particular state, even when one is not actively in that state. It requires being able to autonomously activate a good self-representation.

Object relations models are also *interpersonal* models. Regulation or organization is not just an internal activity. For example, even someone who has developed the ability to regulate self-esteem on his or her own still needs external input. People with emotional object constancy who experience significant narcissistic injury cannot merely activate a good self-representation; usually they need to actively seek out someone else who can remind them that they are good, helping them activate a good-self state. This process is called 'emotional refueling.' Some people, such as those with depressive personality disorder, maintain a bad self-representation state and fail to regulate it internally and externally. Even if they do get external support, it does not help them modulate their mostly bad self-representation. The good relationship in the external world fails to replace the bad relationships in the head.

This version of the object relations model is consistent with dynamical systems theory, in which there is no central organizer, even though there is organization. Organization refers to stable patterns that emerge from the constraints that activated subsystems place on each other. These subsystems exist both internal and external to the organism. This kind of emergent organization has been called self-organization. The major difference between object relations theories and dynamical systems theories is that, in object relations theories, it is not emotional and cognitive subsystems that get organized, but instead positive and negative representations of self and other.

Psychodynamic psychologists have always maintained that early patterns,

once established, have the ability to maintain themselves, even when the system and its components change in drastic ways. So a little girl's need to control her world in response to the chaos of parental alcoholism may be maintained even when she grows up, gets married, and has children of her own. Dynamical systems theorists refer to these as 'preferred states.' Preferred states involve attractors, or a stable relationship between two components.

In object relations theory, preferred states would involve emotionally-laden representations of self and emotionally-laden representations of the other in interaction. The most important preferred states are either interactions that the person experienced as disturbing, or positive interactions that they want to re-establish. In Freud's terms, the *repetition-compulsion* is the organism's attempt to actually enter the preferred state. Psychodynamic models also assume that people have unconscious preferred states, meaning emotionally laden self-other representations that they automatically enact.

The clinically important preferred states are those that appear as maladaptive behavior patterns — for example, someone who has a pattern of establishing immediate intimacy, followed in three weeks by intense jealousy and a need to control the love object. The purpose of therapy is to help people gain insight into the organization of their internal worlds, with the idea that, if these maladaptive preferred states can become more conscious, the person can learn to talk about them rather than act on them. The object relations model also states that once these preferred states are made conscious in the context of a therapy relationship, they can be reorganized in a more adaptive way. This reorganization process is called a corrective emotional experience. To illustrate, the person with the intimacy-jealousy-control pattern learns to feel what it is like to be in a relationship with boundaries, because the boundaries are imposed by the therapist (external regulation). These boundaries are slowly internalized as the client and therapist talk about all the thoughts and feelings that arise with respect to self and other in this interaction (labeling transient emotional states). As internalization proceeds, the person gains the ability to notice and modulate needs for intense closeness, which in turn influences their tendency to react with jealousy and the need for control.

Finally, I suggest that the object relations model requires a shift in how we think about emotion and cognition. Psychologists and philosophers generally agree that emotion and cognition are *basic processes*, and each group asks interesting questions about how those processes are related. Freud also thought that emotion and cognition were associated in development, and sometimes dis-associated by defensive operations such as isolation of affect. The object relations model, however, claims that emotion and cognition develop together. For example, anger depends on the cognitive competence to have a nascent self-other distinction and an awareness that internal expectations have been thwarted.

It can't exist apart from a cognitive appraisal. There are no such things as fully developed emotions and fully developed cognitions that then somehow get associated. They are naturally integrated.

Emotional regulation is not a matter of integrating basic processes; it is a matter of integrating conscious (and unconscious) content. In the philosophy of mind, content refers to representations. What get organized are positive and negative representations of self and other. In the obsessive-compulsive defense mechanism called isolation of affect, for example, emotions are not actually split off from cognitive representations. Instead, good and bad representations of self and other are split. Anger configures a self-representation that is experienced as 'bad,' and it is the composite *'anger = bad self'* that is defensively isolated from conscious awareness. Anger could not be a threat divorced from its implications for the self.

4. Conclusions

Freudian models of psychology are controversial. Contrary to popular assumptions, not every proponent of psychodynamic models is attracted by claims about the universality of the Oedipus Complex or the reducibility of all motivation to infantile sexuality. What attracts most of them is their intuitive sense that psychodynamic models of psychology emphasize all the important components and levels, and also provide some insight into how those components and levels interact. It comes down to a sense that Freud and his followers were in the ball park.

Another reason people are attracted to psychodynamic models is that their descriptions of how the psyche is organized are invariably based on what happens during child development. Freudians generally believe that understanding how infants develop organized psychological lives tells us something important about adult psychology as well. The possibility that patterns laid down in childhood can remain stable over the course of an entire life also provides a certain poetic consistency to human existence.

Under the hegemony of radical behaviorism, scientific psychologists avoided making inferences to psychological states. Although the importance of cognitive states was rediscovered by the majority of psychologists in the 1970s, a widespread recognition that emotional states and consciousness are also important lagged behind. Fortunately for psychology, neuroscientists were not so reluctant to think about emotions and consciousness, and their interest has helped make these topics respectable again in scientific psychology. Psychodynamic thinkers have always been willing to think about how emotion, cognition, and consciousness are organized *vis-à-vis* one another. Having rediscovered emotions and consciousness, it may be easier for cognitive scientists to see that, in addition to

getting into the ballpark, Freudians found pretty good seats from which to view the whole game. Hopefully, some of the ideas presented in this chapter will help philosophers of mind, neuroscientists and experimental psychologists appreciate the strengths of the psychodynamic tradition, and even consider the possibility that this tradition is worth a second look.

Acknowledgments

Glen E. Rat provided very helpful commentary during the writing of this chapter.

References

Ainsworth, M. D. S. 1979. "Infant-Mother Attachment". *American Psychologist 34*: 932–937.

Averill, J. R. 1980. "A Constructivist View of Emotion". In R. Plutchik and H. Kellerman (eds.), *Emotion: Theory, Research and Experience*: 305–399. New York: Academic Press.Bowlby, J. 1969. *Attachment and Loss, Vol.* . New York: Basic Books.

Bridges, K. M. B. 1932. "Emotional Development in Early Infancy". *Child Development 3*: 324–341.

Campos, J. J. and Stenberg, C. R. 1981. "Perception, Appraisal and Emotion: The Onset of Social Referencing". In M. E. Lamb and L. R. Sherrod (eds.), *Infant Social Cognition*: 273–314. Hillsdale, NJ: Erlbaum.

Camras, L. A. 1992. "Expressive Development and Basic Emotions". *Cognition and Emotion 6*: 269–283.

Damasio, A. 1994. *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Avon.

Eagle, M. 1984. *Recent Developments in Psychoanalysis.* Cambridge, MA: Harvard University Press.

Ekman, P. 1992. "An Argument for Basic Emotions". *Cognition and Emotion 6:* 169–200.

Fairbairn, W. R. D. 1952. *An Object Relations Theory of Personality*. New York: Basic Books.

Fogel, A., Nwokah, E., Dedo, J. Y., Messinger, D., Dickson, K. L., Matusov, E. and Holt, S.

1992. "Social Process Theory of Emotion: A Dynamic Systems Approach". *Social Development 1*: 123–142.

Freud, S. 1923/1960. *The Ego and the Id.* (J. Riviere, Trans.). New York. W. W. Norton

Goodenough, F. C. 1931. *Anger in Young Children*. Minneapolis: University of Minnesota Press.

Gopnik, A. 1993. "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality". *Behavioral and Brain Sciences 16*: 1–14.

Greenspan, S. 1997. *The Growth of the Mind.* Reading, MA: Addison-Wesley.

Hamilton, N. G. 1990. *Self and others: Object Relations Theory in Practice*. Northvale, N. J.: Jason Aronson.

Harris, P. 1993. "Understanding Emotion". In M. Lewis and J. M. Haviland (eds.), *Handbook of Emotions*: 237–246. New York: The Guilford Press.

Hoffman, M. L. 1984 "Interaction of Cognition and Affect in Empathy". In C. Izard, J. Kagan, and R. Zajonc (eds.), *Emotions, Cognition, and Behavior*: 103–131. New York: Cambridge University Press.

Izard, C. 1991. *The Psychology of Emotions*. New York: Plenum.

Jacobson, E. 1964. *The Self and Object World*. New York: International Universities Press.

Klein, M. 1964. *Contributions to Psychoanalysis, 1921–1945*. New York: McGraw Hill.

Lewis, M. 1993. "The Emergence of Human Emotions. In M. Lewis and J. M. Haviland (Eds.), *Handbook of Emotions:* 223–235. New York: The Guilford Press.

Lewis, M. D. 1995. "Cognition-Emotion Feedback and the Self-Organization of Developmental Paths". *Human Development 38*: 71–102.

Magai, C. and McFadden, S. H. 1995. *The Role of Emotions in Social and Personality Development*. New York: Plenum Press.

Mahler, M. S. 1979. *The Selected Papers of Margaret S. Mahler*, Vols. 1, 2 and 3. New York: Jason Aronson.

Mahler, M. S., Pine, F., and Bergman, A. 1975. *The Psychological Birth of the Human Infant*. New York: Basic Books.

Piaget, J. 1954. *The Construction of Reality in the Child*. (M. Cook, Trans). New York: Basic Books.

Schafferm, H. R. 1996. *Social Development*. Cambridge, MA: Blackwell.

Sorce, J. F; Emde, R. N. Campos, J. J; and Klinnert, M. D. 1985. "Maternal Emotional Signaling: Its Effect on the Visual Cliff Behavior of 1-Year-Olds". *Developmental Psychology 21*: 195–200.

Stern, D. N. 1985. *The Interpersonal World of the Infant*. New York: Basic Books.

Wellman, H. M. 1988. "First Steps in the Child's Theorizing about the Mind". In J. W. Astington, P. L. Harris, and D. R. Olson (eds.), *Developing Theories of Mind* (pp. 64-92). Cambridge: Cambridge University Press.

Winnicott, D. W. 1965. *The Maturational Process and the Facilitating Environment*. New York: International Universities Press.

CHAPTER 11

Emotions

The Fetters of Instincts and the Promise of Dynamic Systems

Gary Backhaus
Morgan State University

1. Introduction

Since the evolutionary theory of Charles Darwin, the dominant theoretical strategy for the study of the emotions has been to establish a causal relation with instinctual behavior and the physiological mechanisms of instinct. This underpinning has determined an overall research path concerning emotion, commencing with scientific descriptions of taxonomies organized around causal explanations based on phylogenetically earlier experiences, i.e., 'instincts,' or their proposed human counterparts, 'propensities,' 'drives,' or other designations meant to signify biologically determined behavior. Whether or not theorists subscribe to preformationist conceptions (developmental outcomes encoded or pre-scripted in organic matter) of innate mechanisms and behaviors that determine emotional experiences, the parameters from which subsequent research has conceived its problematic and raised its questions had been prefigured through an atomistic-reductionism. This paradigm now needs to be critically examined, for the traditional constructs are no longer viable in light of recent evidence.

Recent neurophysiological research reveals a greater interrelation of the parts and systems of the brain, with an interfusing of emotional processes in all forms of experience. The model of consciousness as a self-organizing process, and the theory of dynamic systems of development based on the general theory of open thermodynamic systems, provide a challenge to the concept of innate mechanisms of instinctual behavior. Furthermore, the newly emerging enactive theory of the mind (see Ellis 1995; Newton 1996; Varela *et al.* 1991/1993) is

based on consciousness as a self-organizing process and the fundamental role of affective valences that direct conation in the fulfillment of emotional needs. The word 'conation' has been traditionally employed in connection with goal-oriented activity or the will. In my view, conation involves the selection process of consciousness that varies in intensity. The intensity of the conation (contextual mapping through selection) concerning the heterogeneous qualities of the intentional process (consciousness of …) is a function of the significance of those qualities in terms of emotional needs. If consciousness is a self-organizing process that modifies its substratum to effect its own maintenance, then a phenomenological analysis of emotional experience is indispensable in order to conduct investigations without theoretical abstractions that prescribe for consciousness what-experience-is-like. Each developmental stage in a hierarchical open dynamic system involves a transformation, by which a new form of organization emerges that is irreducible to the aggregation of its parts. Investigative access to the level of systems organization called the 'field of consciousness' requires a phenomenological methodology as appropriate for the study of its emergent system-characteristics. But the continued deference given to the traditional taxonomies of emotion as reducible to instinctual behavior has operated under the Cartesian epistemological bias of the analytic of simples: the ultimate physiological furnishings of instincts (objective transcendent units) corresponding to the ultimate psychological furnishings of emotions (subjective immanent units). Research that isolates facial expressions and gestures, that starts its problematic with atomized conceptual correlates such as fear/flight, anger/fight, etc., foils the burgeoning paradigm that deals with ordered wholes, hierarchical systems, processes and fields, because interpretations of data continue to construct consciousness. Thus, it is necessary to undertake a critical analysis of the alleged causal relation of instinct and emotion in light of the dicta of self-organization and dynamic systems. The purpose of this paper is to focus on the need to scientifically formulate appropriate questions that will promote effective phenomenological and experimental research on 'emotions.'

2. Enactivism and recent neurophysiology

Ellis (this volume) discusses the mind/body issue in light of theoretical models that accommodate recent advances in neurophysiology while making use of the inter-disciplinary cooperation of neurophysiology, biology, psychology, phenomenology, and philosophy of science in the attempt to get an accurate picture of the organic processes of life and conscious experience. In practice, this means that a hierarchy of emergent systems and subsystems are to be revealed through coordinated efforts of researchers working on the various levels of the dynamic whole.

Enactivism has gained credence only recently with the model of consciousness as a self-organizing process and with the dynamic systems theory of development. Enactive theories dismiss the notion that the elements that make up the substratum cause the process.

> Instead the self-organizing processes that ground emotions and motivations have the power to appropriate and replace the substratum elements needed to maintain the desired pattern. Currently, the most prevalent way of explaining the causal power of self-organizing processes to rearrange and replace their own substratum elements, without at the same time violating the sufficiency of causal chains at the substratum level (e.g., efficient chemical reactions) is to be found in 'dynamic systems theory' (this volume).

Enactivism is compatible with the bi-perspectivism of the systems philosophy of Ervin Laszlo (1972), but the emergent properties of the contents of 'cognitive systems' are permeated with emotion. Dynamic systems theorists eschew computer analogies precisely because of the emotional character of the whole system (what it feels like modifies the system). Thelen and Smith, state, "As emergent, self-organized processes, emotions … are fluid, context-sensitive nonlinear, and contingent" (1998: 320). Since Thelen and Smith claim that "activity in the world, real-time activity, makes development happen" (1998: 338), only an enactive theory of consciousness is compatible with the dynamic systems approach.

According to the dynamic systems theory, self-organizing conscious processes are not incidental. It is essential for consciousness to participate in development in order for adaptive behaviors to emerge. Acts of consciousness contribute to the formation of brain processes and their physiological topography. Hebb's theory of cell assemblies (1949), the topographical dynamics of organization, has gained support through more recently acquired evidence (e.g. Kellman and Spelke 1983). "Perception and learning enter into the formation of cell assemblies" (Laszlo 1972: 158). Evidence shows that it is the global-quality of desire in the system that selects from various possibilities in the attempt to realize emotional needs. Affective processes participate in the morphological development of the brain.

In describing self-organizing processes Ellis argues that even though consciousness is not empirically observable, nevertheless, it is not a non-physical entity. If it were, the enactive theory would merge with dualism, but the agency of consciousness is grounded without positing it as an immaterial agency.

> This is possible only if consciousness is related to empirically observable physiological events as a process relates to the elements of the physical substratum.… E.g., a transverse wave takes physical particles and discrete movements of these particles as its substratum when the wave passes through that particular material medium (a sound wave through a wooden door, for

> example), but the wave is not identical with the door, nor is it caused by the door. The wave could have been the same wave even if some other material medium (e.g., a different door, or a volume of air) had been in a position to serve as its substratum (Ellis 1999: 393).

Conscious processes are multiply realizable, since they manipulate substrata in order to maintain a certain context of conation, i.e., neuronal pattern selection.

Formerly it had been assumed that sensory and motor functions of the body are topographically mapped early in development and remain functionally static through life. Ample evidence has been gathered that reveals developmental plasticity in adults. For example, Merzenich *et al.* state, *"the specific details of cortical 'representations' — of the distributed, selective responses of cortical neurons — are established and are continually remodelled BY OUR EXPERIENCES throughout life"* (1990: 195) [emphasis in original]. By selection and appraisal, which involves satisfying emotional needs, consciousness participates in the development of cortical topography. Unlike sound waves passing through a door, which do not affect its physical structure, conscious processes not only organize the substratum momentarily, they participate to some extent in the formative architecture (refashioning the door) of the substratum.

Thomas Natsoulas (1993) and Nicholas Georgalis (1994) argue that consciousness can no longer be treated as an appendage to unconscious physiological processes. Neurophysiology must be carefully applied to conscious and unconscious processes in that the brain functions quite differently in each of these two modes. LeDoux (1987) has found evidence that sensory projections from the thalamus follow dual pathways. One pathway follows to the cortex where appraisal takes place and the other pathway follows to the amygdala and the hypothalamus, which involves the emotions. These pathways converge at the subcortical level. This evidence reveals the interrelationship between objectifying and evaluating processes of the brain. This is extremely important, for it proves that the presumed Kantian functional distinctions of the psyche, which are neatly partitioned into cognition, conation, and emotion, are constructs or abstractions that do not correlate with the actual interrelation of neurophysiological processes. The taxonomy of the functions of the mind (faculty psychologism) has retarded our understanding of both conscious experience and the workings of the brain. The supposed psychological functions have set physiologists to searching for counterparts in brain tissues. The interrelationships of brain subsystems will only eventually become clear through a systems approach that can synthesize the subsystems, rather than through the analysis of parts (static states and isolated mechanisms). Traditionally theories of emotion have posited the existence of primary emotions as static states that are correlated with static instinct mechanisms. But there are no pure simples, either as functions or states. All conscious experience is inherently an axiological whole by which the organism organizes

contexts of conation according to its emotional needs. This means our taxonomies based on ready-made simples and linear causality have misled us both in descriptions of conscious experience and in explanations of physiology.

> Current neuroanatomical and neurophysiological evidence supports this *integrated view*. The basic wiring of the brain … strongly implicates a value component in all processes of learning and memory. That is, from the start, it appears that areas of the brain involved in emotion, arousal, and vigilance have access to, and are in turn accessed by, nearly all areas of higher brain function (the neocortex) and the structures involved in memory, learning, and action, as well as information about autonomic and endocrine functioning.… This means that processes of neuronal group selection in basic perceptual categorization, memory, recall, and recategorization all take place in this wash of *affective valence* (Thelen and Smith 1998: 316) [italics mine].

