The Emotional Brain Revisited

Edited by: Jacek Dębiec Michael Heller
Bartosz Brożek Joseph LeDoux

Copernicus Center PRESS

The Emotional Brain Revisited

The Emotional Brain Revisited

Edited by:

Jacek Dębiec
Michael Heller
Bartosz Brożek
Joseph LeDoux

Copernicus Center
PRESS

Editing:
Aeddan Shaw

Cover design:
Mariusz Banachowicz

Layout:
Mirosław Krzyszkowski

Typesetting:
MELES-DESIGN

Publication Supported by the John Templeton Foundation Grant "The Limits of Scientific Explanation"

ISBN 978-83-7886-042-6 (cloth)
ISBN 978-83-7886-072-3 (PDF)

Kraków 2014

Copernicus Center PRESS

Publisher: Copernicus Center Press Sp. z o.o.
pl. Szczepański 8, 31-011 Kraków
tel/fax (+48) 12 430 63 00
e-mail: marketing@ccpress.pl
www.ccpress.pl

Printing and binding by Bookpress.eu

Table of Contents

Part I

Part II

Part III

Jacek Dębiec
Michael Heller
Bartosz Brożek
Joseph LeDoux

Introduction

Until modern times, the emotions were entirely within the domain of the humanities. The first written attempts to conceptualize emotions in Western culture come from the philosophers of Ancient Greece. For Plato, the emotions (or passions) were the basic components of the soul along with desire and reason, with reason being in a natural position of ruling over the others. In contrast, Aristotle provided more nuanced characteristics of the emotions and acknowledged their role in shaping judgments and influencing reason. Mastering control over destructive emotions and encouraging ethical conduct in the pursuit of a happy life were at the core of the teachings of the Stoics. In subsequent centuries, regardless of whether they were explicitly addressed or merely implicit in other concepts and ideas, emotions have always been present throughout the history of Western thought. Although philosophical inquiry tackled almost every aspect of the emotions as experienced through introspection and observation, the main questions around which the philosophical pursuit was historically centered upon were their origins and nature, as well as the function of emotions and how they relate to other mental capacities. This approach, however, was closely aligned with the view of emotions as a specifically human achievement which was closely associated with the human capacity for consciousness and introspection.

Since the second half of the nineteenth century (marked symbolically by the publication of Charles Darwin's *The Expression of the Emotions in Man and Animals* in 1872 and William James' article *What is an emotion?* in 1884) emotions began to be an acknowledged and systematically studied subject for the biological sciences and psychology. The introduction of the evolutionary perspective and recognition that some emotion-related processes are preserved across animal species allowed the studying of these processes in animals. Animal studies were, and still are, crucial to progress in relating observable phenomena, such as emotional expressions and associated behaviors, to underlying, detailed brain mechanisms. The emergence and progress of neuroscience in the twentieth century provided grounds for, at least to some degree, the integration of traditional biological and psychological approaches to emotions. It took decades, however, for emotions to become a fully acknowledged topic of investigation within the neuroscientific community.

At the same time, the philosophical inquiry into emotions, which had been formed by centuries of investigations, continued independently of the empirical disciplines. In extreme instances, philosophers developed their concepts and theories of emotions by relying on philosophical or literary texts as the sole source of their thinking. In recent years, however, there has been increased interest among philosophers in the results and methods of neural science. On the other hand, many brain researchers acknowledge the need to understand their work and results in a broader cultural and philosophical context. This reciprocal curiosity and mutual interest has created fertile ground for encounters and dialogue between neuroscientists and philosophers.

The present collection of essays represents one such encounter. Initially meant to be the proceedings of the conference *The Emotional Brain: From the Humanities to Neuroscience and Back Again* which was organized by the Copernicus Center for Interdisciplinary Studies in Krakow, Poland in May 19-20, 2011, this volume has assumed a life of its own. Apart from the lectures delivered during the conference, this collection includes essays which have been directly inspired by the talks and discussions held during the Krakow meeting. A few of the contributions in this volume have previously been published in specialist journals and are re-printed here in order to reach a broader audience.

This collection of articles has been divided into three parts. The first part discusses emotion research in neurosciences. In the opening essay, Joseph E. LeDoux explores brain circuits and functions associated with survival and their relation to emotions. He proposes the survival circuit concept as grounds for understanding common human and animal emotion-related processes. The two articles that follow discuss research on brain circuits contributing to emotions in animal models and humans. Regina M. Sullivan and Margo S. Landers review rodent studies of attachment and discuss how disruptions in early infant attachment contribute to maladaptive emotional states in adult life. Paul Whalen and colleagues explore human research on the amygdala-medial prefrontal cortex circuitry, its role in emotional regulation, and explain how dysregulation in this circuitry may lead to pathological anxiety. In the next paper, Jacek Debiec discusses historical and methodological issues related to the formulation of the brain-based theory of emotions. This part of the volume is concluded by an article by Bram T. Heerebout and R. Hans Phaf which discusses the role of computational modeling in studying the affective regulation of attention.

The second part of the collection is comprised of two essays representing psychological studies of emotions. Nico H. Frijda explores the phenomena of impulsive action and conceptualizes emotions as motive states which interact with each other and contribute to eliciting and controlling impulsive actions. In the next essay, James A. Russell introduces and discusses the concepts of core affect and psychological construction as central ideas in explaining and understanding emotion.

The third and last part of this volume includes philosophical essays on emotions, emotion research in neurosciences and psychology, and their impact on philosophy and society. Łukasz Kurek analyzes the idea of emotion as a theory-binding concept and argues against attempts to eliminate the concept of emotion from philosophical and scientific discourse. Mateusz Hohol and Piotr Urbańczyk review past and current paradigms in neurocognitive sciences to focus on the embodied-embedded mind approach in the study of social cognition and emotion. In the next essay, Wojciech Załuski proposes a conceptual framework to permit us to understand the rationality of emotions. Bartosz Brożek analyzes and questions the concept of 'moral emotions' present in some trends of contemporary psychology and philosophy. Brożek argues that emotions play a fundamental role in regulating social interactions and culture and, as such, they cannot be reduced to mere factors regulating moral behavior. In the last contribution to this volume, Dominika Dudek discusses the concept of mental illness from the historical perspective and draws parallels between the progress in emotion research and advancements in conceptualizing mental illness in psychiatry and contemporary culture.

Part I

Joseph LeDoux
Center for Neural Science and Department of Psychology,
New York University, New York, NY 10003 USA
Emotional Brain Institute, New York University and Nathan Kline Institute,
Orangeburg, NY 10962 USA

Rethinking the Emotional Brain

I propose a reconceptualization of key phenomena important in the study of emotion—those phenomena that reflect functions and circuits related to survival, and that are shared by humans and other animals. The approach shifts the focus from questions about whether emotions that humans consciously feel are also present in other animals, and toward questions about the extent to which circuits and corresponding functions that are present in other animals (survival circuits and functions) are also present in humans. Survival circuit functions are not causally related to emotional feelings but obviously contribute to these, at least indirectly. The survival circuit concept integrates ideas about emotion, motivation, reinforcement, and arousal in the effort to understand how organisms survive and thrive by detecting and responding to challenges and opportunities in daily life.

1. Introduction

Emotion is a major research growth area in neuroscience and psychology today. A search of PubMed citations for the 1960s yields just over 100 papers with the word "emotion" in the title. With each subsequent decade, small increases resulted, until the last decade, when emotion titles grew exponentially—more than 2,000 hits. Emotion has happened.

But what exactly is it that has happened? What is being studied in all these papers on emotion? Actually, the term "emotion" is not well defined in most publications. Perhaps this is not surprising since there is little consensus about what emotion is, and how it differs from other aspects of mind and behavior, in spite of discussion and debate that dates back to the earliest days in modern biology and psychology (e.g., Darwin, 1872; James, 1884; Cannon, 1927, 1931; Duffy, 1934, 1941; Tomkins, 1962; Mandler, 1975; Schachter, 1975; Ekman, 1980, 1984, 1992; Izard, 2007; Frijda, 1986; Russell, 2003; Ekman and Davidson, 1994; LeDoux, 1996; Panksepp, 1998, 2000, 2005; Rolls, 1999, 2005; Damasio, 1994, 1999; Leventhal and Scherer, 1987; Scherer, 2000; Ortony and Turner, 1990; Öhman, 1986, 2009; Johnson-Laird and Oatley, 1989; Ellsworth, 1994; Zajonc, 1980; Lazarus, 1981, 1991a, 1991b; Barrett, 2006a, 2006b; Barrett et al., 2007; Kagan, 2007; Prinz, 2004; Scarantino, 2009; Griffiths, 2004; Ochsner and Gross, 2005; Lyons, 1980).

One point that many writers on this topic accept is that, while there are unique features of human emotion, at least some aspects of human emotion reflect our ancestral past. This conclusion is the basis of neurobiological approaches to emotion, since animal research is essential for identifying specific circuits and mechanisms in the brain that underlie emotional phenomena.

Progress in understanding emotional phenomena in the brains of laboratory animals has in fact helped elucidate emotional functions in the human brain, including pathological aspects of emotion. But what does this really mean? If we don't have an agreed-upon definition of emotion that allows us to say what emotion is, and how emotion differs from other psychological states, how can we study emotion in animals or humans, and how can we make comparisons between species?

The short answer is that we fake it. Introspections from personal subjective experiences tell us that some mental states have a certain "feeling" associated with them and others do not. Those states that humans associate with feelings are often called emotions. The terms "emotion" and "feeling" are, in fact, often used interchangeably. In English we have words like fear, anger, love, sadness, jealousy, and so on, for these feeling states, and when scientists study emotions in humans they typically use these "feeling words" as guideposts to explore the terrain of emotion.

The wisdom of using common language words that refer to feelings as a means of classifying and studying human emotions has been questioned by a number of authors over the years (e.g., Duffy, 1934, 1941; Mandler, 1975; Russell, 1991, 2003; Barrett, 2006a, 2006b; Kagan, 2007; Griffiths, 1997; Rorty, 1980; Dixon, 2001; Zachar, 2006). Whatever problems might arise from using feeling words to study human emotion, the complications are compounded many fold when such words are applied to other animals. While there are certainly emotional phenomena that are shared by humans and other animals, introspections from human subjective experience are not the best starting point for pursuing these. How, then, should the aspects of emotion relevant to animals and humans alike be pursued?

In answering this question it is important to separate the phenomena of interest from the overarching concept of emotion. One set of such phenomena includes responses that occur when an organism detects and responds to significant events in the course of surviving and/or maintaining well-being—for example, responses that occur when in danger or when in the presence of a potential mate or in the presence of food when hungry or drink when thirsty. These are fundamental phenomena that have always interested animal behavior scientists, and would be of interest even if the terms "emotion" and "feelings" never existed. The challenge for emotion researchers is to understand the relation of the phenomena to the field of emotion without redefining them as fundamentally emotional phenomena, and thus infusing the phenomena with confusing implications.

In this paper I, therefore, describe a way of conceiving phenomena important to the study of emotion, but with minimal recourse to the terms emotion or feelings. The focus is instead on circuits that instantiate functions that allow organisms to survive and thrive by detecting and responding to challenges and opportunities. Included, at a minimum, are circuits involved in defense, maintenance of energy and nutritional supplies, fluid balance, thermoregulation, and reproduction. These survival circuits and their adaptive functions are conserved to a significant degree in across mammalian species, including humans. While there are species-specific aspects of these functions, there are also core components of these functions that are shared by all mammals.

By focusing on survival functions instantiated in conserved circuits, key phenomena relevant to emotions and feelings are discussed with the natural direction of brain evolution in mind (by asking to what extent are functions and circuits that are present in other mammals also present in humans) rather than by looking backward, and anthropomorphically, into evolutionary history (by asking whether human emotions/feelings have counterparts in other animals).

Emotion, motivation, reinforcement, and arousal are closely related topics and often appear together in proposals about emotion. Focusing on survival functions and circuits allows phenomena related to emotion, motivation, reinforcement, and arousal to be treated as components of a unified process that unfolds when an organism faces a challenge or opportunity.

What follows is *not* an attempt at explaining or defining emotion. Instead, the aim is to offer a framework for thinking about some key phenomena associated with emotion (phenomena related to survival functions) in a way that is not confounded by confusion over what emotion means. Stepping back from the overarching concept of emotion and focusing instead on key phenomena that make emotion an interesting topic may be the best way out of the conceptual stalemate that results from endless debates about what emotion is.

2. Why do we need to rethink the relation of emotion to survival?

The relation of innate survival functions to emotions is hardly novel, and goes back at least to Darwin (1872). As a result, neuroscientists have long assumed that specific emotional/motivational circuits are innately wired into the brain by evolution, and that these mediate functions that contribute to survival and well-being of the organism (e.g., Cannon, 1929; MacLean, 1949, 1952; Hess, 1954; Stellar, 1954; von Holst and von Saint- Paul, 1962; Flynn, 1967; Olds, 1977; Siegel and Edinger, 1981; Panksepp, 1982, 1998, 2005; Blanchard and Blanchard, 1972; Bolles and Fanselow, 1980; Damasio, 1994, 1999; Berridge, 1999; McNaughton, 1989; Swanson, 2000; Ferris et al., 2008; Choi et al., 2005; Motta et al., 2009; Lin et al., 2011; Öhman, 2009). That certain emotions are wired into the brain is also a major tenet of evolutionary psychology (e.g., Tooby and Cosmides, 1990; Pinker, 1997; Nesse, 1990). If many researchers in the field (past and present) believe this, why do we need to bother with another discussion of the topic?

A major controversy in the field of emotion research today is, in fact, about the issue of whether there are innate emotion circuits in the human brain. This debate is centered on the question of whether emotions are "natural kinds," things that exist in nature as opposed to being inventions (constructions) of the human mind (e.g., Panksepp, 2000; Griffiths, 2004; Barrett, 2006a; Izard, 2007; Scarantino, 2009). Much of the discussion is focused the question of whether so-called "basic emotions" are natural kinds. Basic emotions are those that are said to be universally expressed and recognized in people around the world, conserved in our close animal ancestors, and supposedly hardwired into brain circuits by evolution (Darwin, 1872; Tomkins, 1962; Ekman, 1972, 1980, 1984, 1992, 1999a, 1999b; Izard, 1992, 2007; Damasio, 1994, 1999; Panksepp, 1998, 2000, 2005; Prinz, 2004). Contemporary theories recognize between five and seven of these basic or primary emotions. Ekman's list of six basic emotions is the canonical example (Ekman, 1972) and includes fear, anger, happiness, sadness, disgust, and surprise. This list of putative hard-wired basic emotions in fact serves as the foundation for much research on the neural basis of emotional functions in the human brain—a recent review uncovered 551 studies between 1990 and 2008 that used Ekman's basic emotions faces or variants of these to study functional activity related to emotion in the human brain (see Fusar-Poli et al., 2009).

In spite of being well known and widely applied in research, the basic emotions point of view has been challenged on various grounds (e.g., Averill, 1980; Ortony and Turner, 1990; Russell, 1980, 2003; Barrett, 2006a; Barrett et al., 2007). For one thing, different theories have different numbers of basic emotions, and even different names for similar emotions. In addition, questions have been raised about the methods used to identify basic emotions (e.g., forced choice rather than free labeling of the emotion expressed in a face). Basic emotions theory has also been challenged on the basis of a lack of coherence of the phenomena that constitute individual emotions, and the diversity of states to which a given emotion label can refer. Others argue that emotions, even so-called basic emotions, are psychological/social constructions, things created by the mind when people interact with the physical or social environment, as opposed to biologically determined states. Also relevant is the fact that the main basic emotions theory based on brain research in animals (Panksepp, 1998, 2005) lists emotions that do not match up well with those listed by Ekman or others as human basic emotions.

Of particular relevance here is Barrett's recent challenge to the natural kinds status of basic emotions, and particularly to the idea that the human brain has evolutionarily conserved neural circuits for basic emotions (Barrett, 2006a; Barrett et al., 2007). Her argument is centered on several points: that much of evidence in support of basic emotions in animals is based on older techniques that lack precision (electrical brain stimulation), that basic emotions identified in animals do not map onto the human categories, and that evidence from human imaging studies show that similar brain areas are activated in response to stimuli associated with different basic emotions. I disagree with Barrett's conclusion that the similarity of functional activation in different emotions is an argument against basic emotions since imaging does not have the resolution necessary to conclude that the similarity of activation in different states means similar neural mechanisms. Yet, I concur with her conclusion that the foundation of support for the idea that basic emotions, as conventionally conceived, have dedicated neural circuits is weak. This does not mean that the mammalian brain lacks innate circuits that mediate fundamental phenomena relevant to emotion. It simply means that emotions, as defined in the context of human basic emotions theory, may not be the best way to conceive of the relevant innate circuits. Enter survival circuits.