The emergence of conscious experience entails a subjunctive process of appraisal (Newton 1995) such that the mere process of forming a perceptual image entails a global process of the brain. Since a percept requires 'an attending to' and 'a looking for' (cognition/motivation), consciousness is already modifying brain processes according to emotional needs. But consciousness occurs only when areas of the brain associated with the process of imagination are activated (Richardson 1991). What occurs, then, is that the emotional process modifies the substratum so that what the emotion is about (specifically, vaguely, or generally) can be subjunctively entertained through imaginatively representing possible missing elements. In this way emotional needs are identified or misidentified and can be met or frustrated (Ellis 1995). Subjunctive processing is a factor in the selection processes of consciousness.

Explaining the dynamic systems model, Thelen and Smith write,

> We seek a biologically valid, but nonreductionistic, account of the development of behavior.… When developmental psychologists invoke the 'biological bases' of behavior, they usually mean the neurophysiological, hormonal, or genetic aspects of human functioning: behavior is assumed to be 'based' on these more fundamental processes.… It is … a serious error to partition the contributors to development into those that somehow reside within the organism as biological, genetic, innate, and therefore primary, and those outside the organism, which may include the everyday features of the physical and social environment, as only supportive and nonbiological (1998: xviii).

If, considering recent evidence that calls for this new paradigm, the concept of instinct is replaced by emergent developmental processes, then theories of emotion that have been hitherto based on the concept of instincts are no longer viable. Many researchers have abandoned the notion of instinct, but the fact that it has 'poisoned' the very concept of emotions has not been fully recognized. If emotional experience is explained through a reduction to biological mechanisms

of instincts, then only reactive mind/brain theories can be entertained. From these biased standpoints, consciousness as a whole and thus certainly emotion will never be properly investigated. Instincts, which are defined as fixed patterns of behavior, do not exist in the tentative, fluid, messy, novel, variable, and context-driven processes of open dynamic systems. Yet the fundamental problem is to account for how the species-typical global order, which is an obvious fact, is reconciled with the complex heterochronic details of developmental transitions, which is the evidence that dynamic systems best examines. Systems theory shows that reconciliation is possible without recourse to teleological structures (genetic nativism and vitalism), and without subscription to the denial of instincts from the standpoint of the mechanical linear causality of reflexology (constructionism/behaviorism).

Let us illustrate the difference between an instinct theory explanation and a dynamic systems explanation of a phenomenon, e.g. the lashing out tantrum of the toddler (the terrible twos). Instinct theory would have an innate mechanism for this behavior built into the organism as successful behavior for survival. The extinguishing of this behavior would have to come about through inhibiting devices brought into effect by the acquisition of learned behavior. The uncontrollable lashing out is a consummatory activity, meaning that the behavior itself is the goal of the organism. There is an external stimulus that releases or triggers this 'drive' that awaits release. If the drive intensifies and is not externally triggered, it may manifest spontaneously. In order to reduce the drive's intensification the two-year-old may become involved in exploratory activities that will lead to the experience of a stimulus that will trigger the behavior (the naughty child). The older child must learn to repress the drive to lash out in an emotional fit of anger, because that is the instinctive behavior allegedly hard-wired.

From the standpoint of dynamic systems, development is plastic and emergent. The emergence of the angry lashing out of the two-year-old child has become an attractor in the dynamic landscape; it is not hard-wired. The varying individual characteristics of this behavior in ontogeny shows that development consists of a system of attractors that must be selected by the organism. This 'natural' behavior will no longer be natural when experiences create different attractors for the behavioral manifestation of anger. Lashing out is not the natural instinct that must be repressed by the adult, because lashing out is no longer 'natural' behavior. Traditional theory proceeds as if whatever is ontogenetically earlier must be phylogenetically necessary as a static possession. The modifications of the behavior are neither merely the result of learning nor behavioral conditioning; they are processes of development, transformations of structural qualities that have been set in motion by the organism's self-organizing tendencies, which are the engine for emotional maturation. Maturation consists of transformative processes, not the results of inhibitory mechanisms. It is thought

that instincts are necessary in human beings until habits and learning can take over. But the nature/nurture debate is an ill-conceived problem, because as a dynamic system both physiology and behavior are plastic, emergent, and interrelated processes rather than things.

3. Theoretical constructs versus the phenomenology of emotions

Dynamic systems theory promotes tenets of organicism e.g. emergent qualities that constitute a totality (the whole exhibits qualitative differences that cannot be predicted through parts analysis). Thus, the emergence of a self-organizing consciousness must be understood according to an appropriate regional ontology. The purpose here is to examine more precisely how emotional experience all too often has been theoretically constructed, creating a distorted view, and why phenomenological methodology is necessary in order to study consciousness in a way that is appropriate and fruitful for mind/brain research.

William James eschewed subjective accounts of emotional experience. "I should a lief read verbal descriptions of the rocks on a New Hampshire farm as toil through them [descriptions of emotional life] again" (1890/1950: 448). But phenomenology is not the introspection of pre-theoretical nor theoretical consciousness. Both naïve and scientific introspection remain within the natural attitude and thus use emotional labels in a hypostatized and reified manner. Both simplify consciousness into ready-made states (hypostatization) that are labeled, which serve to blanket over the constituting processes. The natural attitude is the fundamental character of mundane conscious experience, which remains focussed on the object pole of consciousness of …, in a way that the constituting processes remain hidden from it. Phenomenological investigation of the constituting acts (which are open dynamic processes) is not to deal with subjective, non-scientific processes; it is to investigate how within subjectivity, experience of objectivity is constituted. Reification involves treating the contents of consciousness as if they were things, which is inappropriate for the ontological region of consciousness. The natural attitude involves already a low-grade epistemological bias towards object-contents of conscious experience. Only phenomenological methodology brackets (neutralizes) the natural attitude. Phenomenology apprehends the constitutive acts (the dynamic processes of consciousness) that are hidden by the hypostatic constructions of the natural attitude.

The specific mistake of the traditional investigation of emotion needs to be elucidated. In pre-scientific life, even a mundane perceived object is a construct that is formed through time-successions, spatial-relations and the imagination of hypothetical sense presentations that are necessary to complete the object. States of affairs are selected from a context and interpreted from a set of relevances

based on a stock of knowledge 'at hand.' This scheme of reference offers a horizon of familiarity and exhibits the character of taken-for-granted typifications. The stock of knowledge involves an open horizon for the anticipation of experiences based on typification. These typifications remain operable 'until further notice.' Everyday practicality continues in this unquestioned way until a problem arises. Through solving the problem the stock of knowledge is not merely quantitatively modified, because the solution to the problem brings about a modification in the scheme of typification, which again becomes sedimented as operable until further notice.

In terms of emotional life, everyday practicality is satisfied with typifications of emotional experience that are denoted by words such as anger, jealousy, and infatuation. The goal of the sciences is to explain the thought objects of everyday experience through the thought objects of science. Constructs of science such as molecules and moles supersede the concepts of everyday experience such as laundry detergent. In terms of the everyday constructs called the emotions, sciences that are legitimately founded on those constructs are interpretive sociology and clinical or related fields of practical psychology (school psychology). In order to grasp the social world of its participants, interpretive sociology must construct constructs of the constructs of everyday experience, i.e. create ideal types of the typifications employed in immediate experience. Clinical psychology deals with the emotions in a practical way, i.e. to help people through emotionally charged problems such that they can return to conducting their lives from the standpoint of acceptable typifications concerning their stage in life and roles in society.

However, a science that is interested in brain process as correlative to the processes of consciousness cannot merely accept the constructs that are formed on the basis of the practicalities for the successful interpretation of everyday life. These hypostatized constructs must be bracketed and the living stream of experiences must be brought to self-evidence. So the mistake of the traditional theories of emotion is the illegitimate employment of the methodology that creates scientific constructs of the everyday constructs instead of bracketing the everyday constructs such that the acts that constitute them can be brought to intuitive evidence. It is the dynamic process of the acts of consciousness that are concealed from the natural attitude and are the necessary data for research into the enactive theory of mind, self-organizing systems, and the brain physiology that is to be correlated to conscious experience.

The goal for James is the apprehension of the mechanisms that cause the contents of consciousness, which means that the experience of emotion remains hypostatized from within the natural standpoint, i.e. he naively accepts the mundane constructs. "Now, the general causes of the emotions are indubitably physiological" (James 1890/1950: 449). "Every object that excites an instinct

excites an emotion as well. The only distinction one may draw is that the reaction called emotional terminates in the subject's own body, whilst the reaction called instinctive is apt to go farther and enter into *practical relations* with the exciting object" (1892/1910: 373) [italics mine]. James is an epiphenomenalist, i.e. changes in the organism cause modifications of consciousness, which is wholly determined by physiological states. So introspection is superfluous, since there is scientific access into the real mechanisms of behavior. James held that there are more instincts in humans than in animals, for there is a greater variety of emotional responses. So the explanatory ideal would be to identify the instinctual behaviors, which provides the index for the emotions that result from them.

James states, "*Instinct is usually defined as the faculty of acting in such a way as to produce certain ends, without foresight of the ends, and without previous education in the performance*" (1890/1950: 383) [italics original]. He acknowledges that instincts are identified according to the purposes they serve, but James does not view entelechy as an explanation. "The actions we call instinctive all conform to the general reflex type" (1890/1950: 384). For example, "[The dog's] nervous system is to great extent a preorganized bundle of such reactions — they are as fatal as sneezing, and as exactly correlated to their special excitants as it is to its own" (1890/1950: 384). "Every instinct is an impulse" (1890/1950: 385), which means it results as behavior. A complex instinct action will consist of sensation-impulses, perception-impulses, and idea-impulses. A hungry lion seeks, stalks, springs, tears and devours, which are behaviors that involve all three types of impulses. "Seeking, stalking, springing, and devouring are just so many kinds of muscular contractions" (1890/1950: 385). Since the behaviors are stereotypical, then the conscious experience of the emotions must be correlatively stereotypical and predictable.

The following are a few examples of the more than four dozen human instincts discussed by James: sucking, biting, licking, spitting, grasping, pointing, swallowing, crying, smiling, protruding the lips, holding head erect, standing, walking, imitating, emulation or rivalry, hunting, appropriation, constructiveness, play, curiosity, sociability or shyness, secretiveness, cleanliness, love, jealousy. The James-Lange theory of emotion (1922/1967) sets forth that bodily expression follows directly the mental perception of a fact, and the subsequent feeling linked to the bodily expression is the emotion. For example, one apprehends that one is being approached by a rabid woodchuck then one instinctively runs away, and due to the behavior, one then feels the emotion of fright.

Our purpose is to note how experiences are simplified as nameable states, e.g., anger, astonishment, or fear, that are only scientifically significant due to their connection to behavioral impulses, that is, mechanisms of explanatory value. Otherwise description of conscious experience is merely subjective, non-scientific data. The James-Lange theory continues to motivate research centered

on distinguishing emotions based on patterns of peripheral physiological activity, e.g., Levenson *et al.* 1990 and Levenson *et al.* 1991. But the critical point maintained here is that treating the conscious life in this manner is as simplistic as a physiology that would be content with legs/walking, arms/carrying, and hands/grasping. Consciousness is incredibly complex; so is the brain; and so is development; but no contemporary neurophysiologist is content with naming the main parts of the brain and assigning to each a function. By treating emotions as the function of survival responses insured through innate mechanisms, consciousness becomes a bundle of fixed states. It should be quite obvious that if consciousness, as a complicated self-organizing process, modifies its material substrate, then it cannot be reduced by labeling static structures, which is a distorted objectification and simplification. James serves as an exemplar of those who treat emotion as an experiential index that confers subjective value on the already prepared (in the physiology of the animal) survival-related adaptations that prescribe to the animals what to do (Breger 1974).

Robert Plutchik defines emotions as "adaptive devices in the struggle for individual survival at all evolutionary levels" (1962: 56). The first goal involves identifying the basic dimensions applicable to all organismic levels. He enumerates eight prototypic dimensions of emotion: incorporation, rejection, destruction, protection, reproduction, deprivation, orientation, and exploration (1962: 63). Under each prototype he lists the emotions in order of intensity: (destruction — rage, anger, annoyance); (reproduction — ecstasy, joy, happiness, pleasure, serenity, calmness); (incorporation — admission, acceptance, incorporation); (orientation — astonishment, amazement, surprise); (protection — terror, panic, fear, apprehension, timidity); (deprivation — grief, sorrow, dejection, gloominess, pensiveness); (rejection — loathing, disgust, dislike, boredom, tiresomeness); (exploration — anticipation, expectancy, attentiveness, set). Emotions are treated analogously to the color-wheel by providing laws for their admixtures. For example, anticipation, anger, joy, acceptance, surprise, fear, sorrow, and disgust are placed on the wheel adjacent to one another. Diametrically opposed are disgust/acceptance, anticipation/surprise, anger/fear, and joy/sorrow. "A mixture of any two primaries may be called a dyad, of any three primaries, a triad.... If two adjacent primaries are mixed, the resulting combination may be called a primary dyad. Mixtures of two primary emotions which are once removed on the circle may be called secondary dyads, while mixtures of two primaries which are twice removed on the circle may be called tertiary dyads" (1962: 115–16). Some examples of primary dyads: anger + joy = pride; joy + acceptance = love, friendliness; acceptance + surprise = curiosity. Examples of secondary dyads: joy + surprise = delight; acceptance + fear = submission, modesty; surprise + sorrow = embarrassment, disappointment. Examples of tertiary dyads: anger + surprise

= outrage, resentment, hate; joy + fear = guilt; acceptance + sorrow = resignation, sentimentality.

Plutchik's color analogy constructs emotional experience through a mixture of primary building blocks. This theory and others like it heavy-handedly prepattern emotional experience. Its reification of states obfuscates the description of the experiential process; it biases investigation by promoting interpretations that fit the objectified schematic.

A description of emotional experience that does not distort its ontological character, its organizational processes in the hierarchy of systems, is vital to research that seeks to understand correlative events in the brain and the organism. If emotional experience is theoretically constructed through an assumed epistemology, then self-organizing adaptive processes of conscious experience will never be understood according to its systemic properties as a natural-cognitive system; dualism or physicalism remain the only alternatives. As I have argued, phenomenology provides the only access to experience in a manner that does not objectify consciousness. Eugene Gendlin offers corroboration,

> The job of … [phenomenology is] to undercut and open up the supposedly packaged units supposedly given. Supposedly stable things like emotions, perceptions, images, ideas, and experiences cannot be used as a beginning.… One would not want to talk of emotions or experience without such talk touching upon experience.… We have had too much of the sort of thinking which wants to substitute thin schemes with experience (1973: 368).

Phenomenological exploration of the emotional valence of experience as a dynamic process provides the means for a far more sophisticated and adequate research into the mind/brain system.

The basic mistake, that traditional theory has constructed for experience what its emotional life is about, is, nevertheless, recognized by theorists. LeDoux states, "The proper level of analysis of a psychological function is represented in the brain. This leads to a conclusion that clearly falls into the bizarre at first — that the word 'emotion' does not refer to something that the mind or brain really has or does. 'Emotion' is only a label, a convenient way of talking about aspects of the brain and its mind" (1996: 16). Robert Plutchik states,

> A major element … is that an emotion is a subjective feeling of a certain kind — the kind for which labels such as angry, disgusted, and afraid are appropriate. However, there is considerable evidence to suggest that this is too narrow a way to define emotion and that the facts available to us imply the need for a broader conceptualization. In contrast, it may be proposed that an emotion is not a subjective experience per se, but rather a construct or inference based on various classes of evidence (1980: 4).

These later theorists recognize the distortion but do not recognize that the appropriate manner to investigate consciousness is through phenomenological analysis, which means that the epistemological bias of the natural attitude, with its hypostatization and reification, is not neutralized. The dynamic systems model is complemented by phenomenology, but without recognition of this strategy, there is no alternative but to employ the constructs that continue to maintain a no longer viable explanation that had been historically developed through the theory of instincts.

4. Historical synopsis of the paradigm of instinct

From the theoretical linking of the emotions with instincts emerged fundamental debates. "The ideas that emotions are the manifestations of instincts in more complex animals is clearly stated in Darwin's classic, … and may be found in the work of William James, William McDougall, Lloyd Morgan. McDougall, Lorenz, and Freud would all agree that human instincts are manifest whenever one observes emotions" (Breger 1974: 26). A synopsis of the historical research shows how the dynamic systems approach reinterprets evidence in a way that eliminates the traditional concept of instinct, which frees the study of emotion from the constraints of the old paradigm.

After reading Sir Charles Bell's *The Anatomy and Philosophy of Expression* (Bell 1840/1877), Darwin became convinced of the natural origin of the emotions (Darwin 1904), which led to his last published work, *The Expressions of the Emotions in Man and Animals* (1872). Bell held that the muscles used in emotional expression were created expressly for that function. Darwin countered with two doctrines: serviceable associated habits and antithesis. Emotional expression in humans is the remnant (epiphenomenon) of behavior that is immediately useful in animal life. For example, the expression of disgust is the serviceable associated habit that has its origin in the animal's action of spitting out non-delectable food substance. Since not all emotions can be accounted for in this way, Darwin introduced the doctrine of antithesis. A state of mind that involuntarily occurs along with habitual actions is opposite to a state of mind that involuntarily occurs along with opposite habitual actions (Darwin 1872/1965: 28). Darwin attempted cross-cultural studies of facial expressions and the comparative study of non-humans. Since the postulate of universal facial expressions in the expression of emotions suggests patterns for which humans have a native aptitude, this theory leads to the concept of instincts. Appearing after Darwin are the opposing theories of the conditioned responses of the behaviorist school, which deny innate mechanisms of instinct; the phylogenetically determined behavior patterns of ethology, which promote innate sources for observable

behavior; and the non-physiological vitalistic instincts of purposive psychology.

Opposing the theory that human behavior can be reduced to hedonistic motivations, William McDougall postulated his hormic theory (horme means instinct in Greek) in his purposive psychology. McDougall's vitalism rejects physiological explanations in favor of teleology. McDougall regarded the instincts as agencies neither in need of nor accessible to a natural explanation. In *An Introduction to Social Psychology* (1926), McDougall enumerated the principal human instincts and their correlative emotions: flight/fear, repulsion/disgust, curiosity/wonder, pugnacity/anger, self-abasement/subjection, self-assertion/elation, parental instinct/tenderness, and reproduction/sexual desire. In *The Energies of Man* (1933), McDougall employed the word, 'propensities,' rather than 'instincts' in order to account for the greater complexity and plasticity of human behavior over the stereotypical behavior of animals. Yet the continuity between animals and humans is maintained through the innate character of propensities.

Behaviorism arose out of Wilhelm Wundt's association theory and the reflexology of Ivan Pavlov. John B. Watson, the founder of behaviorism, based behavior on the conditioned reflex, which was a linear causal chain of stimuli and responses. Behaviorists hold that all behaviors are learned; there exist no innate capacities or instincts. The polemics between the purposive psychologists and the behaviorists represent the extremes of the nature/nurture debates. If the purposive psychologists are right, then emotions are caused by instincts, which are viewed as vital teleological non-physiological mechanisms. If the behaviorists are right, then the emotions are merely the result of conditioning (learned responses) and can be manipulated through behavioral modification.

Ludwig von Bertalanffy recognized the underlying affinity of these two positions. "Paradoxical though it may seem, the machine theory is the foundation of both biological mechanism and vitalism" (1933/1962: 44). Bertalanffy advocates holism in his theoretical biology.

> The fundamental error of 'classical' mechanism lay in its application of the additive point of view.... It attempted to analyse the vital process into particular occurrences proceeding in single parts or mechanisms independently of one another.... Vitalism ... believed these to be co-ordinated by an immaterial, transcendent entelechy. Neither of these views is justified by the facts (1933/1962: 177).

Bertalanffy is a progenitor of dynamic systems (Thelan and Smith 1998: xix).

> The solution of this antithesis ... is ... system theory ... which ... sees the essence of the organism in the harmony and co-ordination of the processes among one another ... through the forces immanent in the living system itself (Bertalanffy 1933/1962: 178).

"The 'whole' … means the momentary total state of the living system, not the typical end-state to be reached in the future" (1933/1962: 179). This co-ordination of processes is teleonomic, which is not to say that purpose is built in, but that it autochtonously emerges from the adaptive organizational transformations. "Development is, at last, epigenesis, i.e. neo-formation of manifoldness" (1933/1962: 180). At the time of this precocious discussion there was not much hard evidence to support it.

The ethologists presented a solution to the stymied polemics of behaviorism and vitalism based upon evidence that promoted innate and learned interrelations. Charles Otis Whitman (1919) discovered homologous patterns of movement recognized to be as species distinguishing as comparative morphological structures. Just as morphology originates phylogenetically and is encoded in the genome, so are patterns of behavior. Against the purposive psychologists, Konrad Lorenz acquired evidence showing the existence of blind or senseless sequences of innate behaviors. Against the behaviorists, his data showed that not all animals' behaviors are learned.

Nikolaas Tinbergen (1951) described instincts as species-specific and genetically determined programs of behavior. The stereotyped pattern of movement functions as the consummatory activity. The goal of the instinct is not the accomplishment of an end, as it is in conscious pursuit of a goal, but the consummatory activity itself. The consummatory activity is readied by an intensifying drive state in the organism. The consummatory activity is usually triggered by the key stimulus in the external environment. Fixed motor patterns are dependent on the external stimuli only for their release. However, if the drive is intensified and a key stimulus is unavailable, the consummatory activity may spontaneously occur as a 'vacuum activity.' Appetency behavior places the animal in contact with key stimuli. It 'seeks out' the stimuli needed to trigger the consummatory activity.

One notable criticism of the theory of innate action patterns came from Z. Y. Kuo (1932), who argued for abandoning the partitioning of innate and acquired behavior, since all behavior consists of interaction between the organism and its environment. Lorenz (1981: 8) then posed the solution that innate and acquired are not disjunctive opposites. Phylogenesis provides the foundational parameters such that species-preserving behavior patterns are the ones that are learned. This solves the enigma as to how the right processes are learned, or how adaptive improvements are effected. P. K. Anokhin (1964) put forth that a feedback circuit reports success or failure to the initial mechanisms of the antecedent behavior. Behavior patterns are adapted to environmental situations through either phylogenetic mutation in natural selection or though the ontogeny of an individual who acquires pertinent adaptive skills. The phylogenetic apparatus places pressure on the organism to survive by reinforcing teleonomic

successful behavior and extinguishing unsuitable behavior. Thus, there are innate constraints that operate on the organism to insure that its adaptive modifications favor survival.

Irenaus Eibl-Eibesfeldt accomplished much ethological research concerning humans. Cross-cultural analyses confirm that phylogenetic adaptations determine the social interactions of human beings in various ways. Many expressive movements serve to trigger certain behaviors and thus act as signals for innate releasing mechanisms. He argues that friendly encounters are determined extensively through inborn responses. Displays are precautions against domination, but patterns for bonding and resolving conflict mollify display behavior (1980: 77). Eibl-Eibesfeldt advocates the ethological program (1975) to systematically map the human ethogram. Ethology has provided the most plausible arguments and interpretations of data for the existence of innate behavioral patterns of emotional expression. However, plasticity in behavioral development must be given account, for there is also great variability in action.

Conrad Hal Waddington created diachronic biology, a science of embryology-genetics-evolution through which he developed the epigenetic landscape model, which later becomes an integral part of dynamic systems theory. Epigenesis occurs when environmental signals act upon the genome, which bring about expression of all the morphological and behavioral characteristics displayed by individual organisms at different points during the life span. Waddington was a Whiteheadian who recognized the biological equivalent of concrescence, 'chreod,' i.e. a buffered pathway of change. Pathways are canalized, which means developmental processes are continually adjusted. Some are more constrained so that certain characteristics of the phenotype will be produced across environmental differentiation. Other epigenetic processes are loosely canalized, which allows for phenotypic characteristics that are quite variable on the individual level. The landscape image depicts genetic buffering by a ball rolling into progressively deeper valleys over time. Since development is plastic it is possible for a cell to roll into various valleys which only progressively determine its fate. Waddington was able to shed light on how, even though genetic inheritances and environmental conditions exhibit variability, development still produces species-specific phenotypical characteristics (Gilbert 1991: 181–203). But Waddington's landscape does not entail built-in 'end-states' — teleological scripts.

Waddington's epigenetic landscape and open dynamic systems account for constraints without recourse to the agency of innate mechanisms. Pasko Rakic (1991) employs the epigenetic model to describe the ontogenetic processes that produce the primate cerebral cortex. Rakic found that patterns of neuronal interconnections in the cortex are modified. Plasticity is observable in experiments in which the environment of the developing embryo is altered. This

research shows the inadequacy of explanations based on innate mechanisms as agencies of development. Lorenz accounted for plasticity while maintaining constraints on plasticity through innate mechanisms that he claims exist to insure the survival bias of phylogenesis. Waddington's landscape provides a means to a solution that does not rely upon innate mechanisms but still accounts for evolutionary history — phylogenesis.

5. The dynamic landscape and the elimination of innate mechanisms of instincts

Thelen and Smith (1998) apply Waddington's model to post embryological development by creating the dynamic landscape. Their model does not view development as a progression toward increasing stability, but as a series of modifications of stability/instability. More importantly, they argue that the constraints are not ultimately genetic. In their ontogenetic landscape the hills and the valleys both deepen and become shallow such that preferred behaviors emerge and disappear. The dynamic systems model explains how within a species-typical global order of development there are details of development that are fluid, idiosyncratic, messy and complex. "What turning up the microscope reveals is that individual activity — real-time perceiving, moving, remembering — constitutes the driving force of change" (1998: 311). Their theory describes "how the basic processes of neuroembryology were themselves dynamic and contingent, and how these epigenetic processes built a brain wired to benefit from the time-locked properties of the input — the multimodal consequences of experience" (1998: 311).

Epigenesis explains developmental processes as contingent and historical. Properties emerge through interactions, which are not scripted in genetic codes. Replacing instinctual mechanism as the primary developmental agency is the process of selection (Edelman 1988). Disregarding the specific problems inherent in the details of Edelman's theory of neuronal groups, what remains is that selection creates self-organized maps through reciprocal interactions that are activated by the organism's current and past interrelations with its environment. Phylogenetic adaptation can only broadly anticipate what a life ontogenetically will be like. The details must be acquired and are required for and through the processes of development. Examining development at the local level shows that phylogenesis contributes attractor basins. These are not constraints, but graded fields of possibilities. Some basins are quite stable, such as for walking, which is only disrupted by injury or prolonged inactivity (Thelen and Smith 1998: 145). Many systems are much less stable and variability is the key to understanding modifications. These attractor basins are never fixed; very stable developmental

processes are the result of deeper furrows in the landscape. Through ontogenesis individual differences reveal the range of possible state spaces for the system and the possible trajectories between stable attractors (1998: 145) "Organisms are active seekers of solutions, they have biases and goals, they try out various alternatives, and the fit is *ad hoc* rather than prescribed" (1998: 146). "While guided by the genes, complexity of form must arise *during development* in self-organizing fashion: the genes cannot store the information on the time and space position of each cell of each animal" (1998: 153).

Dynamic systems theory is capable of explaining the seemingly contradictory status of the global order and the local details through the role of context — the here and now. The context makes the global order (global order does not make context), because it is through repeated experiences that the global order is developed. Context selects the global order to insure the plasticity of acts. Context adapts the global order by fitting the history of past here and now experiences with the task at hand. So global order is the result of patterns of real-time activities of time-locked and reentrant systems (sensory inputs of the moment, the preceding activity, and the history of activity) (1998: 236). In the example of the two-year-old who lashes out when 'angry,' the enlarging of experiential contexts adapts a global order such that this behavior no longer operates as an attractor yet still sediments in the historical process of the organism. A temptation is to view attractors as innate mechanisms that then become inhibited. However, attractors "are not things; they are trajectories of activity of neuronal groups through time — trajectories built through the reentrant mapping of heterogeneous systems" (1998: 217). Here and now exploratory experiences "widen the attractor basin and enrich the linkages" (1998: 222). This means that emerging ways of dealing with 'anger' are made possible through the enrichment. The experiential context carves out the ridges and valleys in the landscape, which is always in process of revision.

6. Conclusion

A dynamic systems approach does not reduce emotions to preformed innate mechanisms. Dynamic systems are created through the interrelation of subsystems in a hierarchical context. There is no template or static structure that can be called anger or fear. Rather, there exist dynamic assemblies that are a function of global activity. Spatial maps of nerve cell assemblies are the result of the here and now experience as well as past experiences that are assembled in response to a specific environmental situation. Instead of internal instinctual maturation as the guide for emotional development, development takes place through the intrinsic dynamics in ontogenesis. The variability of emotional life must be

accounted for through extensive phenomenological studies of individuals. Novel forms emerge from the self-organizing properties of complex systems, so that the study of emotional development must focus on an individual's history over time. Phenomenological investigation of conscious experience provides the appropriate access for the description of experience such that the study of dynamic systems at work in the brain and the nervous systems can be correlated with the self-organizing processes of consciousness. A team of experts that coordinate the various levels of dynamic systems would best serve future research concerning conscious experience and the role of emotions in the development of human beings. Only the dynamics systems approach respects the complexity of both the organism and consciousness, but from a standpoint that is univocal in its explanation.

References

Anokhin, P. K. 1964. "Systemogenesis as a General Regulator of Brain Development". *Progress in Brain Research* 9: 54–86.

Bell, C. 1840/1877. *The Anatomy and Philosophy of Expression as Connected with the Fine Arts.* London: George Bell.

Bertalanffy, L. von. 1962. *Modern Theories of Development: An Introduction to Theoretical Biology.* New York: Harper & Brothers.

Breger, L. 1974. *From Instinct to Identity: The Development of Personality.* Englewood Cliffs: Prentice-Hall, Inc.

Darwin, C. 1872/1965. *The Expression of the Emotions in Man and Animals.* Chicago: University of Chicago Press.

Darwin, C. 1904. *The Life and Letters of Charles Darwin.* New York: Appleton.

Edelman, G. M. 1988. *Topobiology: An Introduction to Molecular Embryology.* New York: Basic Books.

Eibl-Eibesfeldt, I. 1975. *Ethology: The Biology of Ethology* (2nd ed.). New York: Holt, Rinehart, and Winston.

Eibl-Eibesfeldt, I. 1980. "Strategies of Social Interaction". In R. Plutchik and H. Kellerman (eds), *Emotion Theory, Research, and Experience VoLume 1 Theories and Emotion.* New York: Academic Press, 57–80.

Ellis, R. D. 1995. *Questioning Consciousness.* Amsterdam: John Benjamins.

Ellis, R. D. 1999. "Why Isn't Consciousness Empirically Observable? Emotion, Self-Organization, and Nonreductive Physicalism". *Journal of Mind and Behavior* 20: 391-402.

Gendlin, E. T. 1973. "A Phenomenology of Emotions: Anger". In D. Carr and E. S. Casey (eds), *Explorations in phenomenology*. The Hague: Martinus Nijhoff, 367–398.

Georgalis, N. 1994. "Asymmetry of Access to Intentional States". *Erkenntnis* 40: 185–211.

Gilbert, S.F. 1991. "Induction and the Origins of Developmental Genetics". In S.F. Gilbert (ed), *A conceptual history of modern embryology*. Baltimore: Johns Hopkins University Press, 181–206.

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

James, W. 1890/1950. *The Principles of Psychology: Volume Two.* New York: Dover Publications, Inc.

James, W. 1892/1910. *Psychology.* New York: Henry Holt and Company

James, W. and Lange, C.G. 1922/1967. *The Emotions.* New York: Hafner Publishing Company.

Kellman, P.J., and Spelke, E.S. 1983. "Perception of Partly Occluded Objects in Infancy". *Cognitive Psychology* 15: 483–24.

Kuo, Z.Y. 1932. "Ontogeny of Embryonic Behavior". In Aves I and II, *Journal of Experimental Zoology* 61.

Laszlo, E. 1972. *Introduction to Systems Philosophy.* New York: Harper & Row.

LeDoux, J.E. 1987. "Emotion". In F. Plum & V.B. Mountcastle (Eds), *Handbook of physiology. The nervous system: Vol. 5. Higher Function.* Washington, DC: American Physiological Society, 419–59.

LeDoux, J.E. 1995. "In Search of an Emotional System in the Brain: Leaping from Fear to Emotion and Consciousness". In M.S. Gazzaniga (ed), *The Cognitive Neurosciences*. Cambridge: MIT Press, 1049–61.

Levenson, R.W., Carstensen, L.L., Friesen, W.V., & Ekman, P. 1991. "Emotion, Physiology, and Expression in Old Age". *Psychology and Aging* 6: 28–35.

Levenson, R.W., Ekman, P., & Friesen, W.V. 1990. "Voluntary Facial Action Generates Emotion-specific Autonomic Nervous System Activity". *Psychophysiology* 27: 363–384.

Lorenz, K.Z. 1981. *The Foundations of Ethology.* New York: Springer-Verlag.

McDougall, W. 1926. *An Introduction to Social Psychology.* Boston: T.W. Luce & Co.

McDougall, W. 1933. *The Energies of Men.* New York: C. Scribner's Sons.

Merzenich, M.M., Allard, T.T., and Jenkins, W.M. 1990. "Neural Ontogeny of Higher Brain Function: Implications of Some Recent Neurophysiological Findings". In O. Franzn & P. Westman (eds), *Information Processing in the Somatosensory System.* London: Macmillan, 293–311.

Natsoulas, Thomas. 1993. "What is Wrong with Appendage Theory of Consciousness". *Philosophical Psychology* 6: 137–54.

Newton, N. 1996. *Foundations of Understanding*. Amsterdam: John Benjamins Publishing Company.

Plutchik, R. 1962. *The Emotions: Facts, Theories, and a New Model.* New York: Random House.

Plutchik, R. 1980. "A General Psychoevolutionary Theory of Emotion". In R. Plutchik & H. Kellerman (eds), *Emotion: Theory, Research, and Experience* 3–33 [Volume 1 Theories of Emotion]. New York: Academic Press.

Rakic, P. 1991. "Plasticity of Cortical Development". In S.E. Braugh, W.S. Hall, & R.J. Dooling (eds), *Plasticity of Development*. Cambridge: The MIT Press, 127–61.

Richardson, J. 1991. "Imagery and the Brain". In C. Cornoldi & M. McDaniels (eds), *Imagery and cognition.* New York: Springer-Verlag, 1–46.

Thelen, E. and Smith, L. B. 1994. *A Dynamic Systems Approach to the Development of Cognition and Action.* Cambridge: The MIT Press.

Tinbergen, N. 1989. *The Study of Instinct.* Oxford: Clarendon Press.

Varela, Francisco, Evan Thompson, and Eleanor Rosch. 1991/1993. *The Embodied Mind.* Cambridge: The MIT Press.

Whitman, C. O. 1919. "The Genetic Standpoint in the Study of Instinct". In H. A. Carr (ed), *The Behavior of Pigeons: Posthumous Works of Charles Otis Whitman.* Washington: The Carnegie Institute of Washington, 87–91.

CHAPTER 12

Awareness of Emotions

A Neuropsychological Perspective

Martin Peper
University of Freiburg, Germany

From a neuropsychological perspective, awareness of emotions is a complex function involving several components (perceptual decoding and conceptualization, memory and attention, and psychophysiological responses). Pathological conditions of the nervous system as well as certain experimental procedures in healthy persons may induce dissociations of these components. It is suggested that perceptual awareness of an emotional stimulus requires a correct stimulus identification as well as input monitoring. Awareness of experiential qualities is a more global function involving integration of interoceptive information, formation of emotional schemas or concepts, and recall of episodic memory of past emotions. Perceptual awareness of internal or external stimulus events can be defined and measured by means of psychophysical methods. Experiental qualities, however, are difficult to assess in a reductionist/physicalist framework.

1. Introduction

Recent advances in the neurosciences have extended our understanding of the network mechanisms regulating emotional experience and behavior (e.g., LeDoux 1996; Rolls 1999). As part of this emerging discipline of 'affective neuroscience' (Davidson and Sutton 1995; Panksepp 1991, 1998), neuro-psychological theories have offered explanations of how the brain controls the perception and evaluation of exteroceptive emotional stimulus events, expressive behavior, psychophysiological responding, and emotional experience. These findings have

been integrated into various models of emotion (for reviews, see LeDoux 1995, 1996; Peper and Irle 1997a; Plutchik and Kellerman 1986; Scherer and Peper 2000). Within the suggested neuropsychological theories, the identification of the cerebral basis of 'conscious' versus 'unconscious' emotional processes has been an important question.

Measurement and functional description of consciousness have always been major themes of psychology. Recently, experimental research on 'perception without awareness' appears to have moved back into the focus of mainstream psychology (Bornstein and Pitman 1992; Kihlstrom *et al.* 1992). In recent years, the definitory and methodological precision of experimental approaches regarding consciousness has considerably increased, and is replacing highly speculative terms by constructs with precise operational definitions. Nevertheless, this line of research has oversimplified or rather entirely neglected the underlying neural mechanisms of preattentive emotional phenomena, in contrast to work on 'concious' emotional experience in which neuroscientists interested in emotion and motivation have described the relevant cerebral architecture (see Jacobson 1995; LeDoux 1996; Panksepp 1998; Rolls 1999). Most of these approaches, however, fail to elaborate on the measurement and operationalization of awareness of emotional stimulus events in humans.

The present chapter aims at discussing the concepts of awareness and emotion from a neuropsychological perspective. It contends that a systematic approach must, first, agree upon a definition of emotional awareness. Second, operational definitions and appropriate psychological methodology are needed to assess processes such as preattentive perception of external objects or intrinsic emotional states. Third, possible causes for dissociated component functions of emotions in humans are discussed. Neuropsychological hypotheses, in addition to case histories of brain damaged patients, are considered to elucidate variations of emotional experience which may occur under certain pathological conditions.