3. Survival circuits

It has long been known that the body is a highly integrated system consisting of multiple subsystems that work in concert to sustain life both on a moment to moment to basis and over long time scales (Bernard, 1878–1879; Cannon, 1929; Lashley, 1938; Morgan, 1943; Stellar, 1954; Selye, 1955; McEwen, 2009; Damasio, 1994, 1999; Pfaff, 1999; Schulkin, 2003). A major function of the brain is to coordinate the activity of these various body systems. An important category of life-sustaining brain functions are those that are achieved through behavioral interactions with the environment. As noted, these survival circuits include, at a minimum, circuits involved in defense, maintenance of energy and nutritional supplies, fluid balance, thermoregulation, and reproduction.

Survival circuits have their ultimate origins in primordial mechanisms that were present in early life forms. This is suggested by the fact that extant single-cell organisms, such as bacteria, have the capacity to retract from harmful chemicals and to accept chemicals that have nutritional value (Macnab and Koshland, 1972). With the evolution of multicellular, and multisystem, eukaryotic organisms (Metazoa, or what we usually call animals), fundamental survival capacities increase in complexity and sophistication, in large part due to the presence of specialized sensory receptors and motor effectors, and a central nervous system that can coordinate bodily functions and interactions with the environment (Shepherd, 1988).

The brains of vertebrate organisms vary in size and complexity. Yet, in spite of these differences, there is a highly conserved organizational plan that is characteristic of all vertebrate brains (Nauta and Karten, 1970; Northcutt and Kaas, 1995; Swanson, 2002; Butler and Hodos, 2005; Striedter, 2005). This conservation is most often discussed in terms of central sensory and motor systems. However, sensory motor systems do not exist in isolation, and in fact evolved to negotiate interactions with the environment for the purpose of sustaining life—for example, by maintaining energy and fluid supplies, regulating body temperature, defending against harm, and enabling reproduction.

The survival circuits listed do not align well with human basic emotions. However, my goal is not to align survival circuits with basic emotion categories. It is instead to break free from basic emotion categories based on human emotional feelings (introspectively labeled subjective states) and instead let conserved circuits do the heavy lifting. For example, there is no anger/aggression circuit in the present scheme. This might at first seem like a striking omission. However, it is important to note that aggression is not a unitary state with a single neural representation (Moyer, 1976; Chi and Flynn, 1971; Siegel and Edinger, 1981). Distinct forms of aggression (conspecific, defensive, and predatory aggression) might be more effectively segregated by the context in which the aggression occurs: defense circuitry (aggression in an attempt to protect one's self from harm); reproductive circuitry (aggression related to competition for mates); feeding circuitry (predatory aggression toward prey species). Similarly, a joy/pleasure/happiness kind of circuit is not listed and might seem like a fatal flaw. However, behaviors used to index joy/pleasure/happiness are instead treated products of specific circuits involved in energy and nutrition, fluid balance, procreation, thermoregulation, etc. By focusing on the subjective state, joy/pleasure/happiness, emotion theories tend to gloss over the underlying details of emotional processing for the sake of converging on a single word that symbolizes diverse underlying states mediated by different kinds of circuits.

Each survival circuit may itself need to be refined. For example, it is unlikely that there is a single unified defense or reproductive circuit. The range of functions studied needs to be expanded to more effectively characterize these. Some variations on defense are described below, but still other refinements may be needed.

Another key difference between the survival circuit and basic emotions approaches is this. Basic emotion circuits are meant as an explanation of the feelings for which each circuit is said to be responsible. Survival circuits are not posited to have any direct relation (causal role) in feelings. They indirectly influence feelings, as described later, but their function is to negotiate behavioral interactions in situations in which challenges and opportunities exist, not to create feelings.

Survival circuits help organisms survive and thrive by organizing brain functions. When activated, specific kinds of responses rise in priority, other activities are inhibited, the brain and body are aroused, attention is focused on relevant environmental and internal stimuli, motivational systems are engaged, learning occurs, and memories are formed (e.g., Morgan, 1943; Hebb, 1949; Bindra, 1969; Gallistel, 1980; Scherer, 1984, 2000; Maturana and Varela, 1987; LeDoux, 2002).

In sum, survival circuits are sensory-motor integrative devices that serve specific adaptive purposes. They are tuned to detect information relevant to particular kinds of environmental challenges and opportunities, and they use this information to control behavioral responses and internal physiological adjustment that help bring closure to the situation. All complex animals (invertebrates and vertebrates) have survival circuits. Core components of these circuits are highly conserved in vertebrates. I focus on vertebrates, especially mammals in this article, but consider the relation of invertebrate to vertebrate survival functions toward the end.

4. Nature and nurture in survival circuits

Survival circuits detect key trigger stimuli on the basis of innate programming or past experience. By innate programming I mean genetically specified synaptic arrangements that are established in early development. Innate evaluative networks make possible species-wide stimulus-response connections that allow organisms to respond to specific stimulus patterns in tried and true ways (i.e., with hard-wired/innate reactions) that have been honed by natural selection.

By experience I mean conditions under which associations are formed between novel stimuli and biologically innately significant events, typically innate triggers. These experience-dependent associations allow meaningless stimuli that occur in conjunction with significant events to acquire the ability to activate the innate response patterns that are genetically wired to innate trigger stimuli. The fact that the response patterns are innately wired and initially expressed involuntarily does not mean that they are completely inflexible. Not only can they be coupled to novel stimuli through experience and learning, they can be regulated in terms of their time course and intensity, and perhaps in other ways.

Innate and experience-based evaluative mechanisms are, as noted, circuit-specific. Thus, defense, nutritional, reproductive, thermoregulatory and other survival systems are wired to detect unique innate triggers. By entering into associations with biologically significant stimuli, novel sensory events become learned triggers that activate survival circuits. We will consider innate and learned survival circuit triggers in the context of defense next. In the field of emotion, these are described as unconditioned and conditioned fear stimuli.

5. Defense as an example

The evidence for conservation across mammals of mechanisms underlying survival functions such as defense (e.g., LeDoux, 1996, 2012; Phelps and LeDoux, 2005; Motta et al., 2009; Choi et al., 2005; Kalin et al., 2004; Amaral, 2003; Antoniadis et al., 2007), reproduction (e.g., Pfaff, 1999; Oomura et al., 1988; Blaustein, 2008), thermoregulation (Nakamura and Morrison, 2007), fluid balance (Johnson, 2007; Fitzsimons, 1979), and energy/nutritional regulation (Elmquist et al., 2005; Morton et al., 2006; Saper et al., 2002) is strong. Space does not permit a detailed discussion of these circuits and their functions. Defense circuits in mammals will be used as an initial illustration.

Defense against harm is a fundamental requirement of life. As noted above, even single-cell organisms can detect and respond to harmful environmental stimuli. In complex organisms (invertebrates and vertebrates), threat detection involves processing of innate and learned threats by the nervous system via transmission of information about the threat through sensory systems to specialized defense circuits.

Unconditioned threat stimuli are species-specific. The most common threat triggers are stimuli that signal other animals (predators and potentially harmful conspecifics), and these will obviously be different for different species. Examples of innately wired stimuli for rodents include predator odors (e.g., Motta et al., 2009; Pagani and Rosen, 2009; Blanchard et al., 1990), as well as high-frequency predator warning sounds emitted by conspecifics (e.g., Litvin et al., 2007; Choi and Brown, 2003), high-intensity auditory stimuli (e.g., Bordi and LeDoux, 1992), and bright open spaces (Thompson and LeDoux, 1974; Gray, 1987; Walker and Davis, 2002). In primates, the sight of snakes and spiders has an innate propensity to trigger defense (Amaral, 2003; Öhman, 1986; Mineka and Öhman, 2002). In spite of being genetically specified, innate stimulus processing is nevertheless subject to epigenetic modulation by various factors inside and outside the organism during development, and throughout life (Bendesky and

Bargmann, 2011; Monsey et al., 2011; McEwen et al, 2012; Brown and Hariri, 2006; Casey et al., 2011; Zhang et al., 2004). Indeed, some aspects of defense stimulus processing in primates, including humans, involves preferential rapid learning to certain classes of innately "prepared" stimuli (Seligman, 1971; Öhman, 1986; Mineka and Öhman, 2002). Fearful and aggressive faces of conspecifics are also a potent innate defense trigger in humans and other primates (Adolphs, 2008; Davis et al., 2011).

Recent studies have revealed in some detail the circuits that allow rodents to respond to unconditioned threats, especially odors that signal predators or potentially dangerous conspecifics (Dielenberg et al., 2001; Canteras, 2002; Petrovich et al., 2001; Markham et al., 2004; Blanchard et al., 2003; Motta et al., 2009; Choi et al., 2005; Vyas et al., 2007; Pagani and Rosen, 2009) (Figure 1). The odors are detected by the vomeronasal olfactory system and sent to the medial amygdala (MEA), which connects with the ventromedial hypothalamus (VMH). Outputs of the latter reach the premammillary nucleus (PMH) of the hypothalamus, which connects with dorsal periaqueductal gray (PAGd). But not all unconditioned threats are signaled by odors. Unconditioned threats processed by other (nonolfactory) modalities involve sensory transmission to the lateral amygdala (LA) and from there to the accessory basal amygdala (ABA), which connects with the VMH-PM-PAGv circuitry (Motta et al., 2009). Different subnuclei of the MEA, PMH, and PAGd are involved in processing conspecific and predatory threats. In the case of both olfactory and nonolfactory unconditioned threat signals, the PAGd and its outputs to motor control areas direct the expression of behavioral responses that help promote successful resolution of the threatening event. The PAG is also involved in detection of internal physiological signals that trigger defensive behavior (Schimitel et al., 2012).

Biologically insignificant stimuli acquire status as threat signals results when they occur in conjunction with biologically significant threats. This is called Pavlovian defense conditioning, more commonly known as fear conditioning. Thus, a meaningless conditioned

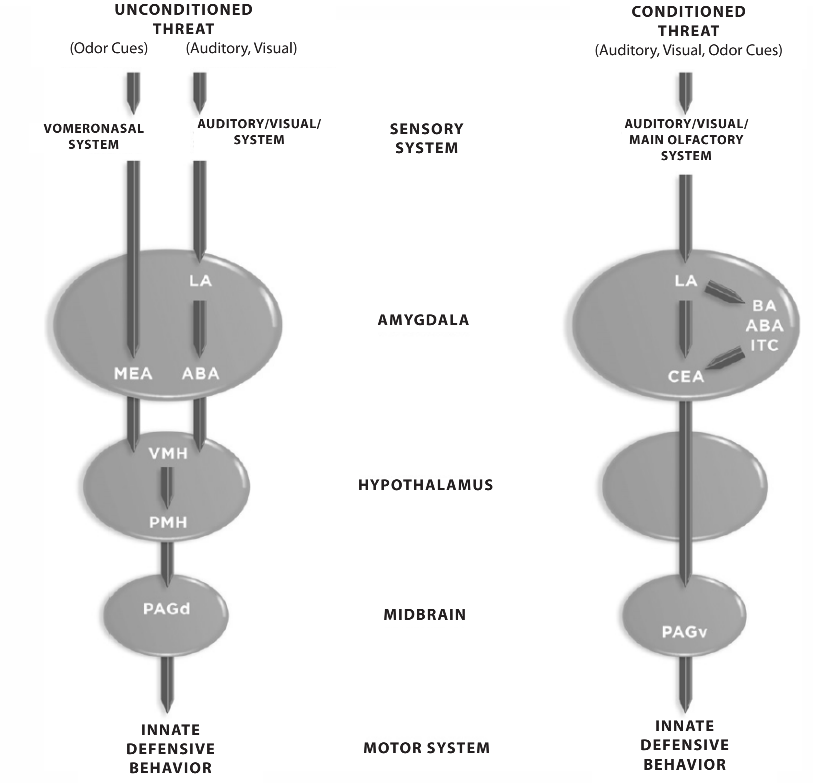


Figure 1. Circuits Underlying Defense Reactions Elicited by Unconditioned (Unlearned) and Conditioned (Learned) Threats. Abbreviations: ABA, accessory basal amygdala; BA, basal amygdala; CEA, central amygdala; LA, lateral amygdala; LH, lateral hypothalamus; MEA, medial amygdala; NAcc, nucleus accumbens; VMH, ventromedial hypothalamus; PAGd, dorsal periaqueductal gray region; PAGv, venral periaqueductal gray region; PMH, premammilary nucleus of the hypothalamus.

stimulus (CS) acquires threat status after occurring in conjunction with an aversive unconditioned stimulus (US). Most studies of Pavlovian defense conditioning involve the use of electric shock as the biologically significant US, though other modalities have been used as well. Typically, auditory, visual, or olfactory stimuli as the insignificant CS. While a strong US can induce learning to most kinds of sensory stimuli, associability is not completely promiscuous—for example, taste stimuli associate more readily with gastric discomfort than with electric shock (Garcia et al., 1968). Once the association is formed, the CS itself has the ability to elicit innate defense responses.

The neural circuit by which a CS (auditory, visual, olfactory) elicits innate defense responses, such as freezing behavior, involves transmission of sensory inputs to the LA, intra-amygdala connections (direct and indirect) linking the LA with the central nucleus of the amygdala (CEA), and connections from the medial CEA (CEm) to the ventrolateral PAG (PAGvl) (Johansen et al., 2011; LeDoux, 2000; Maren, 2001; Fanselow and Poulos, 2005; Davis et al., 1997; Rosenkranz and Grace, 2002; Cousens and Otto, 1998; Paré et al., 2004; Maren and Quirk, 2004; Quirk and Mueller, 2008; Haubensak et al., 2010). The indirect connections between LA and CEA include the basal (BA), AB, and intercalated (ITC) nuclei (Pitkänen et al., 1997; Paré et al., 2004). As with unconditioned threats, PAG outputs to motor control regions direct behavioral responses to the threat. While damage to the PAGvl disrupts defensive freezing behavior, lesions of the PAGdl enhance freezing (De Oca et al., 1998), suggesting interactions between these regions. Whether the CEA and PAG might also be linked via the VMH or other hypothalamic nuclei has not been carefully explored.

While most studies have focused on freezing, this behavior mainly occurs in confined spaces where escape is not possible (Fanselow, 1994; Blanchard et al., 1990; de Oca et al., 2007; Canteras et al., 2010). Little work has been done on the neural basis of defense responses other than freezing that are elicited by a conditioned cues (but see de Oca and Fanselow, 2004).

An important goal for future work is to examine the relation of circuits involved in innate and learned behavior. Electric shock simulates tissue damage produced by predator-induced wounds. However, it is difficult to trace the unconditioned stimulus pathways with this kind of stimulus. Recent studies exploring interactions between circuits processing olfactory conditioned and unconditioned stimuli is an important new direction (Pavesi et al., 2011).

Another form of Pavlovian defense conditioning involves the association between a taste CS and a nausea-inducing US. The circuits underlying so called conditioned taste aversion also involve regions of the amygdala, such as CEA and the basoloateral complex (which includes the LA, BA, and ABA nuclei), as well as areas of taste cortex (Lamprecht and Dudai, 2000). However, the exact contribution of amygdala areas to learning and performance of the learned avoidance response is less clear than for the standard defense conditioning paradigms described above.

While much of the work on threat processing has been conducted in rodents, many of the findings apply to other species. For example, the amygdala nuclei involved in responding to conditioned threats in rodents appear to function similarly in rabbits (Kapp et al., 1992) and nonhuman primates (Kalin et al., 2001, 2004; Antoniadis et al., 2007). Evidence also exists for homologous amygdala circuitry in reptiles (Martínez-García et al., 2002; Davies et al., 2002; Bruce and Neary, 1995) and birds (Cohen, 1974). In addition, functional imaging and lesion results from humans (e.g., Phelps, 2006; Damasio, 1994, 1999; LaBar and Cabeza, 2006; Whalen and Phelps, 2009; Büchel and Dolan, 2000; Mobbs et al., 2009; Schiller and Delgado, 2010) show that the amygdala plays a key role in defense conditioning, and thus suggest that, at least to a first approximation, similar circuits are involved in humans as in other mammals. However, the level of detail that has been achieved in humans pales in comparison to the animal work. Methods available for studying humans are, and are likely to continue to be, limited to levels of anatomical resolution that obscure circuit details.

Because animal research is thus essential for relating detailed structure to function in the brain, it is extremely important that the phenomena of interest be conceptualized in a way that is most conducive to understanding the relation of findings from animal research to the human condition. Survival circuits provide such a conceptualization.

5. Interactions between survival circuit functions

Survival circuits interact to meet challenges and opportunities. Indeed, survival functions are closely intertwined (e.g., Saper, 2006). In the presence of a threat to survival or well-being, the brain's resources are monopolized by the task of coping with the threat. Other activities, such as eating, drinking, and sex, are actively suppressed (Gray, 1987; Lima and Dill, 1990; Blanchard et al., 1990; Fanselow, 1994; Choi et al., 2005). However, increased behavioral activity of any kind (fighting, fleeing, foraging for food or drink, sexual intercourse) expends energy, depleting metabolic resources. At some point, the need to replenish energy rises in priority and overrides defensive vigilance, which might otherwise keep the animal close to home. Foraging for food or liquids often requires exposure to threats and a balance has to be struck between seeking the needed resources and staying put. Metabolic activity during any active behavior (whether fighting, feeding, foraging, fornicating) produces heat that has to be counteracted by lowering body temperature. Thermoregulation is controlled directly by homeostatic alterations that include increased sweating or panting, and by various behavioral means, such as altering fluid intake or seeking shelter. We cannot consider all possible interactions between survival circuits here. Thus, interactions between the energy/nutritional regulation system and the defense system will be discussed in some detail for illustrative purposes.

Across mammalian species, circuits involving the arcuate, ventromeidal, dorsomedial, and lateral hypothalamus, and regulated by leptin, ghrelin, glucose, and insulin, control feeding in relation to energy and nutritional demands (Elmquist et al., 2005; Morton et al., 2006; Saper et al., 2002; Saper, 2006). In satisfying nutritional/energy demands, behavioral responses are guided by the sensory properties of potential food sources and by cues associated with food. For example, auditory or visual cues that occur in connection with food items can modulate the energy/nutritional circuitry (e.g., Petrovich, 2011). Specifically, areas of the amygdala (LA, BA, ABA) process these learned cues associated with food and relay them to the LH. Such cues, if sufficiently potent, can stimulate eating in animals that are sated.

Feeding does not occur in a vacuum. As noted above, when threat levels rise, feeding is suppressed (Gray, 1987; Lima and Dill, 1990; Blanchard et al., 1990; Fanselow, 1994). For example, a tone previously paired with shock inhibits feeding (Petrovich, 2011) and food-motivated instrumental behavior (e.g., Cardinal et al., 2002). Connections from the basolateral amygdala to the LH facilitate feeding by a CS associated with food, while the suppression of feeding by an aversive CS involves outputs of the CEA. The exact target remains to be determined but CEA connects with LH both directly and indirectly (Petrovich et al., 1996; Pitkänen et al., 1997). While threat processing normally trumps feeding, at some point the risk of encountering harm is balanced against the risk of starvation. A similar case can be made for the suppression of other behaviors by threat processing. For example, medial amygdala areas that process threat related odors suppress reproduction via connections to VHM reproductive circuits (Choi et al., 2005).

The fact that the amygdala contributes to appetitive states (e.g., Rolls, 1999, 2005; Everitt et al., 1999, 2003; Gallagher and Holland, 1994; Holland and Gallagher, 2004; Cardinal et al., 2002; Baxter and Murray, 2002; Moscarello et al., 2009) as well as defense (see above) does not mean that the amygdala processes food and threat related cues in the same way. Similarly, the fact that both appetitive and aversive stimuli activate the amygdala in fMRI studies (e.g., Canli et al., 2002; Hamann et al., 2002; Lane et al., 1999) does not mean that these stimuli are processed the same by the amygdala. Recent unit recording studies in primates show that appetitive and aversive signals are processed by distinct neuronal populations of cells in the lateral/ basal amygdala (Paton et al., 2006; Belova et al., 2007; Belova et al., 2008; Morrison and Salzman, 2010; Ono and Nishijo, 1992; Rolls, 1992, 1999, 2005). Molecular imaging techniques with cellular resolution show that similarities in activation at the level of brain areas obscures differences at the microcircuit level (Lin et al., 2011).

6. Circuit functions versus behavioral responses

Because different groups of mammals faced different selective pressures, the behavioral responses controlled by conserved survival circuits can differ. As ethologists have long noted, many survival-related behaviors are expressed in species-specific ways (e.g., Tinbergen, 1951; Lorenz, 1981; Manning, 1967).

Consider escape from a threat. We've seen evidence for conserved defense circuits across mammals and even across vertebrates, but behavioral responses controlled by these circuits can differ dramatically. For example, while most mammals flee on all fours, some use only two legs (humans), others escape by flying (bats), and still others by swimming (whales, seals, and walrus). Similarly, sexual and feeding behavior, while largely conserved at the neural system level, is also expressed behaviorally in diverse ways within mammals. For example, although androgen activity in the hypothalamus is important in all male mammals, the semen delivery process varies in males, in part because of different approaches required given the configuration of the male and female body (e.g., Pfaff, 1999). This is perhaps most dramatically illustrated by the lordosis posture of female rats. The male cannot insert his penis into the vaginal cavity of a female unless she arches her back to adopt this posture, which is regulated by the binding of estrogen during the fertile phase of her cycle (Pfaff, 1999; Blaustein, 2008). Further, some mammals use their snouts when eating and others their paws/hands, but the core circuits described above nevertheless regulate the various homeostatic and behavioral functions required to regulate energy and nutritional supplies.

Thus, the responses used by survival circuits to achieve survival goals can be species-specific even though the circuit is largely species-general (obviously, there must be some differences in circuitry, at least in terms of motor output circuitry for different kinds of behaviors, but the core circuit is conserved). By focusing on the evolved function of a circuit (defense, reproduction, energy and nutrition maintenance, fluid balance, thermoregulation), rather than on the actual responses controlled by the circuit, a species-independent set of criteria emerge for defining brain systems that detect significant events and control responses that help meet the challenges and opportunities posed by those events.

7. Information processing by survival circuits: computation of stimulus significance

A key component of a survival circuits is a mechanism for computing circuit-specific stimulus information. A defense circuit needs to be activated by stimuli related to predators, potentially harmful conspecifcs, and other potential sources of harm, and not be triggered by potential mates or food items. The goal of such computational networks is to determine whether circuit-specific triggers are present in the current situation, and, if a trigger is detected, to initiate hardwired (innate) responses that are appropriate to the computed evaluation. Such responses are automatically released (in the ethological sense—see Tinbergen, 1951; Lorenz, 1981; Manning, 1967) by trigger stimuli.

The nature of behavioral responses released by survival circuit triggers should be briefly discussed. Activation of a survival circuit elicits behavioral responses on the spot in some cases (e.g., in the presence of defense triggers) but in other cases unless the goal object (sexual partner, food, drink) is immediately present, the more general effect is the alteration of information processing throughout the brain in such a way as to mobilize resources for bringing the organism into proximity with suitable goal objects and thus dealing with the opportunity or challenge signaled by the trigger. We will consider a number of different consequences of survival circuit activation below. Here, we focus on information processing related to trigger detection.

Above we briefly noted the species-specific nature of innate trigger stimuli. While the original idea of the ethologists focused on complex Gestalt configural stimuli and pattern recognition, simpler features are now emphasized. Thus, a rat can recognize a predator (cat, fox) by specific chemical constituents of predator odors (Wallace and Rosen, 2000; Vyas et al., 2007; Dielenberg et al., 2001; Markham et al., 2004; Blanchard et al., 2003) and does not have to recognize the predator as a complex perceptual pattern. Moreover, humans can recognize certain emotions by the eyes alone and do not need to process the face as a whole (e.g., Whalen et al., 2004), and evidence exists that this can be handled subcortically (Liddell et al., 2005; Morris et al., 1999; Tamietto et al., 2009; Luo et al., 2007). These findings are consistent with the notion that that relatively simple sensory processing by subcortical areas can provide the requisite inputs to structures such as the amygdala, bypassing or short-circuiting cortical areas (LeDoux, 1996). In contrast to innate trigger stimuli, learned triggers are less restricted by species characteristics. Thus, many (though not all, as noted above) stimuli can be associated with harm and become a trigger of defense circuits later.

In the field of emotion, the term automatic appraisal is sometimes used when discussing how significant stimuli elicit socalled emotional responses automatically (without deliberate control), and is contrasted with cognitive or reflective appraisal, where processing that is deliberate, controlled and often conscious, determines stimulus meaning and predisposes actions (e.g., Arnold, 1960; Bowlby, 1969; Frijda, 1986; Lazarus, 1991a, 1991b; Leventhal and Scherer, 1987; Lazarus and Folkman, 1984; Smith and Ellsworth, 1985; Scherer, 1988; Scherer et al., 2001; Sander et al., 2005; Jarymowicz, 2009).

The stimulus significance evaluations by survival circuits are obviously more in line with automatic, unconscious appraisal mechanisms. However, while stimulus evaluations by survival circuits is clearly an example of automatic appraisal, one should not be too quick to assume that what psychologists refer to as automatic appraisals in humans is identical to survival circuit processing. The latter probably refers to a narrower set of phenomena than the former, at least in humans, if not other species, though the range of phenomena in question clearly overlap.

8. Multiple roles of innate and learned stimuli

So far we've seen that unconditioned and conditioned emotional stimuli can be thought of in other terms, as unconditioned and conditioned survival circuit triggers. In addition, though, they can also be described as incentives—stimuli that motivate instrumental behavior. The same stimuli additionally function as reinforcers—stimuli that strengthen the probability that an instrumental response will be learned and later performed. Motivation and reinforcement are obviously closely aligned with the topic of emotion, though these are often studied separately today. Let's look more closely at how closely intertwined these processes are to one another (Figure 2).

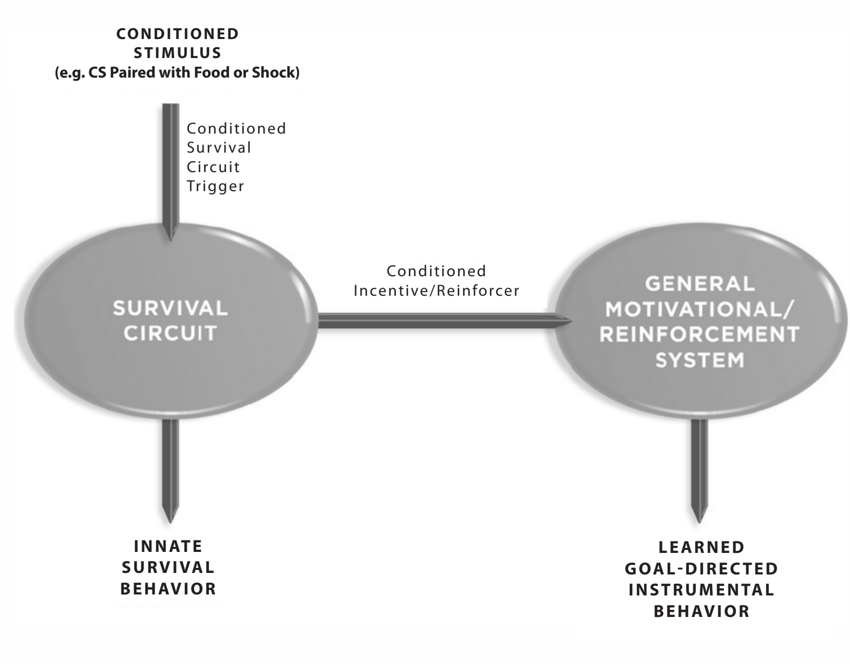


Figure 2. Multiple Roles for a Conditioned Stimulus. A CS functions as a survival circuit trigger (by activating a specific survival circuit related to the US that was used during conditioning), and as a conditioned incentive and a conditioned reinforcer (by way of connections from the survival circuit to motivational and reinforcement systems). Other routes by which a CS might influence motivational and reinforcement circuitry are not shown.

Consider a tone that is paired with food. This is a typical method used to study positive emotional states in animals. The tone in other words is an appetitive Pavlovian CS that elicits innate approach behavior. However, it is also a survival circuit trigger, as it can stimulate eating, even in satiated rats, by activating hypothalamic circuits involved in energy management (Petrovich, 2011). The same CS will also function as a conditioned incentive that can modulate instrumental behaviors (in contrast to the ability of a CS to elicit Pavlovian innate (unconditioned approach) behaviors. Thus, a CS associated with food will facilitate performance of an instrumental response that is also maintained by food (e.g., bar-pressing for food) (Corbit and Balleine, 2005; Cardinal et al., 2002; Balleine and Killcross, 2006). This is called Pavlovian-to-instrumental transfer since the value of the Pavlovian CS is transferred to (alters performance of) the instrumental response. The degree of transfer depends in part on the similarity of the US in the Pavlovian and instrumental tasks. A tone CS can also be used to reinforce the learning of a new instrumental response (e.g., Holland and Rescorla, 1975). Thus, a hungry rat will learn to press a bar simply to receive the tone CS. In this case the tone is considered a reinforcer, a second-order or conditioned reinforcer (a first order or primary reinforcer would be something like food itself rather than a stimulus associated with food).

Similar relations hold for a tone paired with an aversive US, footshock. The tone CS elicits innate freezing behavior (see above) and is thus often described as a conditioned emotional stimulus (conditioned fear stimulus in this case). And just as an appetitive CS enhances bar pressing for food, and aversive CS suppresses food-maintained bar pressing (Estes and Skinner, 1941; Hammond, 1970; Cardinal et al., 2002; Balleine and Killcross, 2006). However, an aversive CS will also facilitate performance of an aversively motivated behavior (Hammond, 1970; Lázaro-Muñoz et al., 2010). Further, just as rats will learn to perform new instrumental responses for the sole reward of receiving an appetitive CS, they will also learn new instrumental responses that are rewarded by the elimination of an aversive CS (e.g., Cain and LeDoux, 2007).

Although we've focused on multiple roles of CSs a similar argument can be made for USs. These are simply stimuli that innately activate survival circuits, promote the performance of consummatory responses (food is eaten, sex is consummated) in their presence, or support Pavlovian associative conditioning or instrumental conditioning.

If we choose, we can thus describe a variety of the effects of so-called "emotional" stimuli without the use of the adjective "emotional". These are innate or learned stimuli that activate survival circuits and trigger the expression of the innate responses controlled by these circuits, that modulate the performance of learned (previously reinforced) instrumental behaviors, and that lead to the reinforcement of new instrumental behaviors (Table 1).

Table 1. Multiple Roles for So-Called "Emotional" Stimuli

1. Survival Circuit Trigger Stimulus	Activates a specific survival circuit
Innate (Unconditioned) trigger	Elicits innate responses to stimuli without the need for prior exposure to the stimulus and mobilizes other brain resources to deal with the opportunity or challenge presented by the innate trigger
Learned (Conditioned) trigger	Potentially elicits innate responses to stimuli after being associated (via Pavlovian conditioning) with an innate trigger; more generally, mobilizes brain resources to deal with the challenge or opportunity signaled by the learned trigger
2. Incentive	Modulates instrumental goal-directed behavior to help meet the opportunity or challenge signaled by the stimulus that is triggering activation of a specific survival circuit
Innate (unconditioned or primary) incentive	Increases approach toward or avoidance of the stimulus in an effort to resolve the challenge or opportunity present
Learned (conditioned or secondary) incentive	Invigorates and guides behavior toward situations where the challenge or opportunity present can be resolved
3. Reinforcer	Supports the learning of Pavlovian or instrumental associations
Innate (unconditioned or primary) reinforce	Induces the formation of associations with neutral stimuli that occur in its presence (through Pavlovian conditioning) and to the formation of associations with responses that lead to the presentation (appetitive stimuli) or removal (aversive stimuli) of the stimulus (through instrumental conditioning)
Learned (conditioned or second-order) reinforce	Induces formation of associations with other stimuli (through Pavlovian second-order conditioning) or with goal directed responses (through second-order instrumental conditioning)

9. Motivation in the survival circuit scheme

Emotion and motivation were traditionally treated as separate topics. Emotion was viewed as a reaction (e.g., a fearful, angry, disgusted, joyful, or sad emotional reaction) to some environmental situation, and motivation as a drive from within (e.g., hunger, thirst, or sexual drive) (e.g., Hull, 1943; Stellar, 1954). In the late 1960s, the emergence of the concept of incentives helped bring these together (Bindra, 1969; Trowill et al., 1969). Bindra (1969), for example, argued that emotion, like motivation, is influenced by internal factors (e.g., hormones) and motivation, like emotion, is impacted by external stimuli (incentives).

Motivation, as assessed behaviorally, involves approach toward desired outcomes and avoidance of undesired outcomes (Tolman, 1932; McClelland et al., 1953; Schneirla 1959, Elliot and Church, 1997; Cofer, 1972; Cofer and Appley, 1964; Miller, 1944; Trowill et al., 1969; Bindra, 1969; Davidson, 1993; Gray, 1982; Lang et al., 1990; Berridge, 2004; Cardinal et al., 2002; Balleine and Dickinson, 1998; Holland and Gallagher, 2004; Gallagher and Holland, 1994; Everitt and Robbins, 2005). So-called approach/avoidance motivation often occurs in two stages: an anticipatory/exploratory/search for goal objects and the performance and consummatory responses (innate responses controlled by survial circuits) once goal objects are in reach (Sherrington, 1906; Tinbergen, 1951; Cardinal et al., 2002; Berridge, 1999, 2007).

The anticipatory/exploratory/search phase is guided by incentives (Bindra, 1968; Trowill et al., 1969; Balleine and Dickinson, 1998; Cardinal et al., 2002; Johnson et al., 2009; Petrovich et al., 2002; Berridge, 1999, 2007, 2004; Rolls, 1999, 2005; Glimcher, 2003). Incentives, as noted, are essentially innate or conditioned emotional stimuli; in other words, stimuli with the potential to activate survival circuits.

One of the key discoveries that led to the rise of incentive views was that stimuli that lacked the ability to satisfy needs and reduce drives (for example, the nonnutritive sugar substitute saccharin) were nevertheless motivating (Sheffield and Roby, 1950; Cofer, 1972). A major consequence was that the connection between motivation and specific functional circuits (what we are calling survival circuits) began to be deemphasized. Motivation became a somewhat generic process by which behavior was invigorated and guided toward goals by incentives.