Aside from methodological problems, research on awareness of emotional qualities has been complicated by metatheoretical difficulties. Neurophilosophy may help to clarify reasons for divergent interpretations and methodological preferences of the respective subdisciplines in the neurosciences. The data appears to support an emergentist materialist position. Neurophilosophical considerations may instigate new impulses for designing advanced behavioral methods and paradigms. For example, an aim of future endeavours could be to capture more precisely the 'experiential space' of the individual. In addition, the conditions giving rise to a coupling or decoupling of emotional experience and automatic emotional reactions may need to be better specified.

2. Awareness of emotions

2.1 *Concepts of emotional awareness*

Emotion is typically interpreted as a multi-facetted construct with several functional components: (a) the detection and perception of an exteroceptive or interoceptive stimulus event; (b) a cognitive component for conceptualization and appraisal; (c) a neurophysiological component for activation and arousal, homeostatic regulation, adaptation and energy provision for instrumental action; (d) a response mediation component for the communication of responses and intentions and for goal-directed action; (e) a motivational component for the planning and direction of instrumental behavior; and (f) a component for subjective feelings and mood (cf. Scherer and Peper 2000).

One important component, that of emotional perception, refers to the decoding of stimulus events. Emotional perception of an exteroceptive object or situation is assumed to include an initial evaluation of novelty, familiarity and personal relevance of the object (Scherer 1993; Scherer and Peper 2000). A complete encoding of an emotional object includes not only the detection of surface features and recognition of its identity, but also an identification of higher-order dimensions such as pleasantness or need significance. The process of conceptualization — that is, of integrating representations of an emotional object into conceptual or knowledge systems — is an integral component of emotional processing (Peper and Irle 1997b, c; see also Rolls 1999).

Emotions have particular importance within the context of social decision making and conduct. It is increasingly recognized that automatic evaluations of norm compatibility use representations of tacit emotional or social knowledge that are not always consciously accessible (Scherer and Peper 2000). Emotional experience has been associated with the component function of subjective feeling and mood (see Averill *et al.* 1994; Buck 1993). The object of emotional experience is not only found in internal 'limbic' or peripheral sources of activation; it also designates intrinsic ('phenomenal') qualities of external emotional objects or events. These experiential qualities are generated in awareness.

2.2 *Dissociations of emotional components*

The subsystems of emotion are assumed to show a temporary synchronization during an emotional episode (Scherer 1993). However, previous research has not unequivocally supported the hypothesis of convergent indicators of introspective, behavioral and psychophysiological components in emotionally relevant situations. Rather, findings from several disciplines provide evidence for dissociations of the data sources. Psychophysiological findings, for example, suggest that

converging subjective reports and autonomic reactions during emotional states appear to be the exception rather than the rule. For example, results of field studies with portable data recorders indicate that healthy humans may show 'silent' autonomic changes of potential emotional origin (e.g., an 'additional heart rate', not related to motor activity) that is not registered and reported by the subject (Fahrenberg and Myrtek 1997; Myrtek *et al.* 1997).

Similarly, decades of psychophysiological research on the personality trait of emotionality have not provided convergent validation of this assumption. An influential hypothesis stated that persons who describe themselves as highly anxious, nervous and depressed should, due to altered arousability of limbic brain structures, show consistently greater reactions in autonomic and central response measures under certain types of situational strain (Eysenck 1967). However, psychophysiological data show that self reports on emotional episodes as assessed by questionnaires are only very weakly associated with central or peripheral response measures. A recent meta-analysis of previous studies suggested that self-descriptions of emotionality are only poorly associated with changes in psychophysiological reactivity (Myrtek 1998).[2] Among other reasons, our limited capacity to monitor intrinsic reactions and behavior as well as to recall episodic memory of emotional events could be responsible for this divergence.

Moreover, many attempts have been made to experimentally induce dissociations of explicit knowledge of an aversive stimulus and autonomic responding. Numerous paradigms have been used to assess perceptual or memory functions, where stimulus qualities or contingencies were not or only partially available to awareness. Examples of such non-declarative or implicit learning procedures are: classical conditioning paradigms (conditioned autonomic responses, evaluative conditioning, subliminal conditioning); instrumental and operant conditioning (verbal conditioning); implicit learning (learning of complex problem-solving strategies, concept formation); and implicit memory (semantic priming research, preference formation, concept identification tasks).

In neuropsychology, anecdotes, case reports and group studies of patients with focal brain damage illustrate the fact that dissociations of functional components of emotions may occur under certain pathological conditions of the nervous system (Scherer and Peper 2000). This field of research has contributed to revealing the neural basis of dissociation phenomena. The findings of divergent emotional components may be attributed to the partial and incomplete coupling of verbal, behavioral and physiological response systems under certain experimental conditions.

2.3 *Operationalization and assessment*

We begin with a discussion of constructs and definitions. The relation between awareness and perception has been extensively studied in cognitive psychology (Bornstein and Pittman 1992). In the mid-eighties, this field was involved in intensive discussions as to whether 'subliminal perception' exists and how it could be operationalized (for reviews, see Bowers 1984; Cheesman and Merikle 1984, 1985, 1986; Dixon 1971, 1981; Holender 1986; Kihlstrom 1987; Kunst-Wilson and Zajonc 1980; Merikle 1982; Schachter 1987). Many studies on subliminal lexical priming effects have been influenced by the work of Marcel (1983). A host of constructs such as subsception, unconscious/nonconscious/subliminal/implicit perception, perception without awareness, preattentive or automatic processing have been suggested for the description of phenomena that escape 'conscious' control (see Bornstein and Pittman 1992; Kihlstrom 1990). These cognitive approaches confront many terminological and definitory difficulties.

The notion of 'subliminal perception' has been rejected since the concept of the limen may have 'unfortunate psychophysical implications' (Kihlstrom *et al.* 1992a,b). Suggestions were made to abandon the concept of the limen in favor of the concept of 'implicit perception' (Kihlstrom 1987, 1990). In contrast, explicit perception "refers to the person's conscious perception of some object or event in the current stimulus environment" (Kihlstrom *et al.* 1992: 22). It can be operationalized by qualitative descriptions of the object, such as form, color, distance or identity, although sensory detection is also required for correct operationalisation. However, implicit perception does not necessarily refer to objects or events in the stimulus environment and "does not require the subject to perceive any object, qua object, at all". Rather, it can be regarded as "any change in experience, thought or action that is attributable to some past experience even in the absence of conscious recollection of that event" (Kihlstrom *et al.* 1992: 22).

Similar definitions of preattentive phenomena have been suggested in the field of learning and memory research (e.g., Berry and Broadbent 1987; Graf and Schacter 1985; Reber 1989; Schacter 1987). Implicit learning characterizes the fact that information can be permanently recorded without attentional control and without insight into stimulus features or contingencies. Thus, it refers to the preattentive retention of stimuli, which indicates that individuals adapt to structural characteristics of the stimulus environment without being asked to do so and without having insight into stimulus features and contingencies (Berry and Broadbent 1987). Explicit learning, on the other hand, refers to reasoned and strategically guided encoding and recognition of an event, which demands active attentional control (see Kihlstrom 1990; Reber 1989). Explicit memory has been interpreted as the "conscious recollection of a previous episode that can be

operationalized by free recall, cued recall and recognition tasks", wherein the participant is asked to "deliberately remember some aspect of the experience" (Kihlstrom *et al.* 1992a).

Another pair of concepts, 'automatic' versus 'controlled' processing, is well established in cognitive psychology (see Neisser *et al.* 1981). Automatic processing is characterized by a reflex-like and rapid course with constant time requirements, even during complex tasks. It is characterized by the lack of active control or attention (it is 'preattentive'). As opposed to controlled processing, which implies a direct and intentionally guided control, this is more time-consuming and requires an active allocation of attentional resources. Socio-cognitive research has adapted this concept of processing to the field of emotional evaluation processes. For example, social impression formation is believed to depend on an implicit learning process (e.g., Bargh 1984; Niedenthal and Kitayama 1994).

Since the many problems regarding methodology and measurement of perception without awareness have not been resolved (Bornstein 1992), the concepts 'awareness' and 'consciousness' must be treated as relatively general descriptions of the research program rather than as constructs with an established validity. Although some of the definitory attempts still contain a problematic mentalistic notion of consciousness, they provide clues for an operational definition of preattentive perception of emotional stimulus events.

Clearly, issues of methodology, operational definition and measurement are *central* to research on preattentive phenomena. Current definitions, at least in the domain of cognitive theories or in psychophysics, have been dominated by information-processing and signal detection approaches. It is generally accepted that 'unconscious' influences in perception and memory are best documented by the relative sensitivity of direct as compared to indirect measures of stimulus perception (see Merikle and Reingold 1991, 1992). When the threshold is defined as the point at which the observer is unable to detect the presence of a stimulus, then two types of thresholds may be identified: the subjective and the objective. *Subjective* thresholds are defined as the "detection level where subjects claim not to be able to discriminate perceptual information at better than chance level" (Cheesman and Merikle 1985:333). The *objective* threshold level denotes "the level of detectability where perceptual information is actually discriminated at chance level" (1985: 333). The subjective threshold setting may be based on the introspective report of the participant (cf. Dixon 1971, 1981) or on discriminative behavior (Holender 1986). Thus an appropriate operational definition could rely on both objective parameters of stimulus identification and aspects of stimulus detection.

The next issue is that of selecting an operational definition. At least two different measures of perception must be obtained to demonstrate dissociation

phenomena (Merikle and Reingold 1992): a first measure demonstrates that information is unavailable to 'consciousness', whereas a second measure shows that information is capable of influencing other processes. Experiments addressing preattentive perception require the participants to make judgments about a preceding encounter with a target event. Thus, stimulus identification is typically viewed in a decision-making process entailing several different components. For example, Mandler (1980) suggested that separate processes of familiarity and retrieval occur conjointly during the recognition of an event. When the participant is asked to make a decision on the prior occurrence of a stimulus event, there are separate and additive processes of familiarity value and a slower search process. The general utility of such a dual process model has been demonstrated by a variety of studies on recall and recognition performance.

Measures of self-report have not been generally accepted as reliable indicators of experience in the field of research on unconscious perception. It is unclear what criteria participants use when introspectively reporting on experience. Reports may be biased by experimental instructions and/or by notions of the person concerning the value of particular types of experiences (Merikle 1984). Introspection may thus represent the person's theory of how perceptual experience guides behavior rather than true perceptual experiences. Thus, an experimental manipulation may be less valid if based on subjective data with poor reliability.

Typical behavioral measures to assess perceptual awareness included forced-choice, present-absent decision tasks (e.g., Marcel 1983), and forced-choice discriminations among a small, known set of stimulus alternatives (e.g., Cheesman and Merikle 1984). However, before dissociations between two indices of perception can be interpreted as evidence for perception without awareness, an assumption of exhaustiveness must be made: "Whenever null sensitivity is equated with null awareness, the implication is that the selected measure provides an exhaustive index of all relevant conscious experience" (Merikle and Reingold 1992: 62). If, however, the selected measure is sensitive to 'conscious' as well as 'unconscious' perceptual processes, establishing null sensitivity can eliminate evidence for 'unconscious' perception. Therefore, an exhaustive measure *exclusively* indexing 'conscious' perceptual experiences is needed.

Previous studies have shown that recognition performance (multiple-choice) may exceed chance, even when a masked stimulus is *not* detected (Merikle and Reingold 1992). Although participants indicated 'no' in a detection task, sufficient information was available in part of the trials to correctly identify the stimulus. This recognition in the absence of stimulus detection was interpreted as evidence for unconscious perception. We have concluded from this discussion that cases in which the choice stimuli are correctly guessed, but at the same time

are not experienced as present, should *not* be treated as instances of perceptual awareness (Peper 1997; Peper and Karcher 2000).

An adequate test of stimulus awareness is thus needed to obtain some kind of convergent validation that an external stimulus has been registered. Such a test may involve the concurrent measurement of stimulus *detection* and stimulus *identification*. This can be accomplished, for example, by means of a stimulus presentation phase followed by a stimulus detection and identification task (cf., Merikle and Reingold 1992). For example, identification of an external stimulus event could be assessed by a forced choice two-alternative identification task (cf., Wong *et al.* 1994). Thus, for the definition of awareness of an external emotional stimulus event, a *necessary* condition is the correct identification of emotional valence (or of any other salient dimension). A *sufficient* condition is that the presence of the emotional stimulus event itself has also been experienced. Therefore, perceptual awareness of stimulus valence may be operationalized by two converging response variables: (1) a *perceptual* measure (correct valence identification), and (2) an *attentional* measure (monitoring stimulus input).

3. Applications of awareness definitions

Several psychophysiological studies investigated whether conditioned emotional stimuli, which could not be 'consciously' recognized, may yet elicit conditioned autonomic responses (e.g., Öhman 1979; Öhman, Dimberg and Esteves 1988). However, the problem of defining a condition of presenting emotional stimuli outside of the subject's awareness has not always been adequately addressed. In some accounts, the manipulation of awareness is not checked at all, precluding any inferences about this factor. Most definitions of awareness are based on introspective data (post hoc verbal report) as an index of declarative memory acquired during the experiment. For example, Bechara *et al.* (1995) contrasted a declarative memory score with non-declarative associative learning performance. As argued above, a definition based on introspective report is not a satisfying criterion of stimulus awareness. Another critical aspect of such an approach is the ex-post-facto study design. Öhman (1986) and Esteves *et al.* (1994) assessed identification and confidence judgments of masked visual stimuli, but did not use this procedure for individual threshold settings. Moreover, the use of a single, fixed condition of presentation for all participants may result in considerable error variance, which is then added to an experiment (Wong *et al.* 1994).

In a half-field study of preattentive perception, we postulated that convergent evidence of stimulus discrimination *and* input monitoring data (presence/absence monitoring task) is necessary to ascertain full awareness of a stimulus (cf. Bornstein and Pittman 1992). These obtained response characteristics were

applied in the context of a visual half-field study of preattentive perception and learning of emotionally valenced facial expressions (see Peper 1997; Peper and Karcher 2000).[2] The resulting response functions were characterized by an initial level at chance which continuously improves until an optimal decoding is reached. Identification performance and input monitoring showed similar functions. The conjunctive function of both measures represented our indicator of awareness. This hovers around chance at stimulus onset asynchronies (SOAs) of 10–20 ms and increases up to an individual maximum at SOAs greater than 20 ms. Thus, the preattentive-attentive dimension is not dichotomous, but rather follows a psychophysical function with continuous response characteristics (Figure 1; see also Figure 2).[3] In summary, attempts to define stimulus awareness show that the development of an adequate procedure to assess is not trivial. The emotional experience of an external object is a multistage process with recognition and attentional mechanisms involved.

4. Awareness of emotional objects: Neuropsychological hypotheses

Early attempts to conceptualize the relation of 'conscious' and 'unconscious' emotions have been hampered by both inadequate preliminary neuroscientific knowledge and problematic psychological constructs or operationalizations (Peper and Markowitsch 2000). Recent neurobiological research has considerably expanded knowledge of the structure and function of systems subserving emotions (e.g., LeDoux 1991, 1996; Rolls 1999). The relationship between emotional experience and brain function has been the focus of studies with healthy and brain damaged human subjects (see Davidson 1993; McGlynn and Schachter 1989).

Neuropsychological models suggesting dissociations of 'conscious' reflexive from 'non-conscious' automatic evaluations have increasingly influenced socio-cognitive theorizing. However, given the complexity of systems subserving, for example, non-declarative functions, many attempts to implement neuropsychological evidence into cognitive models of emotions are premature. General suggestions such as the idea that conscious experience could be associated with the linguistic capabilities of the left hemisphere are not satisfying. In the following sections, several more specific hypotheses are briefly summarized.

4.1 *The amygdala-hippocampus dissociation hypothesis*

Early in the sensory decoding process, rudimentary and adaptively significant evaluation procedures may escape 'conscious' control and precede a detailed cognitive analysis. There is converging evidence from animal experimentation

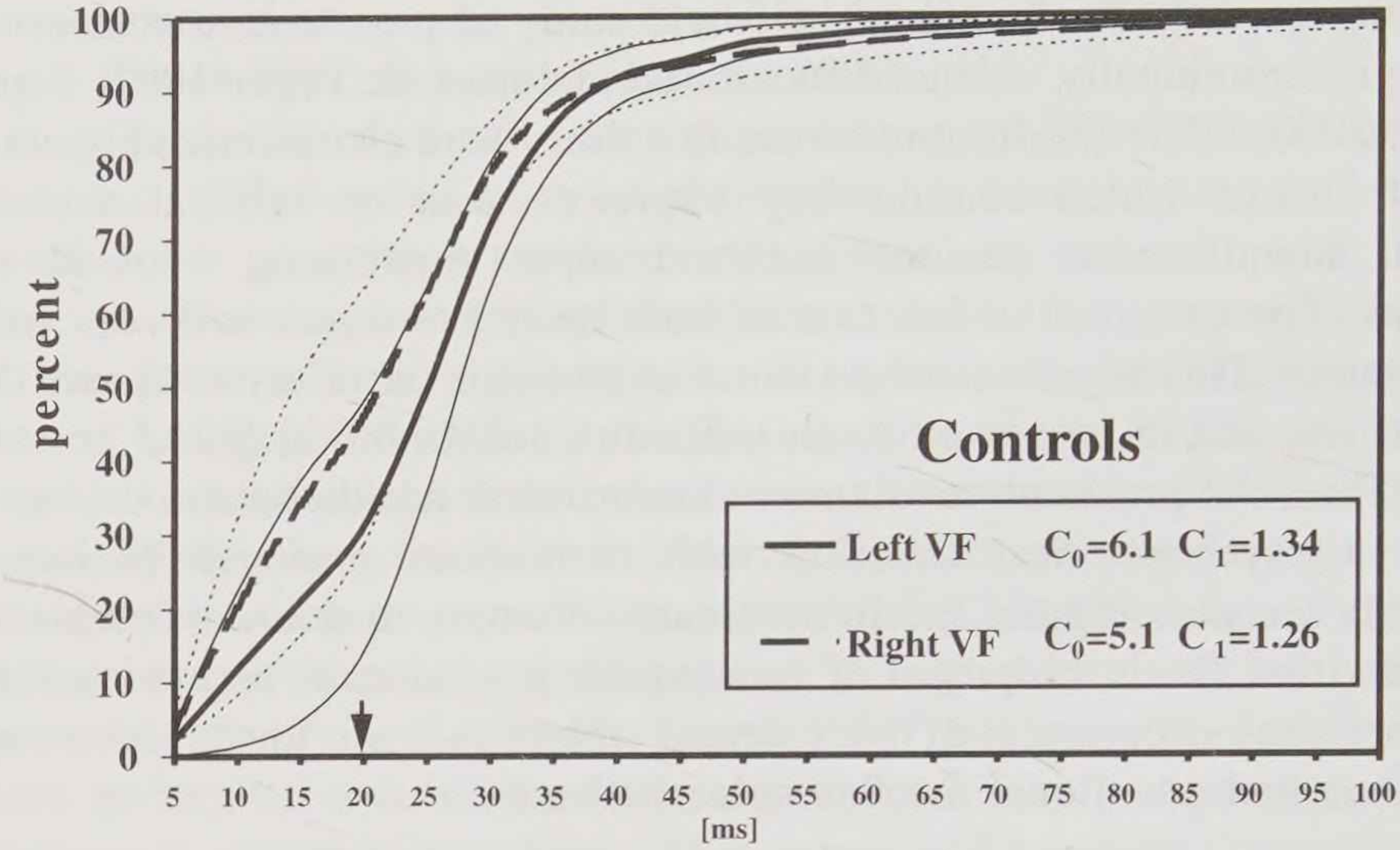


Figure 1.
The relationship of stimulus duration/SOA and identification of emotional valence in negative facial expressions is shown for the left visual field (VF)/right hemisphere and right VF/left hemisphere. The selected measure of perceptual awareness (percent identification corrected for insufficient monitoring, see text[2,3]) followed a sigmoidal curve which can be approximated by a logistic equation ($Y = 100/(1 + 10^{((C_1 - X) * C_0)})$). The parameters C_1 (log[SOA] of 50 percent identification) and C_0 (hillslope factor) were estimated by nonlinear regression (SAS Proc NLIN) for each of the 13 subjects assuming a bottom plateau of 0 and a top plateau of 100 percent ($R^2 > 0.9$). The curves show the predicted values for left or right VF presentations of four negative facial expressions averaged across subjects (95% confidence intervals of the means are also given). An analysis of variance showed a certain perceptual advantage for right as compared to left VF presentations (C_1: $F(1,12) = 3.7$; $p = 0.07$) as well as a more shallow slope factor (C_0: $F(1,12) = 10.7$; $p = 0.007$); no salient interaction with emotion type (positive vs. negative) emerged. This approach was used to obtain similar conditions for all presentation modes and subjects in a subsequent aversive conditioning experiment (Peper et al. 2000b). The selected SOAs (arrow) yielded a suboptimal identification performance of approx. 25–50 percent.

that the conditioning of aversive stimuli is controlled by the amygdala (e.g., Davis 1997; LeDoux 1996). Whereas aversive conditioning of autonomic responses is expected to be associated with amygdala function, the hippocampus has been assigned a role in relational and complex conditional learning both in animals (cf. Cohen and Eichenbaum 1993; LeDoux 1996; O'Keefe and Nadel 1978) and brain damaged humans (e.g., Daum *et al.* 1991). The role of the hippocampus in the declarative memory system is well documented (see, e.g., Squire 1993; Zola-Morgan *et al.* 1991). Neuropsychological case studies have suggested that declarative and nondeclarative aspects of emotional learning and memory could be dissociable functions of the amygdala and the hippocampus.