The nucleus accumbens emerged as a key focal point of this general motivational system (Graybiel, 1976; Mogenson et al., 1980; Balleine and Killcross, 1994; Killcross and Robbins, 1993; Everitt et al., 1999; Cardinal et al., 2002; Ikemoto and Panksepp, 1999; Parkinson et al., 1999; Koob, 2009; Sesack and Grace, 2010; Berridge, 2007, 2009; Berridge and Robinson, 1998; Hyman et al., 2006; Nestler, 2004; Kelley, 2004). Behavioral invigoration or energization was said to be a function of dopamine release in the accumbens and incentive processing by the accumbens was thought to guide behavior toward goals. Other areas involved in incentive motivation, such as the orbito-frontal cortex, are not considered here (see Rolls, 1999, 2005).

A key question is whether motivation is a generic process or whether motivationally specific processing by survival circuits might be significant as well. While there may indeed be generic aspects of motivation (e.g., behavioral invigoration), evidence also supports motivationally specific information processing as well. At the behavioral level, bar pressing for food by a hungry obtain food is facilitated by a conditioned incentive that signals food, is facilitated less by one that signals water and is inhibited by one that signals shock (Corbit and Balleine, 2005; Hammond, 1970), indicating that motivation is tied to specific survival functions. Lateral hypothalamic circuits that control energy maintenance through feeding modulate nucleus accumbens activity (Sears et al., 2010). The accumbens, once thought to be mainly involved in processing appetitive stimuli, is now know to contribute to the processing of aversive incentives as well (Salamone, 1994; Schoenbaum and Setlow, 2003; Roitman et al., 2005; Reynolds and Berridge, 2008). Within the accumbens information processing segregated along motivational lines—aversive and appetitive stimuli are processed separately at the cellular and molecular level (Roitman et al., 2005, 2008). While most work is at the level of appetitive versus aversive states, it would be important to determine whether incentives related to different appetitive survival circuits (e.g., incentives related to food versus sex) are processed separately.

Once incentives have guided the organism to goal objects, innate consummatory responses, which are specific to the particular survival circuit and function, are initiated. Their termination essentially ends the survival (emotional) episode—food is eaten, liquid is drunk, sex is consummated, safety is reached.

Before leaving the topic of motivation of instrumental goal-directed behavior it is important to mention that such behaviors, when repeatedly performed in recurring situations, can become habitual and divorced from the actual attainment of the goal. In such cases of stimulus-response habit formation, the neural control switches from the ventral to the dorsal striatum (Everitt and Robbins, 2005; Wickens et al., 2007; Packard and Knowlton, 2002).

10. Reinforcement and survival circuits

Reinforcement and motivation are closely related. Things that motivate are often reinforcing, and vice versa. Like motivation, reinforcement was once linked to drive states (Hull, 1943), but drifted toward generic mechanisms over the years. The discovery that behavior could be reinforced by electrical stimulation of brain areas (Olds and Milner, 1954), and findings that electrical reinforcement could summate with different natural reinforcers (Coons and White, 1977; Conover and Shizgal, 1994), were compatible with a generic mechanism of reinforcement. Similarly, that addictive drugs and natural or electrical reinforcers interact (Wise, 2006) is also consistent with a generic mechanism. Further, influential mathematical models of reinforcement (e.g., Rescorla and Wagner, 1972; Sutton and Barto, 1987) explained learning with singular learning rules. The modern paradigmatic example of a generic reinforcement mechanism is the role of dopamine in the striatum as a reward prediction error signal (Schultz, 1997).

Nevertheless, there have from time to time been calls for linking reinforcement more directly to specific neurobiological systems. For example, Glickman and Schiff (1967) proposed that reinforcement is a facilitation of activity in neural systems that mediate species-specific consummatory acts. In other words, they proposed a link between reinforcement and motivationally-specific survival circuits. It is therefore of great interest that recent work on the role of dopamine as a reward prediction error signal is beginning to recognize the importance of specific motivational states in modulating the effects of dopamine as a reward prediction error signal (Schultz, 2006; Glimcher, 2011).

The expression of reinforcement as a change in the probability that an instrumental response will be performed may well occur via a generic system in which the reinforcer strengthens the response (e.g., via contributions of dopamine in the striatum to reward prediction errors). But, in addition, survival circuit-specific motivational information is likely to contribute at a fundamental level, providing the stimulus with the motivational value that allows it to ultimately engage the more generic mechanisms that strengthen instrumental responses and that motivate their performance.

Reinforcement principles have been used by some authors to classify emotional states (e.g., Gray, 1982; Rolls, 1999, 2005; Cardinal et al., 2002; Hammond, 1970; Mowrer, 1960). In these models various emotions defined in terms of the presentation or removal of reinforcers. Mowrer (1960), for example, proposed a theory in which fear, hope, relief, and disappointment were explained in these terms. Later authors have attempted to account for more conventional emotions (fear, sadness, anger, pleasure, etc) as products of the presentation or removal of reinforcement. This approach suffers from some of the same problems as basic emotions theory in that it focuses on common language words related to human feelings as the way to identify emotion mechanisms in the brain. Perhaps reinforcement, like motivation, might be fruitfully linked to emotional phenomena through the survival circuit conception.

11. Survival circuits and arousal

Survival circuits are engaged in situations in which challenges and/or opportunities exist, in other words what we commonly call emotional or motivated situations. So far we have focused on two major consequences of survival circuit activation. One is the elicitation of specific kinds of hard-wired behavioral reactions. The second is an increase in the probability that instrumental goal-directed actions relevant to the opportunity or challenge will be learned (reinforced) and performed (motivated)—or, if the situation has been experienced by the individual repeatedly in the past, stimulus-response habits may substitute for incentive guided instrumental goal-directed action.

A third consequence of survival circuit activation is "generalized arousal" (Moruzzi and Magoun, 1949; Lindsley, 1951; Schober et al., 2011; Lang, 1994; Pfaff et al., 2008). As originally conceived, generalized arousal was a function of the brainstem reticular activating system (Moruzzi and Magoun, 1949; Lindsley, 1951). Later, the undifferentiated reticular activating system concept gave way to the notion that distinct populations of chemically specific neurons that underlie sleep-wake cycles and the degree of arousal, attention, and vigilance while awake (Jouvet, 1969, 1999; Steriade, 1995, 2004; Jacobs et al., 1990; Jones, 2003; Aston-Jones, 2005; Monti and Jantos, 2008; Sarter et al., 2005; Arnsten and Li, 2005; Robbins, 2005; Nieuwenhuys, 1985; Nishino, 2011). Specifically, neurons that synthesize and release biogenic amines (norepinephrine, dopamine, serotoinin, or acetylcholine) and peptides (e.g., orexins) are believed to make significant contributions to brain arousal. While these transmitters are released in widespread areas of the brain, their effects are especially profound on neurons that are actively engaged in information processing (Aston-Jones et al., 1991; Foote et al., 1983, 1991; Aston-Jones and Bloom, 1981). That is, they modulate rather than initiate neural activity, regulating neuronal excitability and neurotransmission (Schildkraut and Kety, 1967; Hasselmo, 1995; Lopez and Brown, 1992). Also contributing to generalized arousal are peripheral systems that release hormones into the circulation (e.g., cortisol released from the adrenal cortex, adrenergic hormones, epinephrine and norepinephrine, from the adrenal medulla; and others) (Axelrod and Reisine, 1984; McEwen, 2009; Sapolsky et al., 1986). Cortisol crosses the blood brain barrier and binds to receptors in a variety of areas, while adrenergic hormones affect the CNS indirectly (McGaugh, 2000). The modulatory effects of central modulators are relatively rapid, whereas the effects of peripheral hormones are considerably slower, allowing the prolongation of the survival state for extended periods of time.

Generalized arousal has played a key role in a number of theories of emotion over the years (e.g., Duffy, 1941; Lindsley, 1951; Schachter and Singer, 1962; Schachter, 1975; Schildkraut and Kety, 1967; Mandler, 1975; Lang, 1994; Robbins, 1997) and is also important in contemporary dimensional theories of emotion (Russell, 1980, 2003; Russell and Barrett, 1999) and some neural models of emotion (e.g., Davis and Whalen, 2001; Gallagher and Holland, 1994; Kapp et al., 1994; Lang and Davis, 2006). However, it is important to ask how generalized arousal is triggered in emotional situations, and how the arousal, once present, affects further processing. Again, the defense circuit is useful for illustrative purposes.

The detection of a threat by defense circuits of the amygdala leads to the activation of central neuromodulatory and peripheral hormonal systems (see Gray, 1993; LeDoux, 1992, 1995; Davis, 1992; Rodrigues et al., 2009). Thus, central amygdala outputs target dendritic areas of norepinephrine, dopamine, serotonin, and acetylcholine containing neurons and cause these to release their chemical products in widespread brain areas (e.g., Reyes et al., 2011; Gray, 1993; Weinberger, 1995; Kapp et al., 1994). Central amygdala outputs also target neurons that activate the sympathetic division of the autonomic nervous system, which releases adrenergic hormones from the adrenal medulla, and the hypothalamic-pituitary-adrenal axis, which releases cortisol from the adrenal cortex (Gray, 1993; Talarovicova et al., 2007; Loewy, 1991; Reis and LeDoux, 1987). Threats thus not only elicit specific defense responses but also initiate generalized arousal in the brain and body. Body feedback has played an important role in emotion theory for more than a century (James, 1884; Lange, 1885/1922; Schachter and Singer, 1962; Tomkins, 1962; Adelmann and Zajonc, 1989; Buck, 1980; Damasio, 1994, 1999).

One consequence of this pattern of connectivity is that central and peripheral arousal signals facilitate processing in the survival circuit that triggered the activation of arousal. This establishes a loop in which continued activation of the survival circuit by external stimuli produces continued activation of the modulator release, which in turn facilitates the ability of external stimuli to continue to drive the survival circuit. Indeed, modulators facilitate activity in sensory processing areas (e.g., Hurley et al., 2004), which should enhance attention to external stimuli present during survival circuit activation. Modulators also facilitate processing areas involved in retrieving, forming, and storing memories (McGaugh, 2003; Roozendaal et al., 2009). All of these effects are recapitulated in motivational circuits once the initial reaction begins to give way to goal-directed instrumental actions. For example, dopamine contributes to the invigoration or activation of behavior during the exploratory search phase of a motivated state (Berridge 2004; Berridge and Robinson, 1998; Robbins and Everitt, 2007). Norepinephrine, serotonin, acetylcholine, orexins and other modulators also contribute. While arousal is often discussed in terms of generic (generalized) mechanisms, the possibility that some aspects of arousal might be survival circuit specific should also be explored (Pfaff et al., 2008; Schober et al., 2011).

12. Global organismic states

Survival circuit activation leads to the triggering of arousal responses in the CNS and to the potential expression of innate behaviors (depending on the circumstances), as well as expression of autonomic nervous system and hormonal responses in the body. Behavioral, autonomic, and endocrine responses feedback to the brain and also contribute to arousal. In addition, motivational systems are activated, potentially leading to goal-directed behaviors (Figure 3). The overall result of survival circuit-specific activity, motivational activity, and generalized arousal is the establishment of a state in which brain resources are coordinated and monopolized for the purpose of enhancing the organism's ability to cope with a challenge and/or benefit from opportunities. The organism becomes especially attentive to and sensitive to stimuli relevant to the survival function, memories relevant to the survival function are retrieved, and previously learned instrumental responses relevant to the survival function are potentiated. New learning occurs and new explicit memories (via the hippocampus and related cortical areas) and implicit memories (memories stored in the survival circuit) are formed. Such states will be referred to here as global organismic states. The fact that these states are global does not mean that they completely lack specificity. They include survival circuit-specific components, as well as general motivational components that control instrumental behavior and components that control nonspecific or generalized arousal within the brain and body.

The notion that emotional and motivated states have profound effects on the brain, recruiting widespread areas into the service of the immediate situation, monopolizing and/or synchronizing brain resources, has been proposed previously (Gallistel, 1980; Maturana and Varela, 1987; Scherer, 2000; LeDoux, 2002, 2008). Particularly relevant is the "central motive state" hypothesis (Morgan, 1943; Hebb, 1949; Bindra, 1969). Yet, the exact nature of global organismic states is poorly understood. In part this is likely attributable to the lack of

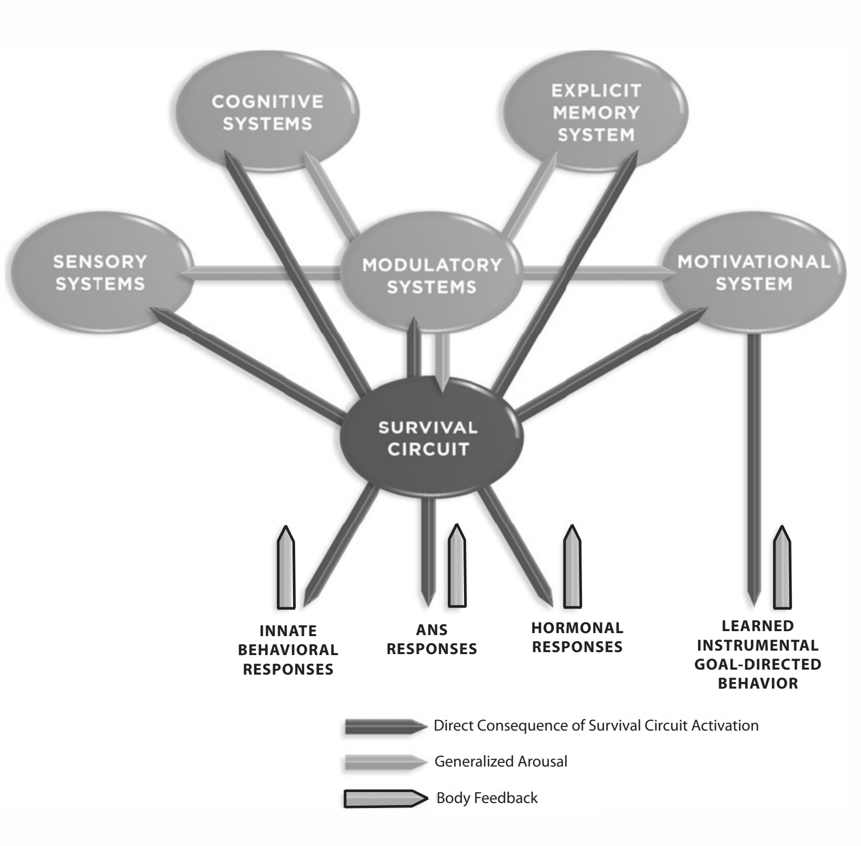


Figure 3. Consequences of Survival Circuit Activation. When a survival circuit trigger activates a survival circuit, a number of consequences follow. (1) Innate behavioral responses are potentially activated, as well as autonomic nervous system (ANS) responses and hormonal responses. These each generate feedback to the brain. (2) Neuromodulator systems are activated and begin to regulate excitability and neurotransmission throughout the brain. (3) Goal-directed instrumental behavior is initiated by the motivation system. (4) Sensory, cognitive, and explicit memory systems are also affected, leading to enhanced attention to relevant stimuli and the formation of new explicit memories (memories formed by the hippocampus and related cortical areas) and implicit memories (memories formed within the survival circuit).

techniques for assessing neural activity across widespread areas of the brain at a sufficiently detailed level of resolution. Measurement of BOLD activity in the brains of humans or animals with fMRI allows whole brain analysis of functional activity, but lacks spatial resolution at the level of cells and circuits. Use of molecular markers, such as the expression of immediate early gene activity, in relation to behavior holds promise. Particularly important would be the development of techniques that could provide widespread simultaneous assessment of changes in body physiology and brain activation and related to survival circuit processing, general-purpose motivational processing, and generalized arousal.

13. Transcending neuroanatomical homology: survival throughout the animal world

Invertebrates do not have the same conserved circuits that vertebrates have. However, they face many of the same problems of survival that vertebrates do: they must defend against danger, satisfy energy and nutritional needs, maintain fluid balance and body temperature, and reproduce. As in vertebrates, specific circuits are associated with such functions, though different invertebrates have different nervous systems and different circuits.