Bechara *et al.* (1995) reported that in a patient with a bilateral lesions to the amygdala, conditioning of autonomic responses but not introspective recall of declarative facts of the learning procedure was affected; the opposite was found in a patient with bilateral lesions of the hippocampus.

Functional brain imaging studies with healthy humans have recently investigated changes of brain activation related to the stimulation with sub-optimally presented fearful or happy facial expressions. Results indicate that the amygdala is relatively sensitive to optimally as well as suboptimally presented negative stimuli with a high emotional intensity. Whalen *et al.* (1998) measured 'nonconscious' processing by 33 ms presentations of an emotional target face masked by a neutral face. 8 out of 10 subjects verbally reported that they had not seen the facial expressions. Using functional magnetic resonance imaging methods, a significant signal increase was observed in the amygdala when a fearful was compared to a happy condition. However, as discussed above, verbal report is not a sufficient criterion of 'conscious' processing. Moreover, since no true experimental design was used, the causal effects of stimulus awareness remain uncertain. Nevertheless, these results suggest that amygdala activation persists when negative facial expressions are suboptimally presented. Using positron emission tomography, Morris, Öhman and Dolan (1998) recently observed hemisphere differences in the processing of masked and unmasked facial expressions. Masked angry faces previously associated with a 1 s burst of white noise were associated with an activation of the right but not the left amygdala. In contrast, an activation of the left amygdala appeared to be related to an identifiable presentation.

This amygdala activation may explain previous psychophysiological findings, which demonstrated that masked threat-related conditioned stimuli (CS's) such as negative facial expressions may show greater autonomic (electrodermal) resistance to extinction effects than positively valenced CS. More salient autonomic responses have been observed for negative expressions presented to the left visual field (LVF), that is, to the right hemisphere (Johnsen and Hugdahl 1991, 1993). Since it remained unclear whether subjects had partially identified the CSs, we used the previously mentioned identification procedure for a control of stimulus awareness in the context of an autonomic conditioning procedure (Peper 1997; Peper and Karcher, 2000). A right hemispheric advantage was observed for negative CS+ presented preattentively during extinction (Figure 2). The observation that conditioned electrodermal responses are elicited most effectively when CS are presented preattentively, i.e., without the subject's awareness of emotional stimulus features (e.g., Esteves *et al.* 1994; Öhman 1986; Wong *et al.* 1994) could be a consequence of an activation of the right amygdala.

If the amygdala is essential for the perception and conditioning of intense emotional stimuli, lateralized medial temporal lesions should have an effect on

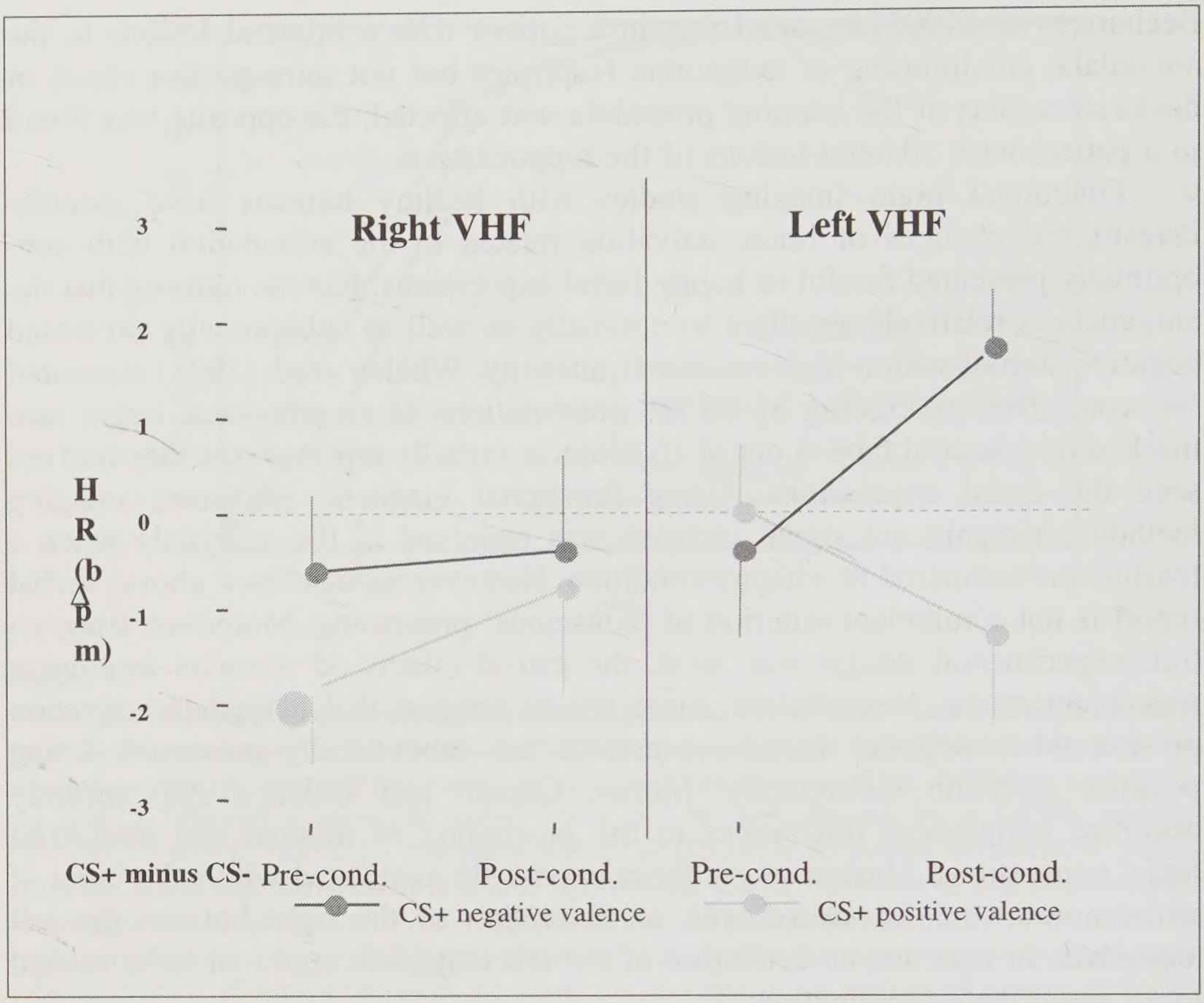


Figure 2.

Previous research on hemispheric asymmetries in aversive learning suggests that a right hemispheric resistance to extinction effect for skin conductance reactions can be expected when negative conditioned stimuli (CS+) are presented preattentively to the left visual field. In the context of a differential conditioning paradigm, Peper and Karcher (2000) examined the effects of stimulus awareness, visual half-field and emotional discriminanda on indicators of autonomic conditioning (bilateral skin conductance reactions, heart rate). N=41 healthy volunteers were investigated in a repeated measures design. Conditioned emotional facial expressions were presented with or without awareness in separate sessions during extinction. SOAs were adapted for each subject and condition prior to conditioning so that identification performance was at chance level (see above description of procedure). Results indicate that a differential effect for preattentive negative facial expressions presented to the LVF could be demonstrated (see accelerative heart rate reactions for in the post-conditioning phase). In contrast, autonomic responding to the CS+ was attenuated under full awareness of the CS during the extinction phase. Thus, in healthy persons, the magnitude of the autonomic conditioning effect reaches a maximum at SOAs of about 10–30 ms. With increasing identification of the CS, the effect declines and appears to attenuate even below the level of the CS–.

emotional decoding and psychophysiological responding during aversive learning. We therefore applied the same paradigm to 14 patients with selective left-temporal amygdalohippocampectomy (AHE), 12 with right AHE, and 13 matched controls. Awareness (identification) of the CSs was examined prior to the conditioning experiment and adapted for each patient and stimulus condition (Figure 3).

When identification of the conditioned negative facial stimulus was set to chance level, AHE patients failed to show autonomic resistance to extinction to preattentive presentations of emotional stimuli in the LVF. This deficit, however, was most likely due to the use of paired stimulus presentation during conditioning. This finding complements previous reports suggesting a particular role of the amygdala in the acquisition and reproduction of autonomic reactions to fearful stimuli in humans.

4.2 *The temporal-frontal cortex dissociation hypothesis*

Automatic defense behavior mediated by the amygdala is only a part of the array of complex behavioral responses to fearful stimuli. Instrumental actions and avoidance behavior are controlled by an interacting system of amygdala, striatum and the prefrontal cortex (see Everitt and Robbins 1992; Rolls 1999). The basolateral amygdala and the hippocampus massively project to the ventral striatum, transmitting information necessary for controlling active avoidance behavior. Frontal brain structures and their striatal-limbic connections are relevant for the evaluation of the reinforcing appetitive or aversive qualities of a stimulus event (e.g., Rolls 1975, 1999; Everitt and Robbins 1992).

In humans, the frontal cortex has been regarded as an important output station for the preparation of flexible behavioral and autonomic responses to environmental and internal needs. Frontal brain regions appear to be involved in the extraction and conceptualization of emotional information. In particular, a right frontal avoidance system appears to provide networks for evaluating the aversiveness dimension of exteroceptive emotional stimuli and regulating withdrawal behavior (Davidson *et al.* 1990; Davidson and Hugdahl 1995). Frontal lesions may affect the decoding of secondary emotional dimensions such as valence even though perceptual functioning remains unimpaired (Peper and Irle 1997b, c). Resections in the region of the ventral frontal cortex are associated with a change towards more negative mood descriptions (e.g., Irle *et al.* 1994). Ventromedial lesions of the frontal cortex also interfere with autonomic (electrodermal) activity elicited by emotional stimuli (Peper *et al.*, 2000a; Tranel and Damasio 1994).

Subjective experience and emotional feelings are believed to be mediated by working memory capacities of the frontal cortex. In particular, a working

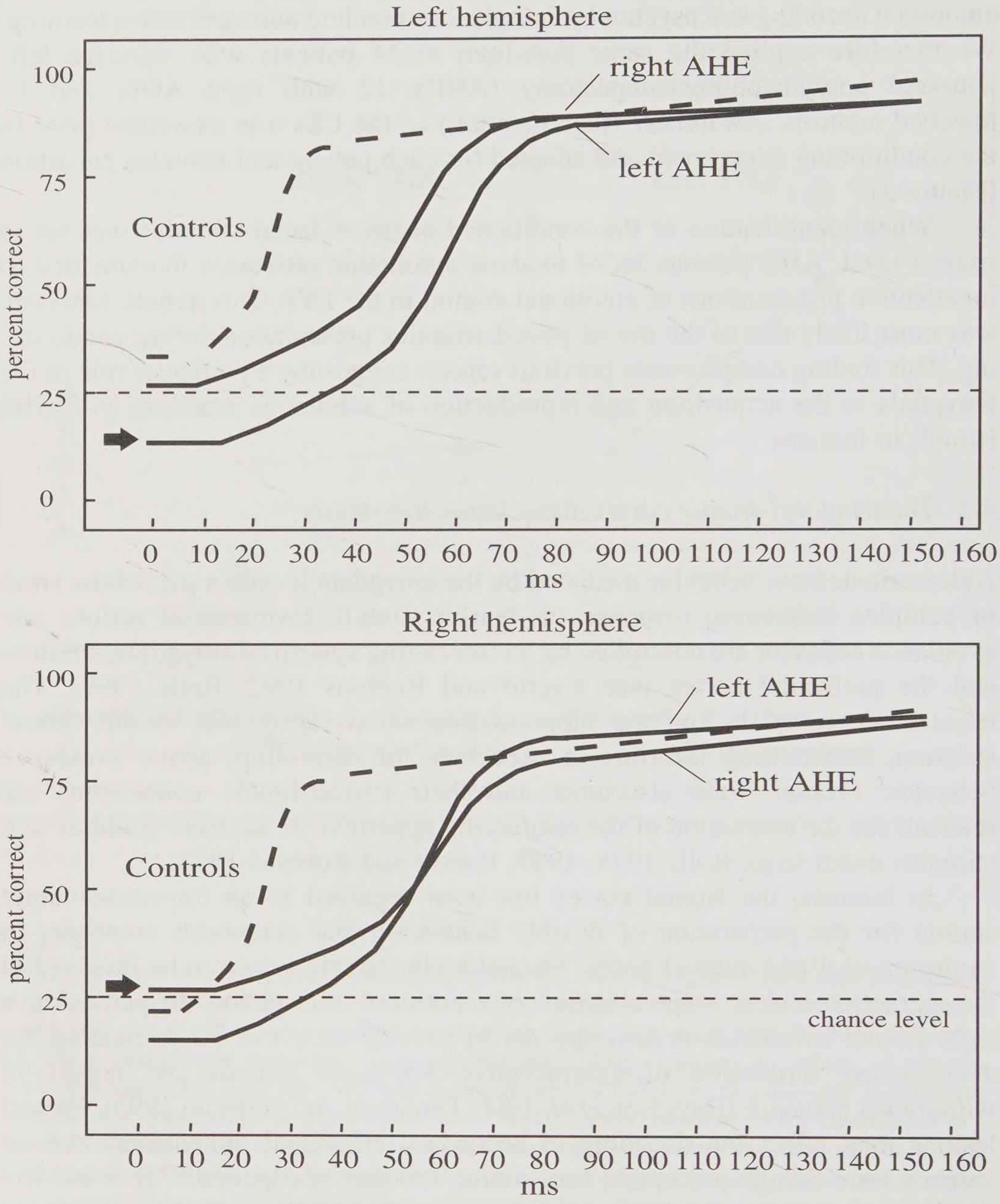


Figure 3. Approximated functions for the identification of negative valence in different facial expressions in patients with left or right unilateral amygdalohippocampectomy and controls. Arrows indicate contralesional presentation. Identification of negative facial emotions was attenuated in both visual fields as compared to controls.

memory executive in the prefrontal cortex is assumed to keep track of short-term sensory buffers, to retrieve relevant information from long-term memory and to register the direct or indirect effects of autonomic arousal (cf. LeDoux 1996). Thus, the temporal lobe networks and the amygdala appear to be involved in the recognition of emotional and social objects as well as in the automatic responding to these stimuli. In contrast, the frontal lobe provides mechanisms for the integration and recall of emotional information in working memory. These frontal capacities to monitor ongoing intero- or exteroceptive events and to relate this information to previously acquired knowledge, constitute an important basis of appropriate social decision making and action.

4.3 *The dorsal–ventral visual stream hypothesis*

From animal studies as well as clinical data, it has been concluded that the ventral stream and the temporal cortex are involved in the decoding of complex stimulus features and emotional significance of objects, whereas the parietal cortex is related to place information associated with visual guidance and integration of action (e.g., Goodale and Milner 1992). If so, sensomotor information of the dorsal stream may influence motor response parameters without necessarily inducing a conscious experience.

A dissociation of subjective awareness of an object on one hand and behavioral parameters (reaction time, error rate) on the other can be induced when a stimulus is masked by metacontrast (that is, when it is embedded in a superimposed stimulus arrangement). This dissociation has been attributed to the differential responding of the ventral and dorsal visual streams (e.g., Klotz and Neumann 1999). Hence, motor or autonomic measures of response readiness could be related to early dorsal output from the visual system, which is relatively independent of the ventral stream of object recognition.

5. Neuropsychological case histories

Emotional awareness may also be interpreted as an intrinsic, qualitative experience. Interoception of autonomic processes, reafferent information from the periphery, perception of anticipatory arousal, etc., may contribute to subjective experience of an emotion. The following neuropsychological case histories of patients with relatively similar lesion localizations in the right medial temporal lobe are included here to illustrate that brain lesions have a considerable impact on intrinsic emotional experience, while leaving other components of the emotional process relatively intact.

Case 1. A 29-year old woman was admitted to a neurosurgery department with a small astrocytoma of the right medial temporal lobe. Despite stereotaxic irradiation, tumor growth and edema occurred and were observed in a 3-year follow up CT (cf. Figure 4). A stereotaxic excision indicated an anaplastic astrocytoma. The tumor was microsurgically resected as far as the temporal regions were involved. However, a small portion of the tumor was already extending into the right basal ganglia. Upon admission, the patient was fully oriented and cooperative. Her neurostatus was regular except for an incomplete hemianopia on the left. Postoperative neuropsychological assessment showed no impairment of verbal cognitive functions nor of short and long term memory, but a significant impairment in speeded visuo-motor tests. Her performance in simple visual memory tasks was average (Benton Test, three errors), but attenuated visual memory functions were evident in complex tasks (Rey Figure, Recurring Figures Test, WMS-R-Visual Associates). In the Benton Face Recognition Test, she was average, but more time was needed to finish the task.

Emotional changes were observed by the patient immediately following tumor extirpation. She reported that she was not able to experience any emotion and described that she felt like a mindless robot or a computer. Significant others, even her child, appeared to have no emotional impact. Nevertheless, she was able to understand jokes and to laugh at them. This, however, would feel like a mechanical act to her as she did not realize the emotional quality of that feeling. Food intake was increased following the operation. The patient was seen by a psychiatrist, but a depressive disorder was not ascertained. Interestingly, she reported physiological arousal: something like apprehension in her body when she thought of problems like canceling the car registration or if she would be able to work after release from the hospital. This was experienced as a diffuse feeling like "something was going on in her stomach or in her nerves". Expressive emotional functions as observed in her facial or vocal behavior appeared to be intact. Perception of facial emotional stimuli as measured by discrimination and conceptualization tasks (Peper and Irle 1997bc) was also unimpaired. However, crossmodal integration of emotions in faces and voices was attenuated.

Loss of emotional qualities is a frequent concomitant of depressive disorders. Difficulties coping with the disease, cognitive deficits and psychosocial stressors could have contributed to a depressive disorder. However, psychiatric examination did not confirm such a disorder in this patient. The sudden onset of symptoms suggested that the right medial temporal lobe damage was responsible for the emotional changes. The lesion to the amygdala and basal ganglia appears to have induced this loss of emotional qualities without severely affecting social behavior or affect expression. This case suggests that awareness of intrinsic aversive or appetitive qualities can be attenuated without perception of emotional stimuli and, presumably, autonomic responding being affected.

Case 2. A 39-year old woman received a right-temporal amygdala-hippocampectomy for epileptic seizures. Previous to the operation, she had reported that the epileptic aura preceding the seizure was accompanied by a feeling within the bowels spreading upward in the body. During seizure, she was unaware of her environment; nevertheless, she was able to carry out relatively automated actions such as getting ready for bed, removing her

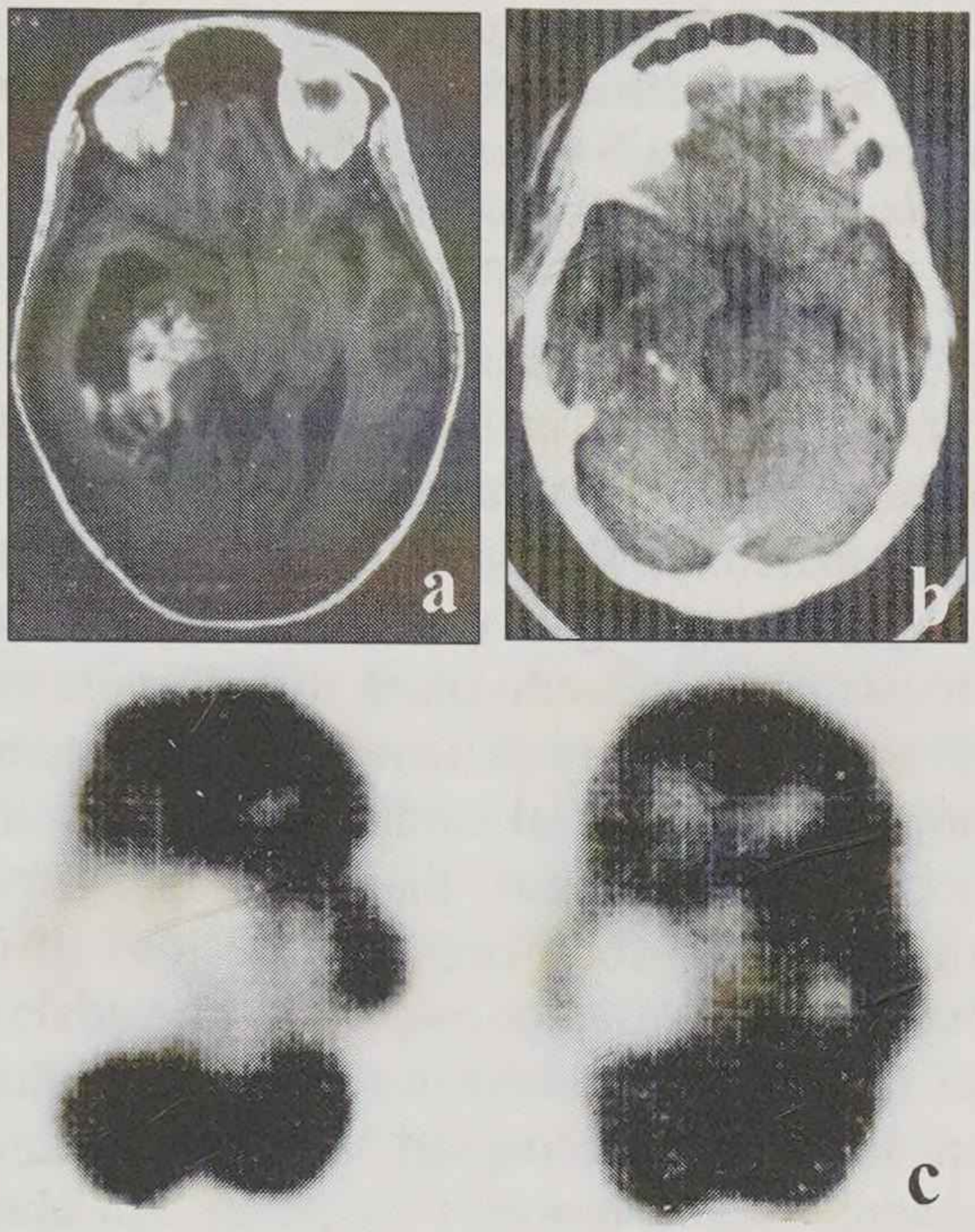


Figure 4. *Lesion of a woman reporting loss of emotional experiential qualities following a brain tumor resection*

a. preoperative MR-scan; b. postoperatively CT-scan showing a limited hypodense area extending posterior into the medial temporal lobe; the lesion involved the amygdala, the hippocampus, part of the basal ganglia, extending posterior up to the walls of the ventricle. c. A postoperative SPECT-scan showed reduced blood flow within the resection area in the medial temporal lobe.

make-up or even crossing a busy street. After the selective unilateral resection of the amygdala and the anterior part of the hippocampus, seizure activity ended. However, her emotional experience had lastingly changed. She felt upset by any distressing events in her environment: for example, after watching an animal being killed on TV, she was not able to get over this event for days. Thus she avoided watching arousing criminal films. Aversive experiences were repeatedly replayed in memory. Hypersexuality was reported. Moreover, a subjectively experienced increase in autonomic responding, in particular to distressing events and negative vocal utterances, was noted. A touch on some part of the body would immediately reverberate as a kind of 'visceral sensation' within the body. Parts of this self-report were confirmed by her spouse.

Behavioral tests of emotion recognition yielded normal verbal naming of emotional facial expressions. However, crossmodal decoding of the negative valence dimension was poor (T=39). In a semantic differential, the degree of negative valence in facial expressions was rated lower compared to controls (T=22), whereas positive faces were adequately appreciated (T=61). An elevated threshold duration for recognizing facial expressions (T>80)

was noted. In an autonomic conditioning paradigm, she showed low average psychophysiological responding to facial CS- or CS+ with poor responce to an aversive acoustic US (T=25). The aversiveness of the US was subjectively rated as not very intense (T=32). The acquisition of electrodermal (T=39) and cardiac reactions (T=35) was also poor. Her declarative memory of facts of the conditioning procedure was normal (T=53).

In contrast to the first case, awareness of emotional qualities was not attenuated as a consequence of the operation in the second subject. Rather, experiential qualities were distorted or amplified: minor aversive events reverberated exaggeratedly in memory. Apparently, this patient had problems in appropriately perceiving and conceptualizing negative stimulus events. Subjective and autonomic responding to aversive unconditioned stimuli were reduced. This corresponds to the observation that many patients with right temporal lesions show difficulties appreciating aversive social stimuli (cf. Peper and Irle 1997b).

The preceding case histories show that emotional experience of stimulus events may be lastingly altered following focal lesions of one cerebral region related to processing of negative emotions (right amygdala). These cases may seem reminiscent of the qualia problem in theoretical philosophy of mind (cf. Bieri 1981; Dennett 1988), resembling an 'absent emotional qualia problem' (case 1) and an 'inverted qualia problem' (case 2). An absent qualia problem occurs when two brain states are functionally identical, but one subject experiences qualia and the other does not. An inverted qualia problem arises when one subject experiences inverted (or distorted) qualia compared to another. These problems are based on the assumption of an identical physical realization of the mental function. A breakdown of the same part of the system should also have similar emotional effects.

According to the *reductionist* view, a qualitative emotional state is identical to a particular brain state. For these cases to be a problem with reductionism, it must be demonstrated that not only the brains were functionally identical, but also that each subject had precisely the same brain lesion. It is difficult to prove, however, that the brain states of the patients were absolutely identical. For example, the effect of different emotions could be explained by the invasive tumor growth into the right basal ganglia or ventral striatum in case 1. This might have been responsible for a disruption of transfering emotional information to working memory through the pathways of the fronto-striatal loop. Since it can not be shown that the brain states are absolutely identical, these examples provide no strong evidence against the physicalist perspective. A traditional criticism pertaining to *functional materialism* suggests that intrinsic qualitative states are not merely functional states. Neural states can have no qualitative character, that is, no particular 'way that it feels' for a subject to be in an emotional state. Since it can not be guaranteed that both patients had completely

identical brain damage, the divergent experiential qualities may not give rise to the sort of qualia problem problematic for functionalism.

Nevertheless, these cases suggest a fine-grained relationship between intrinsic experiential qualities and brain states, even within narrowly defined anatomical systems. Minor physical variations can produce major changes of experiential aspects. The cases exemplify the problem of studying intrinsic properties of emotional experience in affective neuroscience. The impossibility of obtaining a simplistic reduction of qualitative states to physical ones may have motivated the tendency of many emergentists or functionalists to neglect consciousness and emotion in favor of cognitive processes, or to identify emotion as a specific category of cognitive processes. Doing so would allow variations in qualitative emotional experience to be treated as epiphenomena, and hence as playing a nonessential role in the more central matter of cognition. This attitude would be a serious mistake. Therefore, even though the present case histories are not true examples of the 'qualia problem', they illustrate that distortions of emotional awareness *qua* experiential qualities are difficult to assess in the context of a reductionist or functionalist research program. The representation of the 'what is it like' dimension of emotional experience, that is, intrinsic qualitative states such as depression, anxiety, or pain, could not be identified with brain states without the loss of characteristic properties (see also Bieri 1981; Churchland 1986; Kurthen 1990; Panksepp 1991).

The research traditions of neuropsychology, psychophysiology and neurobiological research have partially divergent metatheoretical reference systems. Different metatheoretical presuppositions exist concerning the relationship between physical (neural) and mental (emotional) phenomena. It is beyond the scope of this chapter to discuss the different philosophical convictions related to the mind-body problem (for an overview of the philosophy of mind as a metatheory of empirical sciences, see, for example, Bieri 1981; Bunge 1980 1981; Churchland 1986; Fahrenberg 1992, 2000). Interestingly, the use of mentalist constructs such as 'emotion' is largely accepted, even in the realm of reductionist neuroscience (see SFN proceedings). It is well established that certain emotions can be indirectly inferred from observable spontaneous or acquired behavior or physiological changes in animals (cf., LeDoux 1986; Panksepp 1991). It also appears that the existence of *subjective* phenomena and its relevance for a theory of emotion is no longer questioned. However, the relationship between emotional experience and neural systems is still difficult to assess.

An important methodological implication of the reductionist stance is the necessity of a precise neuroanatomical-neurophysiological analysis of structures related to emotion. Furthermore, operational definitions of spontaneous or acquired emotional behavior have been implemented by animal models of several emotions (cf. Panksepp 1998). Although affective neuroscience has made

considerable progress, knowledge of neuropsychological mechanisms of emotional qualia in humans remains rudimentary. The functional implementation and architectures responsible for mediating emotional qualia must be further clarified. The process by which qualities of emotional experience can be identified as cerebral phenomena need to be more precisely described. The causal role of awareness can only be studied in the context of appropriate experimental approaches. A systematic approach requires a definition of internal emotional states and a method for assessing emotional qualities.

Conscious emotional experience is the result of the activity of only partially and temporarily coupled subsystems for stimulus evaluation and response generation. Internal and external stimulus events may differentially activate these modules or processing streams. Dissociations between introspective report, behavioral data and psychophysiological measures are the rule rather than the exception and can be observed in brain damaged as well as healthy persons. Emotional experience requires a dynamic state of the brain that includes many subsystems. The neuropsychological hypotheses reviewed support the view that there is not a *single* subsystem responsible for preattentive emotional responses and psychophysiological reactions. Rather, depending on situational constraints, a temporary and incomplete coupling of different systems responsible for the monitoring of internal and external events and for behavioral or physiological response generation represents the normal state of the organism.

Incomplete coupling can be experimentally provoked by degraded stimulus input. This may result, for example, in dissociations of stimulus awareness and autonomic responding. Discrimination or identification measures may provide appropriate measures of perceptual awareness of external emotional objects (Peper 1997). With incomplete coupling, 'silent affects' (as suggested by psychophysiological data or implicit behavioral reactions) or 'cold emotions' (such as the conceptual decoding of social stimuli without significant psychophysiological arousal) may occur.

Awareness of internal emotional states appears to be relatively imprecise and subject to interindividual differences. Situational factors, features of the selected paradigm and personality variables (self-schemata or appraisal habits) appear to determine how emotional events are evaluated. The subjective experience of a person having an emotion can not be easily reduced to aspects which are accessible by an external observer. Idiographic psychological methods are needed to quantitatively and qualitatively describe experiential states.

6. Conclusions

Awareness as a psychological construct is related to questions of construct

definition and operationalization. *Perceptual awareness* of emotions is a multicomponent process involving perceptual as well as attentional functions. Awareness of an external emotional stimulus, as conceptualized in the present chapter, involves at least two important components. The first is correct stimulus identification, subserved by the temporal association areas of the ventral visual stream. A second precondition of awareness is input monitoring which includes the ability to allocate attention (phasic alertness) to the event and to hold it through prefrontal working memory resources. Both aspects could be assessed by means of psychophysical methods (e.g., staircase threshold procedures). Awareness in the sense of intrinsic *experiential qualities* is a more complex process of integrating information from different sensory subsystems and sources of interoceptive information. It is also influenced by individual psychological factors.

Results of neuroscientific studies demonstrate considerable advances in unraveling the mechanisms of aversive and appetitive processing by the brain. However, there is no simple answer to questions pertaining to the neural substrate of emotional experience. Some still question whether emotional awareness can be explained by neuroscience at all (e.g., Bieri 1981; Churchland 1986; Rose 1999). Several hypotheses have been suggested to explain how dissociations of 'conscious' experience and automatic responding during emotional episodes could be realized by the brain. A preliminary model suggested that the experience of fear could be realized by the contributions of multiple cerebral systems (LeDoux 1996). According to this view, active sensory buffers storing information about an external stimulus are involved; activation of the amygdala for producing phasic emotional arousal is needed. Activation of the prefrontal working memory executive is required for interpreting the content of short-term memory stores in the context of activated long term memory. Working memory has been interpreted as the gateway to subjective emotional experience and indispensable for 'conscious' emotional feeling. Brainstem networks for arousal mediated by the amygdala as well as feedback information from the viscera and muscles are further important components for the subjective experience of fear. Dissociations of experience and automatic responding could be associated with the partial activation or decoupling of these subsystems under certain pathological or experimental conditions. Under normal circumstances, however, conscious and automatic processes interact in complex ways (see also Rossetti and Revonsuo, 2000).

It is suggested here that the awareness of emotions is only partially accounted for in the context of a strict physicalist affective neuroscience. Nonreductive materialism or methodological physicalism may permit more degrees of freedom. Following the 'decade of the brain', the American Psychological Association has made efforts to launch a 'decade of behavior'. Advancement of biological techniques (e.g. imaging cerebral activity) is needed; more importantly, advanced

behavioral technologies are required to assess subjective experiential qualities and awareness of emotionally-colored stimulus events.

Acknowledgments

Supported by the Deutsche Forschungsgemeinschaft (Pe 499/2,3). I thank N. Newton and J. Fahrenberg for their helpful comments on earlier versions of the manuscript. C. Bargou and M. Spitzer helped in the assessment of the first patient. S. Karcher assisted in collecting the emotional identification data. G. Reinshagen provided acces to patients of the Epilepsy Center freiberg-Kehl, Germany.

Notes

1. The disposition to experience and report psychological distress (elevated emotionality scores obtained by questionnaire) appears not to be strongly related to corresponding psychophysiological reactions. Nevertheless, emotionality appears to explain a great part of the variance found within self-descriptions of emotional behavior. Trait measures such as anxiety, depression and anger are significantly associated with emotionality (Myrtek, 1998).
2. An emotional facial expression was considered identified, if its emotional content was correctly selected in a two-alternative task *and* if the same stimulus was judged as being present. Each picture was presented at three different SOAs: 10 ms, 20 ms and 30 ms in the context of a backward masking procedure. Each SOA was tested for each condition by sixteen trials in four blocks, with SOAs randomly varied within each block, yielding a total of 192 trials. The stimulus sequence was as follows: a fixation cross was permanently visible; a small rectangle served as warning stimulus, which was foveally presented 500 ms prior to the target. Subsequently, one facial expression was presented, followed by the pattern mask. Identification performance was assessed by a two-alternative identification task: Subjects were first requested to press one of two vertically arranged keys symbolizing positive (+) or negative (–) valence, corresponding to the valence of the face seen before (valence identification task). Subsequently, after being prompted by a question mark, the participants confirmed if a stimulus had been registered (presence judgment or input monitoring task), by pressing one of the keys again. If the stimulus was considered absent, no key was pressed. The subjects were trained in this procedure prior to the experiment.
3. The analysis of the awareness measure yielded a SOA main effect ($F(1.5,48) = 51.58$; $p = .0001$). This was due to an identification performance clearly above chance at 30 ms, whereas identification was suboptimal for 10 and 20 ms. A VHF × Valence interaction ($F(1,31) = 4.30$; $p = .047$) was due to a left visual field (right hemisphere) advantage for identifying negative emotions. This was significant only for the identification measure, but disappeared for the presence judgment and the combined measure.