The fact that invertebrate nervous systems are diverse and differ from the canonical vertebrate nervous system does not mean the invertebrates are irrelevant to understanding survival functions (and thus so-called emotional behavior) in vertebrates. Much progress is being made in understanding innate behaviors related to survival functions such as defense, reproduction and arousal in invertebrates such as *Drosophila* (Wang et al., 2011; Lebestky et al., 2009; Dickson, 2008) and *C. elegans* (McGrath et al., 2009; Pirri and Alkema, 2012; Garrity et al., 2010; Bendesky et al. 2011). In these creatures, as in mammals and other vertebrates, G protein-coupled receptors and their regulators play key roles in modulating neuronal excitability and synaptic strength, and in setting the threshold for behavioral responses to incentives associated with specific motivational/emotional states (Bendesky and Bargmann, 2011). Biogenic amines and their G protein-coupled receptors also play a key role in arousal and behavioral decision making in *Drosophila* (Lebestky et al., 2009) and *C. elegans* (Bendesky et al., 2011) as in vertebrates (see above), and genetic mechanisms underlying survival-based learning in invertebrates. For example, such as in *Aplysia californica* many of the neurotransmitters (e.g., glutamate), neuromodulators (e.g., serotonin, dopamine), intracellular signals (e.g., protein kinase A, map kinase), transcription factors (e.g., cyclic AMP response element binding protein) involved in defense conditioning Aplysia (e.g., Hawkins et al., 2006; Kandel, 2001; Carew and Sutton, 2001; Glanzman, 2010; Mozzachiodi and Byrne, 2010) have been implicated in defense conditioning in the mammalian amygdala (see Johansen et al., 2011). Further, studies in *Drosophila* have implicated some of the same intracellular signals and transcription factors in defense-based learning (Dudai, 1988; Yovell et al., 1992; Yin and Tully, 1996; Margulies et al., 2005).

Similarities at the cellular and molecular level, and presumably at the level of genes that encode these processes, across diverse groups of animals is impressive evidence for conserved principles of organization underlying survival functions. However, an important question is whether there might be more fundamental circuit principles that are instantiated at the microcircuit level in nervous systems that are superficially distinct. If so, the key to understanding the relation of survival functions across invertebrates and vertebrates is likely to involve conserved principles of organization at the microcircuit level rather similarity of anatomical structures or molecules (David Anderson, personal communication). Very interesting examples are emerging from studies of olfactory processing, for which analogies in behaviorally relevant peripheral odor-encoding and central representation occur using similar organizational principles in anatomically distinct (nonhomologous) structures in *Drosophila* and rodents (see Bargmann, 2006; Sosulski et al., 2011; Wang et al., 2011).

Survival functions instantiated in specific neural circuits likely reflect conserved neural principles. We should at least be amenable to the possibility that defense, reproduction, and other survival functions in humans, may be related to survival functions in invertebrates.

This notion is not likely to be surprising to card carrying comparative neurobiologist, but might meet more resistance from researchers who study humans since survival functions account for some fundamental emotional functions in humans, and in humans emotions are often equated with or closely tied to feelings. But the thrust of what has been said here is that survival functions should not be treated as qualitatively differently in humans and other mammals, in mammals and other vertebrates, in vertebrates and invertebrates. As noted earlier, a case can even be made that solutions to fundamental problems of survival are in the final analysis derived from solutions to these problems present primordial single-cell organisms.

14. Survival circuits and human feelings: what is an emotional state?

When the term “emotional state” is used, the user typically has the notion of “feeling” in mind. This article is an attempt to redefine the nature of such states, at least the components of such states that are shared across mammalian species (and likely across vertebrates, and to some extent in invertebrates as well). Nevertheless, the history of emotion research and theory is for the most part the history of trying to understand what feelings are and how they come about. It is thus important to comment on the nature of feelings and their relation to survival circuits.

One might be tempted to conclude that global organismic states, or at least the central representation of such sates, constitute neural correlates of feelings. Global organismic states make major contributions to conscious feelings but the two are not the same. Global organismic states are part of the raw material from which certain classes of feelings are constructed (those feelings associated with survival circuit activation). But they could, and likely do, exist, independent of feelings, at least in relation to what humans call feelings. My proposal is that these kinds of feelings (those associated with survival circuit activation) occur in humans when consciousness (1) detects that a survival circuit is active or witnesses the existence of a global organismic state initiated by the activation of a survival circuit in the presence of particular kind of challenge or opportunity and (2) appraises and labels this state. These are not the only kinds of feelings that can occur in humans. Other kinds include feelings associated with higher-order or social emotions (guilt, shame, envy, pride) or sensory feelings (a pleasant touch or an annoying itch).

What about other animals? To the extent that nonhuman organisms have consciousness and cognition, capacities that allow the observation, appraisal, and categorization of survival circuit activity or global organismic states, they can have feelings when survival circuit activity or global organismic states occur. To the extent that the mechanisms of consciousness and cognition differ in different animals (with humans included as an animal), and to the extent that the mechanisms underlying survival circuit or global organismic states themselves differ, feelings will be different. This leaves open the possibility that conscious feelings can be present in other mammals, other vertebrates, or even in invertebrates. But rather than engaging in idle speculation about this, criteria can be offered that can help address the question. Specifically, if we can understand what underlies conscious feelings in humans, we can then search for whether those mechanisms are present, and to what extent they are present, in other animals.

This, you probably noticed, is a different approach from the one advocated earlier for survival circuits. We now ask whether processes in humans are present in other animals. But just as the survival circuit question should be asked about whether mechanisms in other animals are present in humans, the question of whether mechanisms shown to be present in humans are present in other animals seems only addressable in the other direction. We can never know whether another animal has conscious emotional feelings, but we might be able to determine whether the mechanisms that make of consciousness and feelings possible in humans also present in other animals.

The fact is that the brain mechanisms that underlie conscious emotional feelings in humans are still poorly understood. However, this should not stand in the way of understanding survival functions and the states that occur in the brain when the circuits mediating survival functions are activated. There is much work to be done even if we don't have viable solutions to the problems of conscious feelings.

Research on feelings is complicated because feelings cannot be measured directly. We rely on the outward expression of emotional responses, or on verbal declarations by the person experiencing the feeling, as ways of assessing what that person is feeling. This is true both when scientists do research on emotions, and when people judge emotions in their social interactions with one another.

When not wearing a scientific hat, most of us apply introspectively based concepts to other animals. When a deer freezes to the sound of a shotgun we say it is afraid, and when a kitten purrs or a dog wags its tail, we say it is happy. In other words, we use words that refer to human subjective feelings to describe our interpretation of what is going on in the animal's mind when it acts in way that has some similarity to the way we act when we have those feelings. Some authors also claim that similarity of behavior is strongly suggestive of similarity at the level of subjective experience (Panksepp, 1998, 2005) or more generally that humans know what an animal feels from observing its behavior (Bekoff, 2007; Masson and McCarthy, 1996). But it's hard to justify anthropomorphic speculation in science. Panksepp has attempted this (Panksepp, 1982; 1998, 2000; 2005), but few scientists are convinced that this is the way to go, as there is no way to objectively verify what another organism experiences.

So what's the difference, if any, between attributing feelings to other people and to other animals? There is a strong rationalization for assuming all humans have subjective mental states, such as feelings, that are similar in kind. In the absence of genetic mutations of the nervous system or acquired brain damage, each human possesses the same basic kind of brain, a brain with the same basic neural systems, as every other human. As a result we expect that other people have the same kinds of basic brain functions, and corresponding mental capacities, that we have, and we can assume with some confidence that other people experience the same kinds of feelings we do when we they behave the way we behave when we have those feelings (unless they are being intentionally deceitful). We can therefore fairly comfortably apply our introspections about our own feelings to the mental states of other people on the basis of their behavior.

We should not, however, be so comfortable in talking about the mental states of other species because their brains differ from ours. A key question, of course, is whether their brains differ from ours in ways that matter. In other words, do the brain areas responsible for states of consciousness, such as feelings, differ in humans and other animals?

There is considerable support for the idea that states of consciousness are made possible, at least in part, through the representation of experience in a cognitive workspace involving neocortical areas, especially prefrontal and parietal cortical areas (Crick and Koch, 1990, 2004, Dehaene and Changeux, 2004, Baars, 2005; Frith and Dolan, 1996; Frith et al., 1999; Frith, 2008; Shallice, 1988; Shallice et al., 2008). To the extent that feelings are states of consciousness about emotional situations, they should be represented in these cognitive workspace circuits (LeDoux, 1996, 2002, 2008). The idea proposed here is that conscious feelings result when global organismic states are represented in the cognitive workspace. The basic ingredients of the global organismic state would include information about the stimulus and other aspects of the social and physical environment, the survival circuit the stimulus activates, CNS arousal initiated by the survival circuit, feedback from survival responses that are expressed in the body, and long-term memories (episodic and semantic) about the stimulus and about the resulting state (Figure 4). Thus, in the presence of a survival circuit trigger (a.k.a. an emotional stimulus), the various ingredients would be integrated, and the resulting state categorized by matching the state with long-term memory stores. When this occurs, a conscious feeling of the global organismic state begins to exist. Such

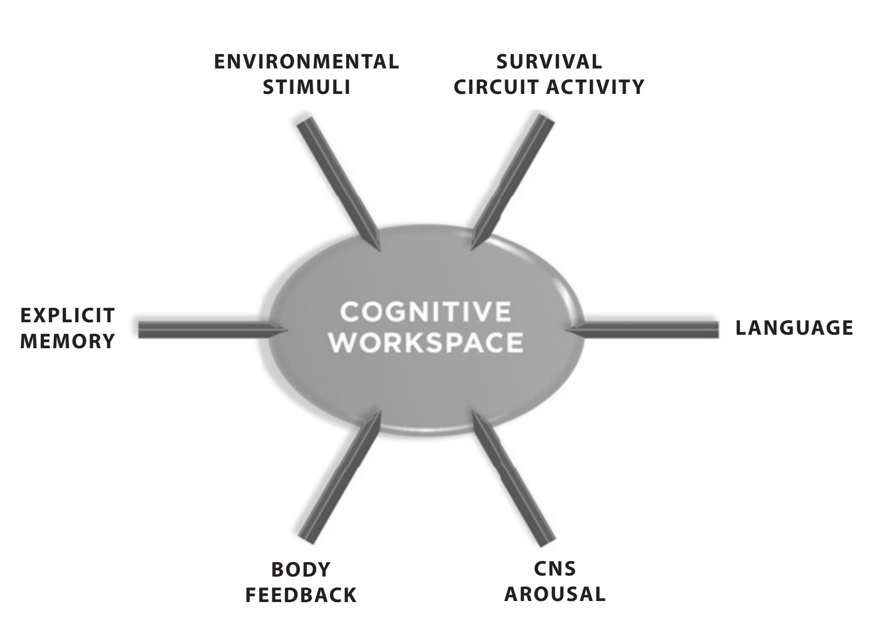


Figure 4. Ingredients of geelings in a cognitive workspace. An emotional feeling is hypothesized to be a representation of a global organismic state initiated by an external stimulus. The representation includes sensory information about the stimulus and the social and physical context, information about the survival circuit that is active, information about CNS arousal, body feedback information, and mnemonic information about the stimulus situation and the state itself. When such a global organismic state is categorized and labeled a conscious feeling of a certain type (e.g. a feeling of fear, pleasure, disgust, etc) results. To the extent that any of these components differ in human and nonhuman species, the nature of the resulting state would differ as well.

a state, having been categorized on the basis of memories of similar states, could be dimensional in nature (just based on arousal and valence) or could take on specific qualities (could be more like what one felt when previously in danger than when frustrated or when enjoying a tasty meal). Labeling of the state with emotion words adds additional specificity to the experience, creating specific feelings (fear, pleasure, disgust, etc).

Dorsolateral prefrontal cortex, a key component of the cognitive workspace, is lacking in most other mammals, and is less developed in nonhuman primates than in humans (Reep, 1984; Braak, 1980; Preuss, 1995; Wise, 2008). In humans, granular prefrontal cortex also has unique cellular features (Semendeferi et al., 2011). Given that feelings are a category of conscious experience, the usual mechanisms of conscious experience should be at work when we have emotional experiences (LeDoux, 1996, 2002, 2008). And given that some of the neural mechanisms involved in conscious representations may be different in humans and other animals, we should be cautious in assuming that the subjectively experienced phenomena that humans label as feelings are experienced by other animals when they engage in behaviors that have some similarity to human emotional behavior. In short, if the circuits that give rise to conscious representations are different in two species, we cannot use behavioral similarity to argue for similarity of conscious feelings functionally. These observations add neurobiological substance to the point famously argued by the philosopher Thomas Nagel. He proposed that only a bat can experience the world like a bat, and only a human can experience the world like a human (Nagel, 1974). We should resist the inclination to apply our introspections to other species.

Also, given that humans are the only organisms with natural language, and language allows a unique mode of information processing in the human brain, we need to be very cautious when we make assumptions about nonhuman behavior regarding processes that language affects. While the idea that language affects thought and conscious experience (Whorf, 1956) was out of favor for a while, it has reemerged as an important principle in recent times (Lakoff, 1987; Lucy, 1997). One way that language is important is that it allows the semantic categorization of experience, including emotional experience. For example, there are more than 30 words in English for gradations of fear (fear, panic, anxiety, worry, trepidation, consternation, etc.) (Marks, 1987). The human brain may be able to categorize emotional states in broad strokes without language but it is unlikely that specific emotions (fear, anger, sadness, joy) could come about without words. Accordingly, lacking language and emotion words, an animal brain cannot partition emotional experience in this way. In short, the language of emotion likely contributes to the experiences one has in emotional situations (Schachter, 1975; Johnson-Laird and Oatley, 1989; Scherer, 1984; Reisenzein, 1995). Indeed, different cultures and their languages express emotions differently (Kitayama and Markus, 1994; Wierzbicka, 1994; Averill, 1980). The dimensional theory of emotion views emotion words as markers in a multidimensional semantic space of feelings (Russell, 1980; Russell and Barrett, 1999). The dimensional theory is incompatible with a basic emotions view, since the latter argues that feelings associated with basic emotions are due to hard-wired circuits, but is compatible with the survival circuit view, which posits indirect and nonobligatory, as opposed to casual, links between survival circuits and feelings.

But the impact of language goes far beyond simple semantics and labeling. We use syntactic processes to evaluate the logical truth of propositional statements. While not all human thought involves propositional statements and logic, syntactic processing provides the human brain and mind with unique features and advantages. Through syntax, the human mind can simulate who will do what to whom in a social situation instantaneously rather than having to learn by trial and error.

So what then might a bat or a rat experience without the kind of cerebral hardware that is characteristic of the human brain? Some have proposed that in addition to full blown feelings that humans talk about, more basic, less differentiated feelings (crude states of positive or negative valence, or maybe even somewhat finer categories based on memory of feelings from the past in similar situations) may exist in other animals. Such states have been called core affects (Panksepp, 1998, 2005; Damasio. 1994, 1999; Barrett et al., 2007; Russell, 2003). While we cannot ask other animals about their feelings, studies of humans can begin to unravel how such states are experienced. Similarity of the structure of these circuits in animals might then provide insight into their function in other animals (Panksepp, 1998, 2005), provided that we do a good job of clarifying the function of the circuits in question (see survival circuit discussion above).

Consciousness and feelings are topics that are best studied in humans. Research on the neural basis of feelings in humans is in its infancy (Panksepp, 1998; 2005; Damasio, 2003; Damasio et al., 2000; Ochsner et al., 2002; Barrett et al., 2007; Rudrauf et al., 2009; Critchley et al., 2004; Pollatos et al., 2007). We will never know what an animal feels. But if we can find neural correlates of conscious feelings in humans (and distinguish them from correlates of unconscious emotional computations in survival circuits), and show that similar correlates exists in homologous brain regions in animals, then some basis for speculating about animal feelings and their nature would exist. While such speculations would be empirically based, they would nevertheless remain speculations.

15. Future directions of research

There are many topics that need further exploration in the study of emotional phenomena in the brain. The following list is meant to point out a few of the many examples and is not meant to be exhaustive.

1. The circuit underlying defense in rodents is fairly well characterized and provides a good starting point for further advancement. An important first step is elucidation of the exact relation between innate and learned defense circuits. Paradigms should be devised that directly compare circuits that are activated by innate and learned cues of the same sensory modality and that elicit similar behavioral defense responses (freezing, escape, attack, etc). Comparisons should proceed in stepwise fashion within a species, with variation in the stimulus and response modalities (though mundane, systematic studies are important).
2. More information is also needed regarding the manner in which external stimuli function as defense triggers, incentives, and reinforcers within defense circuits. Tasks should be developed that can be readily applied across species, but at the same time easily

tailored to species-specific factors. Further exploration of whether there are different circuits for defense in different contexts would also be useful.

3. How do innate and learned stimuli trigger, motivate and reinforce behaviors in nondefense survival circuits? What are the evaluative mechanisms that process such stimuli within specific survival circuits (e.g., energy/nutrition, fluid balance, reproduction, etc)? While incentives and reinforcers have been studied extensively in the context of generic appetitive processes, also of interest is the relation of such stimuli to the specific survival circuit on which they depend.
4. Can survival circuits be further differentiated? For example, to what extent do different forms of defense utilize different circuits? Beyond classifications based on the sensory modality that detects threats and whether the threat is learned or unlearned, are there different circuits for threats related to conspecifics, predators, ingested substances, territory, etc. Similar questions arises for each survival circuit category.
5. When a given brain area is involved in multiple survival functions, an effort should be made to determine the extent to which underlying cellular mechanisms might make distinct contributions. For example, the LA and BA and nucleus accumbens have been implicated in defensive and appetitive behaviors. As noted above, single unit recordings suggest some independence of responses to aversive and appetitive stimuli in these areas. However, questions remain. Do amygdala or accumbens cells simply encode positive and negative valence or are they tied to more specific survival functions? This is readily explored by examining cellular responses to incentive stimuli related different forms of appetitive motivation. For example, do the cells that respond to incentives related to food, drink and reproduction overlap or are they survival-function specific? Single unit recordings are often restricted to one or a small number of brain areas. New molecular imaging techniques are beginning to allow such differences to

be explored at the cellular level across the whole brain. A recent study by Lin et al. (2011) showing distinct populations of cells in the ventromedial hypothalamus that contribute to mating and intruder attack is a prime example. Genetic tools can also be used to provide more detailed information about connectivity, including connectivity at the level of brain areas but also between specific cell types. Such approaches have begun to be used but systematic studies are needed.