References

Averill, J. R., Clore, G. L., LeDoux, J. E., Panksepp, J., Watson, D., Clark, L. A., Ekman, P., and Davidson, R. J. 1994. "What influences the subjective experience of emotion?" In P. Ekman and R. J. Davidson *et al.* (Eds.), *The Nature of Emotion: Fundamental Questions. Series in Affective Science*: 377–407. New York: Oxford University Press.

Bargh, J. A. 1984. "Automatic and conscious processing of social information". In R. S. Wyer and T. K. Srull (eds.), *Handbook of Social Cognition*, Vol. 3: 1–43. Hillsdale, NJ: Erlbaum.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., and Damasio, A. R. 1995. "Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans". *Science* 269: 1115–1118.

Berry, D. C., and Broadbent, D. E. 1987. "The Combination of Explicit and Implicit Learning Processes in Task Control". *Psychological Research* 49: 7–15.

Bieri, P. 1981. *Analytische Philosophie des Geistes*. Königstein: Hain Verlag.

Bornstein, R. F., and Pittman, T. S. 1992. *Perception Without Awareness*. New York: Guilford.

Bowers, D., and Heilman, K. M. 1984. "Dissociation Between the Processing of Affective and Nonaffective Faces: A Case Study". *Journal of Clinical Neuropsychology* 6: 367–379.

Buck, R. 1993. "What is This Thing Called Subjective Experience? Reflections on the Neuropsychology of Qualia". *Neuropsychology* 7: 490–499.

Bunge, M. 1980. *The Mind-Body Problem: A Psychobiological Approach.* 1st Ed. Oxford: Pergamon Press.

Bunge, M. 1981. *Scientific Materialism*. Dordrecht, Holland: D. Reidel Publ.

Cheesman, J., and Merikle, P. M. 1984. "Priming With and Without Awareness". *Perception and Psychophysics* 36: 387–395.

Cheesman, J., and Merikle, P. M. 1985. "Word Recognition and Consciousness". In D. Besner, T. G. Waller and G. E. Mackinnon (eds.), *Reading research: Advances in theory and practice,* Vol. 5: 311–352). New York: Academic Press.

Cheesman, J., and Merikle, P. M. 1986. "Distinguishing Conscious from Unconscious Perceptual Processes". *Canadian Journal of Psychology* 40, 343–367.

Churchland, P. 1986. *Neurophilosophy*. Cambridge, Mass.: MIT Press.

Cohen N. J., and Eichenbaum, H. 1993. *Memory, amnesia, and the hippocampal system*. Cambridge: MIT Press.

Daum, I., Channon, S., Polkey, C. E., and Gray, J. A. 1991. "Classical Conditioning after Temporal Lobe Lesions in Man: Impairment in Conditional Discrimination". *Behavioral Neuroscience* 107: 748–756.

Davidson, R. J. and Sutton, S. K. 1995. "Affective Neuroscience: the Emergence of a Discipline". *Current Opinion in Neurobiology* 5: 217–224.

Davidson, R. J. 1993. "Parsing Affective Space: Perspectives from Neuropsychology and Psychophysiology". *Neuropsychology* 7: 464–475.

Davidson, R. J., and Hugdahl, K. (eds.) 1995. *Brain Asymmetry*. Cambridge, MA: MIT Press.

Davidson, R. J., Ekman, P., Saron, C. D., Senulis, J. A., and Friesen, W. V. 1990. "Approach — Withdrawal and Cerebral Asymmetry: Emotional expression and Brain Physiology I". *Journal of Personality and Social Psychology* 58: 330–341.

Davis, M. 1997. "Neurobiology of Fear Responses: the Role of the Amygdala". *Journal of Neuropsychology and Clinical Neuroscience* 9: 382–402.

Dennett, D. 1988. "Quining Qualia". In A. J. Marcel and E. Bisiach (eds.), *Consciousness in Contemporary Science:* 42–77. Oxford: Clarendon.

Dixon, N. F. 1971. *Subliminal Perception: The Nature of a Controversy*. London: McGraw-Hill.

Dixon, N. F. 1981. *Preconscious Processing*. London: Wiley.

Esteves, F., Parra, C., Dimberg, U., and Öhman, A. 1994. "Nonconscious Associative Learning: Pavlovian Conditioning of Skin Conductance Responses to Masked Fear-Relevant Facial Stimuli". *Psychophysiology* 31: 375–385.

Everitt, B. J., and Robbins T. W. 1992. "Amygdala — Ventral Striatal Interactions and Reward — Related Processes". J. P. Aggleton, Ed., *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction:* 401–430. New York: Wiley — Liss.

Eysenck, H. J. 1967. *The Biological Basis of Personality*. Springfield, Illinois: Thomas.

Fahrenberg, J. 1992. "Komplementarität in der Psychophysiologischen Forschung". In E. P. Fischer, H. S. Herzka, and K. H. Reich (eds.), *Widersprüchliche Wirklichkeit:* 43–77. München: Piper.

Fahrenberg, J., and Cheetham, M. 2000 "The Mind-Body Problem as Seen by Students of Different Disciplines". Manuscript submitted. *Journal of Consciousness Studies*, in press.

Fahrenberg, J., and Myrtek, M. (*eds.*) 1996. *Ambulatory Assessment.* Seattle, WA: Hogrefe and Huber Publishers.

Goodale, M. A., and Milner, A. D. 1992. "Separate Visual Pathways for Perception and Action". *Trends in Neurosciences* 15: 20–25.

Graf, P., and Schacter, D. L. 1985. "Implicit and Explicit Memory for New Associations in Normal and Amnesic Subjects". *Journal of Experimental Psychology: Learning, Memory and Cognition* 11: 501–518.

Holender, D. 1986. "Semantic Activation Without Conscious Identification in Dichotic Listening, Parafoveal Vision, and Visual Masking: A Survey and Appraisal". *Behavioral and Brain Sciences* 9: 1–66.

Irle, E., Peper, M., Wowra, B., and Kunze, S. 1994. "Mood Changes After Surgery for Tumors of the Cerebral Cortex". *Archives of Neurology* 51: 164–174.

Jacobson, M. 1995. *Foundations of Neuroscience*. New York: Plenum Press.

Johnsen, B. H., and Hugdahl, K. 1991. "Hemispheric Assymmetry in Conditioning to Facial Emotional Expressions". *Psychophysiology* 28: 154–162.

Johnsen, B. H., and Hugdahl, K. 1993. "Right Hemisphere Representation of Autonomic Conditioning to Facial Emotional Expressions". *Psychophysiology* 30: 274–278.

Kihlstrom, J. F. 1987. "The Cognitive Unconscious". *Science* 237: 1445–1452.

Kihlstrom, J. F. 1990. "The Psychological Unconscious". In L. A. Pervin (Ed.), *Handbook of Personality:* 445–464 . New York: Guilford.

Kihlstrom, J. F., Barnhardt, T. M., and Tataryn, D. J. 1992a. "Implicit Perception". In R. F. Bornstein and T. S. Pittman *et al.* (eds.), *Perception Without Awareness: Cognitive, Clinical, and Social Perspectives:* 17–54 . New York: Guilford Press.

Kihlstrom, J. F., Barnhardt, T. M., and Tataryn, D. J. 1992b. "The Psychological Unconscious". *American Psychologist* 47: 788–791.

Klotz, W., and Neumann, O. 1999. "Motor Activation Without Conscious Discrimination in Metacontrast Masking". *Journal of Experimental Psychology: Human Perception and Performance* 25: 976–992.

Kunst-Wilson, W. R., and Zajonc, R. B. 1980. "Affective Discrimination of Stimuli That Cannot be Recognized". *Science* 207: 557–558.

Kurthen, M. 1990. *Das Problem des Bewußtseins in der Kognitionswissenschaft.* Stuttgart: Enke.

LeDoux, J. E. 1986. "A Neurobiological View of the Psychology of Emotion". In J. E. LeDoux and W. Hirst (*eds.*), *Mind and Brain:* 355–358. New York: Cambridge University Press.

LeDoux, J. E. 1991. "Emotion and the Limbic System Concept". *Concepts in Neuroscience* 2: 169–199.

LeDoux, J. E. 1992. "Brain Mechanisms of Emotion and Emotional Learning". *Current Opinion in Neurobiology* 2: 191–197.

LeDoux, J. E. 1995. "Emotion: Clues From the Brain". *Annual Review of Psychology* 46: 209–235.

LeDoux, J. E. 1996. *The emotional brain.* New York: Simon and Schuster.

Mandler, G. 1980. "Recognizing: The Judgment of Previous Occurrence". *Psychological Review* 87: 252–271.

Marcel, A. J. 1983. "Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition". *Cognitive Psychology* 15: 197–237.

McGlynn, S. M. and Schacter, D. L. 1989. "Unawareness of Deficits in Neuropsychological Syndromes". *Journal of Clinical and Experimental Neuropsychology* 11: 143–205.

Merikle, P. M. 1982. "Unconscious Perception Revisited". *Perception and Psychophysics* 31: 298–301.

Merikle, P. M. 1984. "Toward a Definition of Awareness". *Bulletin of the Psychonomic Society* 22: 449–450.

Merikle, P. M., and Reingold, E. M. 1991. "Comparing Direct (Explicit) and Indirect (Implicit) Measures to Study Unconscious Memory". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17: 224–233.

Merikle, P. M., and Reingold, E. M. 1992. "Measuring Unconscious Perceptual Processes". In R. F. Bornstein and T. S. Pittman (eds.), *Perception without awareness:* 55–80. New York: Guilford Press.

Morris, J. S., Öhman, A., and Dolan, R. J. 1998. "Conscious and Unconscious Emotional Learning in the Human Amygdala". *Nature 393*: 467–470.

Myrtek, M. 1998. "Metaanalysen zur Psychophysiologischen Persönlichkeitsforschung". In F. Rösler (Ed.), *Ergebnisse und Anwendungen der Psychophysiologie, Enzyklopädie der Psychologie,* Vol. CI5: S. 285–344. Göttingen: Hogrefe.

Myrtek, M., Hilgenberg, B., Brügner, G., and Müller, W. 1997. "Influence of Sex, College Major, and Chronic Study Stress on Psychophysiological Reactivity and

Behavior: Results of Ambulatory Monitoring in Students". *Journal of Psychophysiology* 11: 124–137.

Neisser, U., Hirst, W., and Spelke, E. S. 1981. "Limited Capacity Theories and the Notion of Automaticity: Reply to Lucas and Bub". *Journal of Experimental Psychology, General* 110: 499–500.

Niedenthal, P. M., and Kitayama, S. 1994. *The Heart's Eye*. New York: Academic Press.

O'Keefe, J., and Nadel, L. 1978. *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.

Öhman, A. 1979. "Fear Relevance, Autonomic Conditioning and Phobias: A Model Laboratory Analyses of Emotion". In P.-O. Sjöden, S. Bates and W. S. Dockens III (eds.), *Trends in Behavior Therapy:* 107–134. New York: Academic Press.

Öhman, A. 1986. "Face the Beast and Fear the Face: Animal and Social Fears as Prototypes for Evolutionary Analyses of Emotion". *Psychophysiology* 23: 123–145.

Öhman, A. Dimberg, U., and Esteves, F. 1988. "Preattentive Activation of Aversive Emotions". In T. Archer and L. Nilsson (eds.), *Aversion, Avoidance and Anxiety. Perspectives on Aversively Motivated Behavior:* 169–194. Hillsdale, NJ: Erlbaum.

Panksepp, J. 1991. "Affective Neuroscience: A Conceptual Framework for the Neurobiological Study of Emotions". In K. T. Strongman (Ed.), *International Review of Studies of Emotion*. Vol. 1: 58–99. Chichester: Wiley.

Panksepp, J. 1998. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press.

Peper, M. 1997, 2000. *Aversive Learning: A Neuropsychological Perspective.* University of Freiburg: Habilitationsschrift, monograph in revision.

Peper, M., Häußler, K., Madlinger, A. and Zeutner, J. 2000a. *Acquisition and Extiction of Autonomic reactions in aversive learning: Effects os frontal brain lesions*. 4th Conference of The Federation of European Psychophysiology Societies, Amsterdam, May 24–27, 2000.

Peper, M. and Irle, E. 1997a. "Die Neuropsychologie der Emotionen". In H. J. Markowitsch (Ed.), *Klinische Neuropsychologie. Enzyklopädie der Psychologie, Volume C12:* 741–896. Göttingen: Hogrefe.

Peper, M. and Irle, E. 1997b. "The Decoding of Emotional Concepts in Patients with Focal Cerebral Lesions". *Brain and Cognition* 34: 360–387.

Peper, M., and Irle, E. 1997c. "Categorical and Dimensional Decoding of Emotional intonations in patients with focal brain lesions". *Brain and Language, 58,* 233–264.

Peper, M., and Karcher S. 2000. "Differential conditioning of emotional facial expressions: Effects of Hemispheric Asymmetries and CS-identification." *Psychophysiology,* accepted for publication.

Peper, M., and Markowitsch, H. J. "Pioneers of Affective Neuroscience and Early Concepts of the Emotional Brain". *The Journal of the History of the Neurosciences,* in press.

Peper, M., Karcher, S., Wohlfarth, R., Reinshagen, G., and LeDoux J. E. 2000b. "Emotional Associative Learning in Patients with Unilateral Lesions of the Amygdala and Hippocampus". *Biological Psychology*, in revision.

Plutchik, R. and Kellerman, H. (eds.) 1986. *Biological Foundations of Emotion*. Orlando: Academic Press.

Reber, A. S. 1989. "Implicit Learning and Tacit Knowledge". *Journal of Experimental Psychology 118*: 219–235.

Rolls, E. T. 1975. *The Brain and Reward.* Oxford: Pergamon.

Rolls, E. T. 1999. *The Brain and Emotion.* Oxford: University Press.

Rose, H. 1999. "Who Gets to Explain Consciousness? And Who Might in the Future"? In C. Taddei-Ferretti and C. Musio (eds.), *Neuronal Bases and Psychological Aspects of Consciousness. Series on Biophysics and Biocybernetics, Vol. 8.* London: World Scientific.

Rossetti, Y., and Revonsuo, A. (Eds.) 2000. *Beyond dissociation.* Philadelphia: John Benjamins.

Schacter, D. L. 1987. "Implicit Expressions of Memory in Organic Amnesia: Learning of New Facts and Associations". *Human Neurobiology* 6: 107–118.

Scherer, K. R. 1993. "Neuroscience Projections to Current Debates in Emotion Psychology". *Cognition and Emotion* 7: 1–41.

Scherer, K. R., and Peper, M., 2000. "Psychological Theories of Emotion and Neuropsychological Research". To appear in F. Boller and J. Grafman (eds.). *Handbook of Neuropsychology*. Vol. 5 (Emotional behavior and its disorders, ed. G. Gainotti). Amsterdam: Elsevier.

Squire, L. R., Knowlton, B., and Musen, G. 1993. "The Structure and Organization of Memory". *Annual Review of Psychology* 44: 453–495.

Tranel, D., and Damasio, H. 1994. "Neuroanatomical Correlates of Electrodermal Skin Conductance Responses". *Psychophysiology* 31: 427–438.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., and Jenike, M. A. 1998. "Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity Without Explicit Knowledge". *Journal of Neuroscience* 18: 411–418.

Wong, P. S., Shevrin, H., and Williams, W. J. 1994. "Conscious and Nonconscious Processes: An ERP Index of an Anticipatory Response in a Conditioning Paradigm Using Visually Masked Stimuli". *Psychophysiology* 3: 87–101.

Zola-Morgan, S., Squire, L. R., Alvarez-Royo, P. and Clower, R. P. 1991. "Independence of Memory Functions and Emotional Behavior: Separate Contributions of the Hippocampal Formation and the Amygdala". *Hippocampus* 1: 207–220.

Index

A

Abandonment fears, 209, 12
Acetylcholine, 40, 60–62, 64–66, 79
Action planning, 21
Action potential (in biological systems), 136, 139
Action readiness, 31, 38,
Action selection, 152, 161
Action, voluntary (*see also* Voluntary action), 62
Afferent responses, 5, 65, 83, 85
Affordances, xi, 21
Agency, x, xii, 207, 225, 238
Aggression, 49,63
Aims, xi, xiii–xvii, 6–8, 13, 15, 17, 18, 20, 166
Alcoholism, 63, 74, 219
Alzheimer's Disease, 66
Amygdala, x, xiii, xix, 13–15, 46, 62, 64, 66–71, 73–77, 83–87, 89, 125, 127, 153–4, 174–5, 177–9, 181–2, 186–91, 226, 251ff
Anger, xv, 9, 13, 17, 20, 58, 66, 69–71, 75, 113, 120, 123, 171, 191, 208, 212, 214, 216, 220, 222, 224, 228, 230, 232–3, 235, 240
Angiotensin (and thirst), 40
Anhedonia, 88
Animal vs. human models, 36, 37, 46, 51–53
Anterior cingulate, 5, 15, 29, 39, 57, 62, 68, 73–4, 76–7, 79, 135, 151, 159–60
Anticipatory interaction, 169
Anxiety, 23, 25, 40, 51, 62, 74, 85, 110, 119, 123, 183, 190–1, 209–10, 212
Appendage theory, 23, 25, 242
Appraisal theory, 31, 167, 185, 188, 220–1, 226–7
Arousal/affective circuits, 57, 59
Artificial intelligence, xviii, 22, 33, 176
Attachment, 42, 49–51, 206, 209, 210, 221
Attention, ix, x, xii, xiii, xv, xix, xx, 5–7, 9, 17–21, 33, 39, 45, 56, 61, 66, 71–77, 79, 89, 100, 133, 135, 149, 152, 158, 159, 165, 170, 189, 207, 209, 211, 216
Attractors (*see also* Basins of attraction), 94, 96, 98, 129, 194–7, 228, 239–40
Autism, 42, 74, 101
  and opiod excess, 42
Autocatalytic systems, 94ff
Automatic processing, 169–70
Autonomic motor system, 57–8
Autonomy, 177, 206–11, 216
Awareness definitions, 172, 245ff
Axons, 74

B

Baars, Bernard, 30, 38, 48, 55–6, 61, 133, 158
Basal forebrain, 23, 39, 61–2, 65–6, 76, 90
Basal ganglia, 62, 68, 72, 160, 180–3

Basins of attraction (*see also* Attractors), 4
Behaviorism, 8, 10, 26, 36, 221, 228, 235–6
Bertalanffy, Ludwig von, 235–6, 241
Bifurcations, 197
Blindsight, x
Block, Ned, 27, 48, 55, 69, 84, 164, 191
Brain stem, x, 64, 144, 151

C

C-fos immunohistochemistry, 30
Catalysis, 93, 96
Central motor processes, and consciousness, 35
Chalmers, David, 7–8, 23, 53, 189–90, 198, 200
Chaos, deterministic vs. entropic, 197
Chaos, xix, 22, 94, 179, 181, 191, 193–201, 219
Chaotic attractors, 129–30
Chirping, and tickle response, 44, 53
Cingulate (*see also* Anterior and posterior cingulate), xix, 5, 15, 29, 39, 57, 62–79, 135, 151–6, 159–60
Clark, Andy, xi, xxi, 3, 167, 177, 187
Classical conditioning, 14–15, 168–9, 187
Competence, 25, 39, 127, 173, 211, 220
Computational perspective on emotion, x, xi, xix
Computer modelling, xix, 133ff
Conditioned avoidance (and amygdala lesions), 69
Connectionism, 8, 133ff
Consummatory activity, xvii, 44, 95, 228, 236–7
Consummatory responses, 44, 95, 228, 236–7
Contagion phase, 209–10
Contagious emotion, 109–10
Continuum hypothesis, 123
Core consciousness, 56, 79
Corticothalamic loops, xix, 5, 133ff
Corticotrophic Releasing Factor (CRF), 40, 42
Cross-cultural differences, 53, 120, 130–1, 235–7
Curiosity, xi, xvii, 173, 213, 231, 233, 235

D

Damasio, Antonio, ix, x, xxi, 9, 23, 28, 37, 49, 74, 83, 95–6, 104, 178, 187, 191, 205, 217, 221
Darwin, Charles, 37, 49, 205, 223, 234–5, 241
Davidson, Donald, 10, 23, 25, 165, 174–8, 187
Dendrites, 66, 139
Developmental plasticity, 226
Diencephalon, 65, 151
Dissociation, 168, 171–87, 121–14, 245–51
Distributed parallel processing, xi
Dopamine, 50–53, 60–69, 86–88
Dorsal-ventral visual streams, 179, 257ff
Drive, xvii, 8–9, 66, 126, 144, 157, 228, 236
Dualism, 7, 10, 29, 104, 225, 233
Dynamic landscape, 228, 238
Dynamic equilibrium (*see also* Homeostasis), xviii, xix, 58
Dynamical systems, xii, xviii–xx, 3–5, 22, 91, 129, 136, 217–19
Dynamical uncertainty, 173

E

Ecological approach to emotion, xviii, 108, 113
Edelman, Gerald, x, xii, xxi, 87, 239, 241
EEGs, xxi, 8, 22–24, 60, 69, 84, 160, 194, 198, 201
Efferent activity, 5, 65, 83
Electrostatic charges, 8
Emergent property, 165, 192

Emotion recognition (*see also* Facial espression), 171, 182
Emotional object constancy, 213–18
Empathy, 125–7, 215
Enactive approach to emotion, 6, 12–16, 92, 99
Encodingist models, 166, 174
Endocrine systems,59, 227
Endorphins, 60, 63–4, 74
Entorhinal cortex, 73, 194–6
Entropic chaos, 197
Entropy, xvii, 92, 93
Epicritic vs. protopathic sensations, 30
Epigenetic landscape, 237–8
Epileptic seizures, 181
Epinephrine, 58–9, 122
Epiphenomenalism, 10ff, 104
Epistemological behaviorism, 8
Equalia (definition), 32, 46
Ethology, 54, 114, 235–7, 241–2
Evolution, xvii, 28–30, 32, 41, 46–54, 63, 93, 163, 166, 171, 174–6, 217
Existential loneliness, 212
Explanatory gap, 29, 179, 182, 185–193, 196, 199
Exploration, ix, xvii, 93, 110, 156, 172–3, 211, 232–3
Extended amygdaloid nuclei, 39
Extended reticular thalamic activating system (ERTAS), 38–9, 60, 76
Eysenck, H.J., 8, 23, 168, 188

F

Face recognition, 70, 139, 180
Face-receptive cells in amygdala (*see also* Face recognition), 70
Facial expression, as emotional indicator, 31, 120, 130, 156, 173, 175–77, 182, 190–1, 208, 224, 235, 253ff
Fear, xiii, xv, 13–14, 31, 40, 46–71, 84, 123–7, 131, 186
Felt sense, 17–21
First person science, 109ff
Fodor, Jerry, 166–7, 177
Free will, 109
Freeman, Walter, x, xvii, xxi, 3–4, 12, 23, 40, 49, 54, 89, 160, 183–4, 193–5, 198–201
Freud, Sigmund, 8, 23, 27, 35, 47, 49, 119, 206, 217, 220, 222, 234
Frontal eye fields, xix, 143–8, 153–9
Frontal lobe, 9, 25, 61–2, 66–8, 72, 79, 179, 205, 255ff
Frontal lobectomy, x

G

GABA, 60, 66
GABAergic neurons (*see also* GABA), 60, 66
GENESIS (GEneral NEtwork Simulation System), 136–9,
Gibson, John J., xiii, xxi, 9, 24
Globus, Gordon, x, xxi, 3–4, 12, 22–4, 199, 200
Glutamate, 42, 60, 63
Goldman, Alvin, 10, 24, 145, 159
Guilt, 172, 178, 216, 233

H

Habituation, 42, 66, 169
Hippocampus, 15, 46, 62, 66–77, 154, 174–5, 178, 181, 187–9, 251ff
Homeostasis (*see also* Dynamic equilibrium), xvii, 8, 58, 93, 100
Hunger, 21, 31, 35, 40, 48, 78, 84, 97, 170, 173–5
Husserl, Edmund, 19, 21, 24, 115
Hyperactivity, 45, 74
Hypersexuality, 182
Hypervigilance, 15–6
Hypothalamus, x, 39, 51, 59–74, 77, 87, 169, 226
Hysterical conditions, 9, 86, 160

I

Imagery, xi, xv, xxi, 6, 19, 21, 23, 72, 95–8, 100–4, 242
Implicit learning, 168–70, 187–90

Implicit memory, 168
Implicit perception, 169, 188
Inhibition, 14–15, 50, 58, 60, 77, 95–6, 100–1, 121, 130, 154, 157–8
Inhibitory interaction, 169
Instinct,223–242
Intentional object, xiv, 9, 15, 17–19, 21
distinguished from cause, 17–19;
from current perceptual obj., 17–19, 21
Intentionality, xiv, xvi–xx, 6, 13–17, 22, 165, 184, 198
Interactive flow, 162–3, 173
Interoceptive feedback, 59, 257ff
Intralaminar nuclei thamamus (ILN), 61–79
Introjection, 215
Invariances (among interactive webs), 164
Isomorphism, xiv

J

Jackson, Frank, xi, xxi, 11, 24, 30, 50, 217
James, William, 37, 59, 84–7, 100, 121–36, 159, 215, 229–34
James-Lange theory, 231–2
Jealousy, 219, 230–1
Jeannerod, Mark, 96, 105
Joy, 27, 37, 41–5, 53, 124–5, 208, 232–3

K

Kaszniak, Al, 38, 47, 51, 53
Kauffman, Stuart, xi, xvii, xxi, 3, 11–12, 24, 93–6, 105, 116–18
Kim, Jaegwon, xii, xxi7, 10, 24
Knowledge Argument, xi
Kohut, Heinz, 216

L

Language of thought hypothesis, 166
Laszlo, Ervin, 225, 242
Lateral intraparietal area, xix
Laughter, 43–5, 52–3, 124, 209
LeDoux, Joseph, 14, 24, 28, 36ff, 51, 57, 74, 124, 131, 252ff
Leutinizing Hormone-Releasing Hormone (LH-RH), and sexuality, 40
Libet, Benjamin, 109–11, 118
Limbic system, 124, 135, 151–3, 189–95, 205
Love, xiv, 18, 21, 49, 101, 113, 120–3, 200, 210, 213, 219, 231–3
LSD, 63
Luria, ix, x, xii, xxi, 24, 217

M

Mac Cormac, Earl, x, 12
Mack, Arien, x, xxi, 5
MacLean, P. D., 37–8, 51
Mahler, Margaret, 206–15, 222
Mammalian emotions, 27, 37, 41–6, 50–2
Mastery, 173, 210–13
Maturation, xx, 210, 229–40
McDougall, William, 234–5, 242
Medulla, adrenal, 59–61
Mental causation, 7, 24, 179, 192–6
Metaphorical imagery, 19–21
Microgenesis, 162, 168–70
Midbrain, 39, 48, 60–79, 144, 151
Mind/body interactionism, 35
Mind-body problem, 35, 184, 187–8, 201
Monod, Jacques, xi, xvii, xxi, 3, 11, 25, 116–18
Mood, 6–7, 64, 73–4, 102, 119, 167–8, 188, 210–11
Motivation vs. emotion, ix–xv, 3–6, 11–13, 22–25
Motivation, xix, 3–6, 11–13, 22–25, 43, 49, 55–6, 63–6, 77–9, 95–7, 133–5, 149–50
Multiple personality disorder, 214
Multiple realizability, xix, 11

N
Narcissism, 212ff
Narcissistic personality, 212
Natsoulas, xv, 4, 9, 25, 55, 87, 226, 242
Nielsen, Lis, 38, 47, 51
Nonequilibrium homeostatic states, 93
Nonlinear dynamics, 179, 191–3
Norepinephrine, 12, 40, 58–64, 83
Novelty, 71, 167, 170–73
Nuclei of reticular thalamus (nRT), 65–6, 76
Nucleus accumbens, 50, 62, 65–6, 69, 76, 88
Nurturance, xvii, 40ff

O
Object Relations Theory, xx, 205, 217–22
Object-relatedness of emotions (*see* Intentionality)
Obsessive Compulsive Disorder, 62, 74
Occipital lobe, x, 5
Olfactory bulb, 74, 183, 194–5
Operant conditioning, 66, 168
Operational definitions, 166, 184, 247ff
Opioids (endogenous), 40–3, 51–2
Overcausation, 12
Oxytocin, 40–3, 47, 50–3

P
Pain, 22, 31, 43, 57–8, 63–4, 67, 74, 78–9, 101, 122, 125–6, 169, 173–5, 183
Panic, 41, 78, 171, 232
Parallel vs. serial processing, xi, xix, 133
Parietal lobe, 67, 73
Periaqueductal gray (PAG), 39–42, 48, 63–74, 79
Phase transitions, 197
Phenomenology, xvii, 4, 16, 22–5, 135, 157, 224, 229, 233, 234, 241
Phineas Gage condition, 205
Phobias, 114, 119, 124–5, 189
Play, xvii, 20–1, 41, 43–5, 51–4, 59, 63–5, 68, 96–7, 100, 117, 172, 183, 186, 196, 215
Pleasure, xvii, 8, 23, 42, 48, 64, 95, 122, 125, 130, 232
Pons, 60–1
Posner, Michael, ix, x, xii, xxii, 5, 9, 25, 61, 66, 76, 85, 88, 100, 105, 151, 160
Posterior cingulate, 73, 76
Preattentive perception, 166, 170–3, 247ff
Preconscious imagery, xii
Preconscious processing, xii, 9, 17, 18, 102, 188
Prefrontal cortex, 72, 77, 145, 178
Preoptic areas, 42ff
Presaccadic neurons, 144ff
Priming effects, 168–9, 187
Process theory, 221
Prolactin, 40, 42
Propositional attitude theory, 31
Proprioceptive imagery, 21
Prosopagnosia, 70
Prozac, 10, 119
Psychodynamic models, xx, 206–21
Psychophysical identity, 10–11

R
Raphe nuclei, 60
Rapid eye movements (REM), 56, 60–2, 64–5
Rapprochement, 211–13, 215
Reafference, 195
Redundant causal systems, 12
Reflexive awareness, 100, 172–5
Representation (role in emotion of), xiv, xv, xix, 6, 19, 39–41, 96–8, 100–1, 134, 143, 161–4, 183, 188, 165–8, 174–5, 181–5, 213–20
Reticular formation (*see also* ERTAS), 60–3, 69, 87

Reticular activating system, 39
Reward vs. motivational circuits, 65–8
Rorschach test, 114

S

Saccades, — purposeful vs. unintentional, 143–5, 152ff
Schizophrenia, 56, 63, 66, 74, 77, 84–88
Searle, John, xii, xxii9, 25, 45, 53, 182, 198, 201
SEEKING system, 29, 44ff
Self-awareness, 91
Self-esteem, 13
Self-organization (*see also* Dynamical systems), x, xi. xiii, 92, 93, 96
Selfobjects, 216
Semantic priming research, 168
Separation distress, 40ff, 50–2
Separation-individuation, 205ff
Serotonin, 40, 60–4, 74, 83, 86–8, 96, 1234
Sex, xvii, 66, 189
Shame, 120, 127, 178, 216
Shunt mechanisms, xviii, 12
Significance Activation Mechanism (SAM), 150–158
Social bonding, xvii, 41
Solms, M., 31, 38, 40, 45, 46, 54
Spinal cord injury, 131
Strange attractors, 197
Stria terminalis, 42, 46
Subjective thresholds, 170
Substantia nigra, 60–2, 144
Suicide, 62, 64, 102
Superior colliculus, 144
Supervenience, 12, 40, 186, 190, 199
Symbiosis, 207ff
Synaptic weight values, 136

T

Target nodes (in frontal eye field sequential processing), 145–7
Temporal lesions, and aversive learning, 177, 182, 255ff
Temporal lobe, 68, 73, 83–4, 89, 179–81, 187, 169, 255ff
Thalamus, xii, 13, 15, 58, 60–76, 226
Thelen, Esther, xi, xxii, 3, 12, 25, 118, 175, 178, 225, 227, 238–9, 242
Transitional objects, 210
Trigger stimulus, — distinguished from intentional object, 20–1
Truth value, 163

U

U-shaped learning curves, 127
Unconscious processes, xii–xvii, xx, 9, 87, 95–9, 112, 118, 169, 171–2, 179–85, 187–9, 196–8, 200, 217–20, 226

V

Valence, 62–4, 69, 172, 177–8, 182, 191–2, 227, 233
Varela, Francisco, xiii, xxii, 3, 5, 25, 223, 242
Variable threshold control, 146
Vector spaces (in dynamical systems), 136, 139, 141–3
Vector subtraction, 141ff
Ventral striatum, 178, 183
Ventral tegmental area (VTA), 60–2
Vitality affects, 208
Voluntary action (*see also* Action, voluntary), 62
Vygotsky, Lev, 120, 131

W

Wakefulness (*see also* Arousal), 55–6, 60–2, 89
Watt, Douglas, x, xxii, 3, 26, 38–9, 54, 56, 57, 61, 62, 64, 66, 69, 71, 74, 76, 77, 79, 89
Working memory, 38, 72, 73, 83, 85, 96, 133, 145–58, 165, 178–9, 183–6

Z

Zoloft, 10

In the series ADVANCES IN CONSCIOUSNESS RESEARCH (AiCR) the following titles have been published thus far or are scheduled for publication:

1. GLOBUS, Gordon G.: *The Postmodern Brain*. 1995.
2. ELLIS, Ralph D.: *Questioning Consciousness. The interplay of imagery, cognition, and emotion in the human brain*. 1995.
3. JIBU, Mari and Kunio YASUE: *Quantum Brain Dynamics and Consciousness. An introduction*. 1995.
4. HARDCASTLE, Valerie Gray: *Locating Consciousness*. 1995.
5. STUBENBERG, Leopold: *Consciousness and Qualia*. 1998.
6. GENNARO, Rocco J.: *Consciousness and Self-Consciousness. A defense of the higher-order thought theory of consciousness*. 1996.
7. MAC CORMAC, Earl and Maxim I. STAMENOV (eds): *Fractals of Brain, Fractals of Mind. In search of a symmetry bond*. 1996.
8. GROSSENBACHER, Peter G. (ed.): *Finding Consciousness in the Brain. A neuro-cognitive approach*. n.y.p.
9. Ó NUALLÁIN, Seán, Paul MC KEVITT and Eoghan MAC AOGÁIN (eds): *Two Sciences of Mind. Readings in cognitive science and consciousness*. 1997.
10. NEWTON, Natika: *Foundations of Understanding*. 1996.
11. PYLKKÖ, Pauli: *The Aconceptual Mind. Heideggerian themes in holistic naturalism*. 1998.
12. STAMENOV, Maxim I. (ed.): *Language Structure, Discourse and the Access to Consciousness*. 1997.
13. VELMANS, Max (ed.): *Investigating Phenomenal Consciousness. Methodologies and Maps*. n.y.p.
14. SHEETS-JOHNSTONE, Maxine: *The Primacy of Movement*. 1999.
15. CHALLIS, Bradford H. and Boris M. VELICHKOVSKY (eds.): *Stratification in Cognition and Consciousness*. 1999.
16. ELLIS, Ralph D. and Natika NEWTON (eds.): *The Caldron of Consciousness. Motivation, affect and self-organization – An anthology*. 2000.
17. HUTTO, Daniel D.: *The Presence of Mind*. 1999.
18. PALMER, Gary B. and Debra J. OCCHI (eds.): *Languages of Sentiment. Cultural constructions of emotional substrates*. 1999.
19. DAUTENHAHN, Kerstin (ed.): *Human Cognition and Social Agent Technology*. 2000.
20. KUNZENDORF, Robert G. and Benjamin WALLACE (eds.): *Individual Differences in Conscious Experience*. 2000.
21. HUTTO, Daniel D.: *Beyond Physicalism*. 2000.
22. ROSSETTI, Yves and Antti REVONSUO (eds.): *Beyond Dissociation. Interaction between dissociated implicit and explicit processing*. n.y.p.
23. ZAHAVI, Dan (ed.): *Exploring the Self. Philosophical and psychopathological perspectives on self-experience*. 2000.
24. ROVEE-COLLIER, Carolyn, Harlene HAYNE and Michael COLOMBO: *The Development of Implicit and Explicit Memory*. n.y.p.
25. BACHMANN, Talis: *Microgenetic Approach to the Conscious Mind*. n.y.p.
26. Ó NUALLÁIN, Seán (ed.): *Spatial Cognition. Selected papers from Mind III, Annual Conference of the Cognitive Science Society of Ireland, 1998*. n.y.p.

27. McMILLAN, John and Grant R. GILLETT: *Consciousness and Intentionality.* n.y.p.
28. ZACHAR, Peter: *Psychological Concepts and Biological Psychiatry. A philosophical analysis.* n.y.p.
29. VAN LOOCKE, Philip (ed.): *The Physical Nature of Consciousness.* n.y.p.

These new studies by prominent neuroscientists, psychologists and philosophers work toward a coherent framework for understanding emotion and its contribution to the functioning of consciousness in general, as an aspect of self-organizing, embodied subjects. Distinguishing consciousness from unconscious information processing hinges on the role of motivating emotions in all conscious modalities, and how emotional brain processes interact with those traditionally associated with cognitive function. Computationally registering/processing sensory signals (e.g. in the occipital lobe or area V4) by itself does not result in perceptual *consciousness*, which requires subcortical structures such as amygdala, hypothalamus, and brain stem. This interdisciplinary anthology attempts to understand the complexity of emotional intentionality; *why* the role of motivation in self-organizing processes is crucial in distinguishing conscious from unconscious processes; how emotions account for 'agency'; and how an adequate approach to emotion–motivation can address the traditional mind–body problem through a holistic understanding of the conscious, behaving organism.

John Benjamins Publishing Company

ISBN 90 272 5136 3 (Eur) / 1 55619 196 0 (US)