6. It is generally assumed that circuits underlying defense, energy, fluid balance, reproduction, thermoregulation, and other survival functions interact, but this has been not been studied to any significant degree. This is a particularly important topic that is best pursued by methods that allow evaluation of concurrent activity in the multiple brain areas, such as fMRI in humans and fMRI and molecular imaging in animals. Studies comparing activity patterns across the whole brain in response processing signals related to various survival circuit functions could provide very important information, especially if animal and human projects use related behavioral paradigms. If homologies are found at the level of brain areas between humans and other mammals, molecular imaging can be used in animal studies to search for unique microcircuits that differentiate between functions and the cellular and synaptic level.
7. Techniques are needed to assess physiological activity at the cellular level across the whole brain and throughout the body (global organismic states) in the presence of biological significant stimuli (triggers of survival circuit activity or motivating incentives,) and during the performance of innate or learned survival responses as well as goal directed responses. It will be especially useful to develop analytic tools that will be able to separate contributions that are survival circuit specific from more general purpose mechanisms, such as nonspecific arousal, and generic aspects of reinforcement and instrumental behavior control.

8. More comparative work is needed to elucidate similarities and differences in survival functions and circuits between various groups of vertebrates. Particularly pressing are studies of non--mammalian vertebrates.
9. What is the relation of survival functions in invertebrates to vertebrates? Are there conserved molecules or genes, or conserved computational principles, that underlie anatomically distinct kinds of circuits but that perform similar survival functions in vertebrates and invertebrates?
10. Explorations of the mechanisms underlying conscious emotional feelings in humans should be pursued more vigorously, including full-blown conscious feelings (feelings of fear, joy, sadness, shame, embarrassment) and coarser conditions (pleasant or unpleasant feelings). This information is important because feelings are such a defining feature of human mental life. But in addition with such information it will then be possible to ask if the required mechanisms of a given kind of conscious state are present in other animals. Even if the mechanisms are present, the results would not allow the conclusion that other animals have feelings that are homologous with human feelings. However, such a result would at least provide a basis for saying whether there is a physical possibility for such states in other animals.

16. Conclusion

The survival circuit concept provides a conceptualization of an important set of phenomena that are often studied under the rubric of emotion—those phenomena that reflect circuits and functions that are conserved across mammals. Included are circuits responsible for defense, energy/nutrition management, fluid balance, thermoregulation, and procreation, among others. With this approach, key phenomena relevant to the topic of emotion can be accounted for without assuming that the phenomena in question are fundamentally the same or even similar to the phenomena people refer to when they use emotion words to characterize subjective emotional feelings (like feeling afraid, angry, or sad). This approach shifts the focus away from questions about whether emotions that humans consciously experience (feel) are also present in other mammals, and toward questions about the extent to which circuits and corresponding functions that are relevant to the field of emotion and that are present in other mammals are also present in humans. And by reassembling ideas about emotion, motivation, reinforcement, and arousal in the context of survival circuits, hypotheses emerge about how organisms negotiate behavioral interactions with the environment in process of dealing with challenges and opportunities in daily life.

References

Adelmann, P.K., and Zajonc, R.B. (1989). Facial efference and the experience of emotion. Annual Review of Psychology *40*, 249–280.

Adolphs, R. (2008). Fear, faces, and the human amygdala. Current Opinion in Neurobiology *18*, 166–172.

Amaral, D.G. (2003). The amygdala, social behavior, and danger detection. Annals of the New York Academy of Sciences *1000*, 337–347.

Antoniadis, E.A., Winslow, J.T., Davis, M., and Amaral, D.G. (2007). Role of the primate amygdala in fear-potentiated startle: effects of chronic lesions in the rhesus monkey. Journal of Neuroscience *27*, 7386–7396.

Arnold, M.B. (1960). Emotion and Personality (New York: Columbia University Press).

Arnsten, A.F., and Li, B.M. (2005). Neurobiology of executive functions: catecholamine influences on prefrontal cortical functions. Biological Psychiatry *57*, 1377–1384.

Aston-Jones, G. (2005). Brain structures and receptors involved in alertness. Sleep Medicine *6* (Suppl 1), S3–S7.

Aston-Jones, G., and Bloom, F.E. (1981). Norepinephrine-containing locus coeruleus neurons in behaving rats exhibit pronounced responses to non-noxious environmental stimuli. Journal of Neuroscience *1*, 887–900.

Aston-Jones, G., Shipley, M.T., Chouvet, G., Ennis, M., van Bockstaele, E., Pieribone, V., Shiekhattar, R., Akaoka, H., Drolet, G., Astier, B., et al. (1991). Afferent regulation of locus coeruleus neurons: anatomy, physiology and pharmacology. Progress in Brain Research *88*, 47–75.

Averill, J. (1980). Emotion and anxiety: Sociocultural, biological, and psychological determinants. In: Explaining Emotions, A.O. Rorty, ed. (Berkeley: University of California Press).

Axelrod, J., and Reisine, T.D. (1984). Stress hormones: their interaction and regulation. Science *224*, 452–459.

Baars, B.J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Progress in Brain Research *150*, 45–53.

Balleine, B.W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology *37*, 407–419.

Balleine, B., and Killcross, S. (1994). Effects of ibotenic acid lesions of the nucleus accumbens on instrumental action. Behavioural Brain Research *65*, 181–193.

Balleine, B.W., and Killcross, S. (2006). Parallel incentive processing: an integrated view of amygdala function. Trends in Neurosciences. *29*, 272–279.

Bargmann, C.I. (2006). Comparative chemosensation from receptors to ecology. Nature *444*, 295–301.

Barrett, L.F. (2006a). Are Emotions Natural Kinds? Perspectives on Psychological Science *1*, 28–58.

Barrett, L.F. (2006b). Solving the emotion paradox: categorization and the experience of emotion. Personality and Social Psychology Review *10*, 20–46.

Barrett, L.F., Lindquist, K.A., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J., and Brennan, L. (2007). Of mice and men: Natural kinds of emotions in the mammalian brain? A response to Panksepp and Izard. Perspectives on Psychological Science *2*, 297–311.

Baxter, M.G., and Murray, E.A. (2002). The amygdala and reward. Nature Reviews Neuroscience *3*, 563–573.

Bekoff, M. (2007). The Emotional Lives of Animals: A Leading Scientist Explores Animal Joy, Sorrow, and Empathy – and Why They Matter (Novato, CA: New World Library).

Belova, M.A., Paton, J.J., Morrison, S.E., and Salzman, C.D. (2007). Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. Neuron *55*, 970–984.

Belova, M.A., Paton, J.J., and Salzman, C.D. (2008). Moment-to-moment tracking of state value in the amygdala. Journal of Neuroscience *28*, 10023–10030.

Bendesky, A., and Bargmann, C.I. (2011). Genetic contributions to behavioural diversity at the gene-environment interface. Nature Reviews Genetics *12*, 809–820.

Bendesky, A., Tsunozaki, M., Rockman, M.V., Kruglyak, L., and Bargmann, C.I. (2011). Catecholamine receptor polymorphisms affect decision-making in C. elegans. Nature *472*, 313–318.

Bernard, C. (1878–1879). Lecons sur les phenomenes de la vie communs aux animaus et aux vegetaux (Paris: Bailliere). Berridge, K.C. (1999). Pleasure, pain, desire, and dread: hidden core processes of emotion. In: Well-Being, D. Kahnemann, E. Diener, and N. Schwarz, eds. (New York: Russell Sage Foundation).

Berridge, K.C. (2004). Motivation concepts in behavioral neuroscience. Physiology & Behavior *81*, 179–209.

Berridge, K.C. (2007). The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology (Berl.) *191*, 391–431.

Berridge, K.C. (2009). 'Liking' and 'wanting' food rewards: brain substrates and roles in eating disorders. Physiology & Behavior *97*, 537–550.

Berridge, K.C., and Robinson, T.E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Research Reviews *28*, 309–369.

Bindra, D. (1968). Neuropsychological interpretation of the effects of drive and incentive-motivation on general activity and instrumental behavior. Psychological Review *75*, 1–22.

Bindra, D. (1969). A unfied interpretation of emotion and motivation. Annals of the New York Academy of Sciences *159*, 1071–1083.

Blanchard, D.C., and Blanchard, R.J. (1972). Innate and conditioned reactions to threat in rats with amygdaloid lesions. Journal of Comparative and Physiological Psychology *81*, 281–290.

Blanchard, D.C., Blanchard, R.J., and Rodgers, R.J. (1990). Pharmacological and neural control of anti-predator defense in the rat. Aggressive Behavior *16*, 165–175.

Blanchard, D.C., Li, C.I., Hubbard, D., Markham, C.M., Yang, M., Takahashi, L.K., and Blanchard, R.J. (2003). Dorsal premammillary nucleus differentially modulates defensive behaviors induced by different threat stimuli in rats. Neuroscience Letters *345*, 145–148.

Blaustein, J.D. (2008). Neuroendocrine regulation of feminine sexual behavior: lessons from rodent models and thoughts about humans. Annual Review of Psychology *59*, 93–118.

Bolles, R.C., and Fanselow, M.S. (1980). A perceptual-defensive-recuperative model of fear and pain. Behavioral and Brain Sciences *3*, 291–323.

Bordi, F., and LeDoux, J. (1992). Sensory tuning beyond the sensory system: an initial analysis of auditory response properties of neurons in the lateral amygdaloid nucleus and overlying areas of the striatum. Journal of Neuroscience *12*, 2493–2503.

Bowlby, J. (1969). Attachment, Vol. 1: Attachment and Loss (New York: Basic Books).

Braak, E. (1980). On the structure of IIIab-pyramidal cells in the human isocortex. A Golgi and electron microscopical study with special emphasis on the proximal axon segment. Journal für Hirnforschung *21*, 437–442.

Brown, S.M., and Hariri, A.R. (2006). Neuroimaging studies of serotonin gene polymorphisms: exploring the interplay of genes, brain, and behavior. Cognitive, Affective, & Behavioral Neuroscience *6*, 44–52.

Bruce, L.L., and Neary, T.J. (1995). The limbic system of tetrapods: a comparative analysis of cortical and amygdalar populations. Brain, Behavior and Evolution *46*, 224–234.

Büchel, C., and Dolan, R.J. (2000). Classical fear conditioning in functional neuroimaging. Current Opinion in Neurobiology *10*, 219–223.

Buck, R. (1980). Nonverbal behavior and the theory of emotion: the facial feedback hypothesis. Journal of Personality and Social Psychology *38*, 811–824.

Butler, A.B., and Hodos, W. (2005). Comparative Vertebrate Neuroanatomy: Evolution and Adaptation, Second Edition (Hoboken: John Wiley and Sons, Inc.).

Cain, C.K., and LeDoux, J.E. (2007). Escape from fear: a detailed behavioral analysis of two atypical responses reinforced by CS termination. Journal of Experimental Psychology: Animal Behavior Processes *33*, 451–463.

Canli, T., Sivers, H., Whitfield, S.L., Gotlib, I.H., and Gabrieli, J.D. (2002). Amygdala response to happy faces as a function of extraversion. Science *296*, 2191.

Cannon, W.B. (1927). The James-Lange theory of emotions: a critical examination and an alternative theory. American Journal of Psychology *39*, 106–124.

Cannon, W.B. (1929). Bodily Changes in Pain, Hunger, Fear, and Rage (New York: Appleton).

Cannon, W.B. (1931). Again the James-Lange and the thalamic theories of emotion. Psychological Review *38*, 281–295.

Canteras, N.S. (2002). The medial hypothalamic defensive system: hodological organization and functional implications. Pharmacology Biochemistry & Behavior *71*, 481–491.

Canteras, N.S., Resstel, L.B., Bertoglio, L.J., Carobrez, Ade.P., and Guimarães, F.S. (2010). Neuroanatomy of anxiety. Current Topics in Behavioral Neurosciences *2*, 77–96.

Cardinal, R.N., Parkinson, J.A., Hall, J., and Everitt, B.J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neuroscience & Biobehavioral Reviews *26*, 321–352.

Carew, T.J., and Sutton, M.A. (2001). Molecular stepping stones in memory consolidation. Nature Neuroscience *4*, 769–771.

Casey, B.J., Ruberry, E.J., Libby, V., Glatt, C.E., Hare, T., Soliman, F., Duhoux, S., Frielingsdorf, H., and Tottenham, N. (2011). Transitional and translational studies of risk for anxiety. Depression and Anxiety *28*, 18–28.

Chi, C.C., and Flynn, J.P. (1971). Neural pathways associated with hypothalamically elicited attach behavior in cats. Science *171*, 703–706.

Choi, J.S., and Brown, T.H. (2003). Central amygdala lesions block ultrasonic vocalization and freezing as conditional but not unconditional responses. Journal of Neuroscience *23*, 8713–8721.

Choi, G.B., Dong, H.W., Murphy, A.J., Valenzuela, D.M., Yancopoulos, G.D., Swanson, L.W., and Anderson, D.J. (2005). Lhx6 delineates a pathway mediating innate reproductive behaviors from the amygdala to the hypothalamus. Neuron *46*, 647–660.

Cofer, C.N. (1972). Motivation and Emotion (Glenview, IL: Scott, Foresman).

Cofer, C., and Appley, M. (1964). Motivation, Theory and Research (New York: John Wiley and Sons, Inc.).

Cohen, D.H. (1974). The neural pathways and informational flow mediating a conditioned autonomic response. In: Limbic and Autonomic Nervous System Research, L.V. Di Cara, ed. (New York: Plenum Press), pp. 223–275.

Conover, K.L., and Shizgal, P. (1994). Differential effects of postingestive feedback on the reward value of sucrose and lateral hypothalamic stimulation in rats. Behavioral Neuroscience 108, 559–572.

Coons, E.E., and White, H.A. (1977). Tonic properties of orosensation and the modulation of intracranial self-stimulation: the CNS weighting of external and internal factors governing reward. Annals of the New York Academy of Sciences *290*, 158–179.

Corbit, L.H., and Balleine, B.W. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of pavlovian-instrumental transfer. Journal of Neuroscience *25*, 962–970.

Cousens, G., and Otto, T. (1998). Both pre- and posttraining excitotoxic lesions of the basolateral amygdala abolish the expression of olfactory and contextual fear conditioning. Behavioral Neuroscience *112*, 1092–1103.

Crick, F., and Koch, C. (1990). Toward a neurobiological theory of consciousness. Neurosciences *2*, 263–275.

Crick, F.C., and Koch, C. (2004). A framework for consciousness. In: The Cognitive Neurosciences, Third Edition, M.S. Gazzaniga, ed. (Cambridge: MIT Press), pp. 1133–1143.

Critchley, H.D., Wiens, S., Rotshtein, P., Ohman, A., and Dolan, R.J. (2004). Neural systems supporting interoceptive awareness. Nature Neuroscience *7*, 189–195.

Damasio, A. (1994). Descarte's Error: Emotion, Reason, and the Human Brain (New York: Gosset/Putnam).

Damasio, A.R. (1999). The Feeling of what Happens: Body and Emotion in the Making of Consciousness (New York: Harcourt Brace).

Damasio, A. (2003). Feelings of emotion and the self. Annals of the New York Academy of Sciences *1001*, 253–261.

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

Darwin, C. (1872). The Expression of the Emotions in Man and Animals (London: Fontana Press).

Davidson, R.J. (1993). Cerebral asymmetry and emotion: Conceptual and methodological conundrums. Cognition and Emotion *7*, 115–138.

Davies, D.C., Martínez-García, F., Lanuza, E., and Novejarque, A. (2002). Striato-amygdaloid transition area lesions reduce the duration of tonic immobility in the lizard Podarcis hispanica. Brain Research Bulletin *57*, 537–541.

Davis, M. (1992). The role of the amygdala in conditioned fear. In: The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction, J.P. Aggleton, ed. (NY: Wiley-Liss, Inc.), pp. 255–306.

Davis, M., and Whalen, P.J. (2001). The amygdala: vigilance and emotion. Molecular Psychiatry *6*, 13–34.

Davis, M., Walker, D.L., and Lee, Y. (1997). Amygdala and bed nucleus of the stria terminalis: differential roles in fear and anxiety measured with the acoustic startle reflex. Philosophical Transactions of the Royal Society B: Biological Sciences *352*, 1675–1687.

Davis, F.C., Somerville, L.H., Ruberry, E.J., Berry, A.B., Shin, L.M., and Whalen, P.J. (2011). A tale of two negatives: differential memory modulation by threat-related facial expressions. Emotion *11*, 647–655.

de Oca, B.M., and Fanselow, M.S. (2004). Amygdala and periaqueductal gray lesions only partially attenuate unconditional defensive responses in rats exposed to a cat. Integrative Physiological and Behavioral Science *39*, 318–333.

de Oca, B.M., DeCola, J.P., Maren, S., and Fanselow, M.S. (1998). Distinct regions of the periaqueductal gray are involved in the acquisition and expression of defensive responses. Journal of Neuroscience *18*, 3426–3432.

de Oca, B.M., Minor, T.R., and Fanselow, M.S. (2007). Brief flight to a familiar enclosure in response to a conditional stimulus in rats. Journal of General Psychology *134*, 153–172.

Dehaene, S., and Changeux, J.-P. (2004). Neural mechanisms for access to consciousness. In: The Cognitive Neurosciences, Third Edition, M.S. Gazzaniga, ed. (Cambridge: MIT Press), pp. 1145–1157.

Dickson, B.J. (2008). Wired for sex: the neurobiology of Drosophila mating decisions. Science *322*, 904–909.

Dielenberg, R.A., Hunt, G.E., and McGregor, I.S. (2001). “When a rat smells a cat”: the distribution of Fos immunoreactivity in rat brain following exposure to a predatory odor. Neuroscience *104*, 1085–1097.

Dixon, B.A. (2001). Animal emotion. Ethics and the Environment *6*, 22–30.

Dudai, Y. (1988). Neurogenetic dissection of learning and short-term memory in Drosophila. Annual Review of Neuroscience *11*, 537–563.

Duffy, E. (1934). Emotion: an example of the need for reorientation in psychology. Psychological Review *41*, 184–198.

Duffy, E. (1941). An explanation of “emotional” phemomena without the use of the concept “emotion”. Journal of General Psychology *25*, 283–293.

Ekman, P. (1972) Universals and cultural differences in facial expressions of emotions. In: Nebraska Symposium on Motivation, 1971 (Cole J, ed), pp 207–283. Lincoln, Nebraska: University of Nebraska Press.

Ekman, P. (1980). Biological and cultural contributions to body and facial movement in the expression of emotions. In: Explaining Emotions, A.O. Rorty, ed. (Berkeley: University of California Press).

Ekman, P. (1984). Expression and nature of emotion. In: Approaches to Emotion, K. Scherer and P. Ekman, eds. (Hillsdale, NJ: Earlbaum), pp. 319–343.

Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion *6*, 169–200.

Ekman, P. (1999a). Basic emotions. In: Handbook of Cognition and Emotion, T. Dalgleish and M. Power, eds. (Chichester: John Wiley and Sons, Co.), pp. 45–60.

Ekman, P. (1999b). Annotated update of Charles Darwin's "The Expression of the Emotions in Man and Animals" (New York: Harpercollins).

Ekman, P., and Davidson, R. (1994). The nature of emotion: fundamental questions (New York: Oxford University Press).

Elliot, A., and Church, M. (1997). A hierarchical model of approach motivation and avoidance achievement motivation. Journal of Personality and Social Psychology *72*, 218–232.

Ellsworth, P.C. (1994). William James and emotion: is a century of fame worth a century of misunderstanding? Psychological Review *101*, 222–229.

Elmquist, J.K., Coppari, R., Balthasar, N., Ichinose, M., and Lowell, B.B. (2005). Identifying hypothalamic pathways controlling food intake, body weight, and glucose homeostasis. Journal of Comparative Neurology *493*, 63–71.

Estes, W.K., and Skinner, B.F. (1941). Some quantitative properties of anxiety. Journal of Experimental Psychology *29*, 390–400.

Everitt, B.J., and Robbins, T.W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nature Neuroscience *8*, 1481–1489.

Everitt, B.J., Parkinson, J.A., Olmstead, M.C., Arroyo, M., Robledo, P., and Robbins, T.W. (1999). Associative processes in addiction and reward. The role of amygdala-ventral striatal subsystems. Annals of the New York Academy of Sciences *877*, 412–438.

Everitt, B.J., Cardinal, R.N., Parkinson, J.A., and Robbins, T.W. (2003). Appetitive behavior: impact of amygdala-dependent mechanisms of emotional learning. Annals of the New York Academy of Sciences *985*, 233–250.

Fanselow, M.S. (1994). Neural organization of the defensive behavior system responsible for fear. Psychonomic Bulletin and Review *1*, 429–438.

Fanselow, M.S., and Poulos, A.M. (2005). The neuroscience of mammalian associative learning. Annual Review of Psychology *56*, 207–234.

Ferris, C.F., Stolberg, T., Kulkarni, P., Murugavel, M., Blanchard, R., Blanchard, D.C., Febo, M., Brevard, M., and Simon, N.G. (2008). Imaging the neural circuitry and chemical control of aggressive motivation. BMC Neuroscience *9*, 111.

Fitzsimons, J.T. (1979). The physiology of thirst and sodium appetite. Monographs of the Physiological Society *35*, 1–572.

Flynn, J.P. (1967). The neural basis of aggression in cats. In: Biology and Behavior: Neurophysiology and Emotion, D.G. Glass, ed. (New York: Rockefeller University Press and Russell Sage Foundation), pp. 40–60.

Foote, S.L., Bloom, F.E., and Aston-Jones, G. (1983). Nucleus locus ceruleus: New evidence of anatomical and physiological specificity. Physiological Reviews *63*, 844–914.

Foote, S.L., Berridge, C.W., Adams, L.M., and Pineda, J.A. (1991). Electrophysiological evidence for the involvement of the locus coeruleus in alerting, orienting, and attending. Progress in Brain Research *88*, 521–532.

Frijda, N. (1986). The Emotions (Cambridge: Cambridge University Press).
Frith, C.D. (2008). The social functions of consciousness. In: Frontiers of Consciousness: Chichele lectures, L. Weiskrantz and M. Davies, eds. (Oxford: Oxford University Press), pp. 225–244.

Frith, C., and Dolan, R. (1996). The role of the prefrontal cortex in higher cognitive functions. Brain Research: Cognitive Brain Research *5*, 175–181.

Frith, C., Perry, R., and Lumer, E. (1999). The neural correlates of conscious experience: An experimental framework. Trends in Cognitive Sciences (Regul. Ed.) *3*, 105–114.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., Benedetti, F., Abbamonte, M., Gasparotti, R., Barale, F., et al. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. Journal of Psychiatry & Neuroscience *34*, 418–432.

Gallagher, M., and Holland, P.C. (1994). The amygdala complex: multiple roles in associative learning and attention. Proceedings of the National Academy of Sciences USA *91*, 11771–11776.

Gallistel, R. (1980). The Organization of Action: A New Synthesis (Hillsdale, NJ: Erlbaum).

Garcia, J., McGowan, B.K., Ervin, F.R., and Koelling, R.A. (1968). Cues: their relative effectiveness as a function of the reinforcer. Science *160*, 794–795.

Garrity, P.A., Goodman, M.B., Samuel, A.D., and Sengupta, P. (2010). Running hot and cold: behavioral strategies, neural circuits, and the molecular machinery for thermotaxis in C. elegans and Drosophila. Genes & Development *24*, 2365–2382.

Glanzman, D.L. (2010). Common mechanisms of synaptic plasticity in vertebrates and invertebrates. Current Biology *20*, R31–R36.

Glickman, S.E., and Schiff, B.B. (1967). A biological theory of reinforcement. Psychological Review *74*, 81–109.

Glimcher, P.W. (2003). Decisions, Uncertainty, and the Brain: The science of neuroeconomics (Cambridge: MIT Press). Glimcher, P.W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proceedings of the National Academy of Sciences USA *108* (*Suppl 3*), 15647–15654.

Gray, J.A. (1982). The Neuropsychology of Anxiety (New York: Oxford University Press).

Gray, J.A. (1987). The Psychology of Fear and Stress (New York: Cambridge University Press).

Gray, T.S. (1993). Amygdaloid CRF pathways. Role in autonomic, neuroendocrine, and behavioral responses to stress. Annals of the New York Academy of Sciences *697*, 53–60.

Graybiel, A. (1976). Input-Output Anatomy of the Basal Ganglia (Toronto, Canada: Lecture at the Society for Neuroscience).

Griffiths, P.E. (1997). What Emotions Really Are: The Problem of Psychological Categories (Chicago: University of Chicago Press).

Griffiths, P.E. (2004). Is emotion a natural kind? In: Thinking about Feeling: Contemporary Philosophers on Emotions, R.C. Solomon, ed. (Oxford: Oxford University Press), pp. 233–249.

Hamann, S.B., Ely, T.D., Hoffman, J.M., and Kilts, C.D. (2002). Ecstasy and agony: activation of the human amygdala in positive and negative emotion. Psychological Science *13*, 135–141.

Hammond, L.J. (1970). Conditional emotional states. In: Physiological Correlates of Emotion, P. Black, ed. (New York: Academic Press).

Hasselmo, M.E. (1995). Neuromodulation and cortical function: modeling the physiological basis of behavior. Behavioural Brain Research *67*, 1–27.

Haubensak, W., Kunwar, P.S., Cai, H., Ciocchi, S., Wall, N.R., Ponnusamy, R., Biag, J., Dong, H.W., Deisseroth, K., Callaway, E.M., et al. (2010). Genetic dissection of an amygdala microcircuit that gates conditioned fear. Nature *468*, 270–276.

Hawkins, R.D., Kandel, E.R., and Bailey, C.H. (2006). Molecular mechanisms of memory storage in Aplysia. Biological Bulletin *210*, 174–191.

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

Hess, W.R. (1954). Functional Organization of the Diencephalon (New York: Grune and Stratton).

Holland, P.C., and Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy. Current Opinion in Neurobiology *14*, 148–155.

Holland, P.C., and Rescorla, R.A. (1975). Second-order conditioning with food unconditioned stimulus. Journal of Comparative and Physiological Psychology *88*, 459–467.

Hull, C.L. (1943). Principles of Behavior (New York: Appleton-Century-Crofts).

Hurley, L.M., Devilbiss, D.M., and Waterhouse, B.D. (2004). A matter of focus: monoaminergic modulation of stimulus coding in mammalian sensory networks. Current Opinion in Neurobiology *14*, 488–495.

Hyman, S.E., Malenka, R.C., and Nestler, E.J. (2006). Neural mechanisms of addiction: the role of reward-related learning and memory. Annual Review of Neuroscience *29*, 565–598.

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

Izard, C.E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. Psychological Review *99*, 561–565.

Izard, C.E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. Perspectives on Psychological Science *2*, 260–280.

Jacobs, B.L., Wilkinson, L.O., and Fornal, C.A. (1990). The role of brain serotonin. A neurophysiologic perspective. Neuropsychopharmacology *3*, 473–479.

James, W. (1884). What is an emotion? Mind *9*, 188–205.

Jarymowicz, M. (2009). Emotions as evaluative processes from primary affects to appraisals based on deliberative thinking. In: Closer to Emotions, A. Blachnio and A. Przepiorka, eds. (Lublin: Wydawnictwo KUL), pp. 55–72.

Johansen, J.P., Cain, C.K., Ostroff, L.E., and LeDoux, J.E. (2011). Molecular mechanisms of fear learning and memory. Cell *147*, 509–524.

Johnson, A.K. (2007). The sensory psychobiology of thirst and salt appetite. Medicine and Science in Sports and Exercise *39*, 1388–1400.

Johnson, A.W., Gallagher, M., and Holland, P.C. (2009). The basolateral amygdala is critical to the expression of pavlovian and instrumental outcome-specific reinforcer devaluation effects. Journal of Neuroscience *29*, 696–704.

Johnson-Laird, P.N., and Oatley, K. (1989). The language of emotions: An analysis of a semantic field. Cognition and Emotion *3*, 81–123.

Jones, B.E. (2003). Arousal systems. Frontiers in Bioscience *8*, s438–s451.

Jouvet, M. (1969). Biogenic amines and the states of sleep. Science *163*, 32–41.

Jouvet, M. (1999). Sleep and serotonin: an unfinished story. Neuropsychopharmacology *21* (2, Suppl), 24S–27S.

Kagan, J. (2007). What is emotion? History, Measures, and Meanings (New Haven: Yale University Press).

Kalin, N.H., Shelton, S.E., Davidson, R.J., and Kelley, A.E. (2001). The primate amygdala mediates acute fear but not the behavioral and physiological components of anxious temperament. Journal of Neuroscience *21*, 2067–2074.

Kalin, N.H., Shelton, S.E., and Davidson, R.J. (2004). The role of the central nucleus of the amygdala in mediating fear and anxiety in the primate. Journal of Neuroscience *24*, 5506–5515.

Kandel, E.R. (2001). The molecular biology of memory storage: a dialog between genes and synapses. Bioscience Reports *21*, 565–611.

Kapp, B.S., Whalen, P.J., Supple, W.F., and Pascoe, J.P. (1992). Amygdaloid contributions to conditioned arousal and sensory information processing. In: The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction, J.P. Aggleton, ed. (New York: Wiley-Liss), pp. 229–254.

Kapp, B.S., Supple, W.F., Jr., and Whalen, P.J. (1994). Effects of electrical stimulation of the amygdaloid central nucleus on neocortical arousal in the rabbit. Behavioral Neuroscience *108*, 81–93.

Kelley, A.E. (2004). Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. Neuroscience & Biobehavioral Reviews *27*, 765–776.

Killcross, A.S., and Robbins, T.W. (1993). Differential effects of intra-accumbens and systemic amphetamine on latent inhibition using an on-baseline, within-subject conditioned suppression paradigm. Psychopharmacology (Berl.) *110*, 479–489.

Kitayama S. and Markus H.R., eds. (1994). Emotion and Culture: Empirical Studies of Mutual Influence (Washington, DC: American Psychological Association).

Koob, G.F. (2009). Brain stress systems in the amygdala and addiction. Brain Research *1293*, 61–75.

LaBar, K.S., and Cabeza, R. (2006). Cognitive neuroscience of emotional memory. Nature Reviews Neuroscience *7*, 54–64.

Lakoff, G. (1987). Women, Fire, and Dangerous Things: What Categories Reveal about the mind (Chicago: The University of Chicago Press).

Lamprecht, R., and Dudai, Y. (2000). The amygdala in conditioned taste aversion: it's there, but where? In: The Amygdala, J. Aggleton, ed. (Oxford: Oxford U. Press).

Lane, R.D., Chua, P.M., and Dolan, R.J. (1999). Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. Neuropsychologia *37*, 989–997.

Lang, P.J. (1994). The varieties of emotional experience: a meditation on James-Lange theory. Psychological Review *101*, 211–221.

Lang, P.J., and Davis, M. (2006). Emotion, motivation, and the brain: reflex foundations in animal and human research. Progress in Brain Research *156*, 3–29.

Lang, P.J., Bradley, M.M., and Cuthbert, B.N. (1990). Emotion, attention, and the startle reflex. Psychological Review *97*, 377–395.

Lange, C.G. (1885/1922). The emotions. In: The Emotions, E. Dunlap, ed. (Baltimore: Williams and Wilkins).

Lashley, K.S. (1938). Experimental analysis of instinctive behavior. Psychological Review 45, 445–471.

Lázaro-Muñoz, G., LeDoux, J.E., and Cain, C.K. (2010). Sidman instrumental avoidance initially depends on lateral and basal amygdala and is constrained by central amygdala-mediated Pavlovian processes. Biological Psychiatry *67*, 1120–1127.

Lazarus, R.S. (1981). A cognitivist's reply to Zajonc on emotion and cognition. American Psychologist *36*, 222–223.

Lazarus, R.S. (1991a). Cognition and motivation in emotion. American Psychologist *46*, 352–367.

Lazarus, R.S. (1991b). Emotion and Adaptation (New York: Oxford University Press).

Lazarus, R.S., and Folkman, S. (1984). Stress, Appraisal and Coping (New York: Springer).

Lebestky, T., Chang, J.S., Dankert, H., Zelnik, L., Kim, Y.C., Han, K.A., Wolf, F.W., Perona, P., and Anderson, D.J. (2009). Two different forms of arousal in Drosophila are oppositely regulated by the dopamine D1 receptor ortholog DopR via distinct neural circuits. Neuron *64*, 522–536.

LeDoux, J.E. (1992). Emotion and the amygdala. In: The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction, J.P. Aggleton, ed. (New York: Wiley-Liss, Inc.), pp. 339–351.

LeDoux, J.E. (1995). Seting "Stress" into Motion. In: Neurobiological and Clinical Consequences of Stress: From Normal Adaptation to PTSD, M.J. Friedman, D.S. Charney, and A.Y. Deutch, eds. (New York: Raven Press), pp. 125–134.

LeDoux, J.E. (1996). The Emotional Brain (New York: Simon and Schuster).

LeDoux, J.E. (2000). Emotion circuits in the brain. Annual Review of Neuroscience *23*, 155–184.

LeDoux, J.E. (2002). Synaptic Self: How Our Brains Become Who We Are (New York: Viking).

LeDoux, J.E. (2008). Emotional colouration of consciousness: how feelings come about. In: Frontiers of Consciousness: Chichele Lectures, L. Weiskrantz and M. Davies, eds. (Oxford: Oxford University Press), pp. 69–130.

LeDoux, J.E. (2012). Evolution of human emotion: A view through fear. Progress in Brain Research *195*, 431–442.

Leventhal, H., and Scherer, K. (1987). The relationship of emotion to cognition: a functional approach to a semantic controversy. Cognition and Emotion *1*, 3–28.

Liddell, B.J., Brown, K.J., Kemp, A.H., Barton, M.J., Das, P., Peduto, A., Gordon, E., and Williams, L.M. (2005). A direct brainstem-amygdala-cortical 'alarm' system for subliminal signals of fear. Neuroimage *24*, 235–243.

Lima, S.L., and Dill, L.M. (1990). Behavioral decision made under the risk of predation: a review and prospectus. Canadian Journal of Zoology *68*, 619–640.

Lin, D., Boyle, M.P., Dollar, P., Lee, H., Lein, E.S., Perona, P., and Anderson, D.J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. Nature *470*, 221–226.

Lindsley, D.B. (1951). Emotions. In: Handbook of Experimental Psychology, S.S. Stevens, ed. (New York: Wiley), pp. 473–516.

Litvin, Y., Blanchard, D.C., and Blanchard, R.J. (2007). Rat 22kHz ultrasonic vocalizations as alarm cries. Behavioural Brain Research *182*, 166–172.

Loewy, A.D. (1991). Forebrain nuclei involved in autonomic control. Progress in Brain Research *87*, 253–268.

Lopez, H.S., and Brown, A.M. (1992). Neuromodulation. Current Opinion in Neurobiology 2, 317–322.

Lorenz, K.Z. (1981). The Foundations of Ethology: The Principal Ideas and Discoveries in Animal Behavior (New York: Simon and Schuster).

Lucy, J.A. (1997). Linguistic Relativity. Annual Review of Anthropology *26*, 291–312.

Luo, Q., Holroyd, T., Jones, M., Hendler, T., and Blair, J. (2007). Neural dynamics for facial threat processing as revealed by gamma band synchronization using MEG. Neuroimage *34*, 839–847.

Lyons, W. (1980). Emotion (Cambridge: Cambridge University Press). MacLean, P.D. (1949). Psychosomatic disease and the visceral brain; recent developments bearing on the Papez theory of emotion. Psychosomatic Medicine *11*, 338–353.

MacLean, P.D. (1952). Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). Electroencephalography and Clinical Neurophysiology *4*, 407–418.

Macnab, R.M., and Koshland, D.E., Jr. (1972). The gradient-sensing mechanism in bacterial chemotaxis. Proceedings of the National Academy of Sciences USA *69*, 2509–2512.

Mandler, G. (1975). Mind and Emotion (New York: Wiley).

Manning, A. (1967). An introduction to Animal Behaviour (Reading, MA: Addison-Wesley).

Maren, S. (2001). Neurobiology of Pavlovian fear conditioning. Annual Review of Neuroscience *24*, 897–931.

Maren, S., and Quirk, G.J. (2004). Neuronal signalling of fear memory. Nature Reviews Neuroscience *5*, 844–852.

Margulies, C., Tully, T., and Dubnau, J. (2005). Deconstructing memory in Drosophila. Current Biology *15*, R700–R713.

Markham, C.M., Blanchard, D.C., Canteras, N.S., Cuyno, C.D., and Blanchard, R.J. (2004). Modulation of predatory odor processing following lesions to the dorsal premammillary nucleus. Neuroscience Letters *372*, 22–26.

Marks, I. (1987). Fears, Phobias, and Rituals: Panic, Anxiety and Their Disorders (New York: Oxford University Press).

Martínez-García, F., Martínez-Marcos, A., and Lanuza, E. (2002). The pallial amygdala of amniote vertebrates: evolution of the concept, evolution of the structure. Brain Research Bulletin *57*, 463–469.

Masson, J.M., and McCarthy, S. (1996). When Elephants Weep: The Emotional Lives of Animals (New York: Delacorte).

Maturana, H., and Varela, F. (1987). The Tree of Knowledge: The Biological Roots of Human Understanding (Boston: New Science Library).

McClelland, D., Atkinson, J., Clark, R., and Lowell, E. (1953). The Achievement Motive (New York: Irvington Publishers).

McEwen, B.S. (2009). The brain is the central organ of stress and adaptation. Neuroimage *47*, 911–913.

McEwen, B.S., Eiland, L., Hunter, R.G., and Miller, M.M. (2012). Stress and anxiety: structural plasticity and epigenetic regulation as a consequence of stress. Neuropharmacology *62*, 3–12.

McGaugh, J.L. (2000). Memory—a century of consolidation. Science *287*, 248–251.

McGaugh, J.L. (2003). Memory and Emotion: The Making of Lasting Memories (London: The Orion Publishing Group).

McGrath, P.T., Rockman, M.V., Zimmer, M., Jang, H., Macosko, E.Z., Kruglyak, L., and Bargmann, C.I. (2009). Quantitative mapping of a digenic behavioral trait implicates globin variation in C. elegans sensory behaviors. Neuron *61*, 692–699.

McNaughton, N. (1989). Biology and Emotion (Cambridge: Cambridge University Press).

Miller, N. (1944). Experimental studies of conflict. In: Personality and the Behavioral Disorders, N.S. Endler and J.M. Hunt, eds. (New York: Ronald Press), pp. 431–465.

Mineka, S., and Öhman, A. (2002). Phobias and preparedness: the selective, automatic, and encapsulated nature of fear. Biological Psychiatry *52*, 927–937.

Mobbs, D., Marchant, J.L., Hassabis, D., Seymour, B., Tan, G., Gray, M., Petrovic, P., Dolan, R.J., and Frith, C.D. (2009). From threat to fear: the neural organization of defensive fear systems in humans. Journal of Neuroscience *29*, 12236–12243.

Mogenson, G.J., Jones, D.L., and Yim, C.Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. Progress in Neurobiology *14*, 69–97.

Monsey, M.S., Ota, K.T., Akingbade, I.F., Hong, E.S., and Schafe, G.E. (2011). Epigenetic alterations are critical for fear memory consolidation and synaptic plasticity in the lateral amygdala. PLoS ONE 6, e19958.

Monti, J.M., and Jantos, H. (2008). The roles of dopamine and serotonin, and of their receptors, in regulating sleep and waking. Progress in Brain Research *172*, 625–646.

Morgan, C.T. (1943). Physiological Psychology (New York: McGraw-Hill).

Morris, J.S., Ohman, A., and Dolan, R.J. (1999). A subcortical pathway to the right amygdala mediating “unseen” fear. Proceedings of the National Academy of Sciences USA *96*, 1680–1685.

Morrison, S.E., and Salzman, C.D. (2010). Re-valuing the amygdala. Current Opinion in Neurobiology *20*, 221–230.

Morton, G.J., Cummings, D.E., Baskin, D.G., Barsh, G.S., and Schwartz, M.W. (2006). Central nervous system control of food intake and body weight. Nature *443*, 289–295.

Moruzzi, G., and Magoun, H.W. (1949). Brain stem reticular formation and activation of the EEG. Electroencephalography and Clinical Neurophysiology *1*, 455–473.

Moscarello, J.M., Ben-Shahar, O., and Ettenberg, A. (2009). Effects of food deprivation on goal-directed behavior, spontaneous locomotion, and c-Fos immunoreactivity in the amygdala. Behavioural Brain Research *197*, 9–15.

Motta, S.C., Goto, M., Gouveia, F.V., Baldo, M.V., Canteras, N.S., and Swanson, L.W. (2009). Dissecting the brain’s fear system reveals the hypothalamus is critical for responding in subordinate conspecific intruders. Proceedings of the National Academy of Sciences USA *106*, 4870–4875.

Mowrer, O.H. (1960). Learning Theory and Behavior (New York: Wiley).

Moyer, K.E. (1976). The Psychobiology of Aggression (New York: Harper and Row).

Mozzachiodi, R., and Byrne, J.H. (2010). More than synaptic plasticity: role of nonsynaptic plasticity in learning and memory. Trends in Neurosciences. *33*, 17–26.

Nagel, T. (1974). What is it like to be a bat? Philosophy Review *83*, 4435–4450.

Nakamura, K., and Morrison, S.F. (2007). Central efferent pathways mediating skin cooling-evoked sympathetic thermogenesis in brown adipose tissue. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology *292*, R127–R136.

Nauta, W.J.H., and Karten, H.J. (1970). A general profile of the vertebrate brain, with sidelights on the ancestry of cerebral cortex. In: The Neurosciences: Second Study Program, F.O. Schmitt, ed. (New York: The Rockefeller University Press), pp. 7–26.

Nesse, R. (1990). Evolutionary explanations of emotions. Human Nature *1*, 261–289.

Nestler, E.J. (2004). Molecular mechanisms of drug addiction. Neuropharmacology *47* (*Suppl 1*), 24–32.

Nieuwenhuys, R. (1985). Chemoarchitecture of the Brain (Berlin: Springer-Verlag). Nishino, S. (2011). Hypothalamus, hypocretins/orexin, and vigilance control. Handbook of Clinical Neurology *99*, 765–782.

Northcutt, R.G., and Kaas, J.H. (1995). The emergence and evolution of mammalian neocortex. Trends in Neurosciences. *18*, 373–379.

Ochsner, K.N., and Gross, J.J. (2005). The cognitive control of emotion. Trends in Cognitive Sciences (Regul. Ed.) *9*, 242–249.

Ochsner, K.N., Bunge, S.A., Gross, J.J., and Gabrieli, J.D. (2002). Rethinking feelings: an FMRI study of the cognitive regulation of emotion. Journal of Cognitive Neuroscience *14*, 1215–1229.

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. (2009). Of snakes and faces: an evolutionary perspective on the psychology of fear. Scandinavian Journal of Psychology *50*, 543–552.

Olds, J. (1977). Drives and Reinforcement (New York: Raven).

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

Ono, T., and Nishijo, H. (1992). Neurophysiological basis of the Kluver-Bucy Syndrome: Responses of monkey amygdaloid neurons to biologically significant objects. In: The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction, J.P. Aggleton, ed. (New York: Wiley- Liss, Inc.), pp. 167–190.

Oomura, Y., Aou, S., Koyama, Y., Fujita, I., and Yoshimatsu, H. (1988). Central control of sexual behavior. Brain Research Bulletin *20*, 863–870.

Ortony, A., and Turner, T.J. (1990). What's basic about basic emotions? Psychological Review *97*, 315–331.

Packard, M.G., and Knowlton, B.J. (2002). Learning and memory functions of the Basal Ganglia. Annual Review of Neuroscience *25*, 563–593.

Pagani, J.H., and Rosen, J.B. (2009). The medial hypothalamic defensive circuit and 2,5-dihydro-2,4,5-trimethylthiazoline (TMT) induced fear: comparison of electrolytic and neurotoxic lesions. Brain Research *1286*, 133–146.

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

Panksepp, J. (1998). Affective Neuroscience (New York: Oxford U. Press).

Panksepp, J. (2000). Emotions as natural kinds within the mammalian brain. In: Handbook of Emotions, Second Edition, M. Lewis and J.M. Haviland-Jones, eds. (New York: The Guilford Press), pp. 137–156.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. Consciousness and Cognition *14*, 30–80.

Paré, D., Quirk, G.J., and Ledoux, J.E. (2004). New vistas on amygdala networks in conditioned fear. Journal of Neurophysiology *92*, 1–9.

Parkinson, J.A., Olmstead, M.C., Burns, L.H., Robbins, T.W., and Everitt, B.J. (1999). Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. Journal of Neuroscience *19*, 2401–2411.

Paton, J.J., Belova, M.A., Morrison, S.E., and Salzman, C.D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature *439*, 865–870.

Pavesi, E., Canteras, N.S., and Carobrez, A.P. (2011). Acquisition of Pavlovian fear conditioning using b-adrenoceptor activation of the dorsal premammillary nucleus as an unconditioned stimulus to mimic live predator-threat exposure. Neuropsychopharmacology *36*, 926–939.

Petrovich, G.D. (2011). Learning and the motivation to eat: forebrain circuitry. Physiology & Behavior *104*, 582–589.

Petrovich, G.D., Risold, P.Y., and Swanson, L.W. (1996). Organization of projections from the basomedial nucleus of the amygdala: a PHAL study in the rat. Journal of Comparative Neurology *374*, 387–420.

Petrovich, G.D., Canteras, N.S., and Swanson, L.W. (2001). Combinatorial amygdalar inputs to hippocampal domains and hypothalamic behavior systems. Brain Research Reviews *38*, 247–289.

Petrovich, G.D., Setlow, B., Holland, P.C., and Gallagher, M. (2002). Amygdalo- hypothalamic circuit allows learned cues to override satiety and promote eating. Journal of Neuroscience *22*, 8748–8753.

Pfaff, D.W. (1999). Drive: neurobiological and molecular mechanisms of sexual motivation (Cambridge, MA: MIT Press).

Pfaff, D., Ribeiro, A., Matthews, J., and Kow, L.M. (2008). Concepts and mechanisms of generalized central nervous system arousal. Annals of the New York Academy of Sciences *1129*, 11–25.

Phelps, E.A. (2006). Emotion and cognition: insights from studies of the human amygdala. Annual Review of Psychology *57*, 27–53.

Phelps, E.A., and LeDoux, J.E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. Neuron *48*, 175–187.

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

Pirri, J.K., and Alkema, M.J. (2012). The neuroethology of C. elegans escape. Current Opinion in Neurobiology, in press. Published online January 4, 2012. 10.1016/ j.conb.2011.12.007.

Pitkänen, A., Savander, V., and LeDoux, J.E. (1997). Organization of intraamygdaloid circuitries in the rat: An emerging framework for understanding functions of the amygdala. Trends in Neurosciences. *20*, 517–523.

Pollatos, O., Gramann, K., and Schandry, R. (2007). Neural systems connecting interoceptive awareness and feelings. Human Brain Mapping *28*, 9–18.

Preuss, T.M. (1995). Do rats have prefrontal cortex? The Rose-Woolsey-Akert program reconsidered. Journal of Cognitive Neuroscience *7*, 1–24.

Prinz, J. (2004). Which emotions are basic? In: Emotion, Evolution and Rationality, P. Cruise and D. Evans, eds. (Oxford: Oxford University Press), pp. 69–87.

Quirk, G.J., and Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. Neuropsychopharmacology *33*, 56–72.

Reep, R. (1984). Relationship between prefrontal and limbic cortex: a comparative anatomical review. Brain, Behavior and Evolution *25*, 5–80.

Reis, D.J., and LeDoux, J.E. (1987). Some central neural mechanisms governing resting and behaviorally coupled control of blood pressure. Circulation *76*, I2–I9.

Reisenzein, R. (1995). On Oatley and Johnson-Laird's Theory of emotion and hierarchical structures in the affective lexicon. Cognition and Emotion *9*, 383–416.

Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Classical Conditioning II: Current Research and Theory, A.A. Black and W.F. Prokasy, eds. (New York: Appleton-Centry-Crofts), pp. 64–99.

Reyes, B.A., Carvalho, A.F., Vakharia, K., and Van Bockstaele, E.J. (2011). Amygdalar peptidergic circuits regulating noradrenergic locus coeruleus neurons: linking limbic and arousal centers. Experimental Neurology *230*, 96–105.

Reynolds, S.M., and Berridge, K.C. (2008). Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. Nature Neuroscience *11*, 423–425.

Robbins, T.W. (1997). Arousal systems and attentional processes. Biological Psychology *45*, 57–71.

Robbins, T.W. (2005). Chemistry of the mind: neurochemical modulation of prefrontal cortical function. Journal of Comparative Neurology *493*, 140–146.

Robbins, T.W., and Everitt, B.J. (2007). A role for mesencephalic dopamine in activation: commentary on Berridge (2006). Psychopharmacology (Berl.) *191*, 433–437.

Rodrigues, S.M., LeDoux, J.E., and Sapolsky, R.M. (2009). The influence of stress hormones on fear circuitry. Annual Review of Neuroscience *32*, 289–313.

Roitman, M.F., Wheeler, R.A., and Carelli, R.M. (2005). Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. Neuron *45*, 587–597.

Roitman, M.F., Wheeler, R.A., Wightman, R.M., and Carelli, R.M. (2008). Real-time chemical responses in the nucleus accumbens differentiate rewarding and aversive stimuli. Nature Neuroscience *11*, 1376–1377.

Rolls, E.T. (1992). Neurophysiology and functions of the primate amygdala. In: The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction, J.P. Aggleton, ed. (New York: Wiley-Liss,Inc.), pp. 143–165.

Rolls, E.T. (1999). The Brain and Emotion (Oxford: Oxford University Press).

Rolls, E.T. (2005). Emotion Explained (New York: Oxford University Press).

Roozendaal, B., McEwen, B.S., and Chattarji, S. (2009). Stress, memory and the amygdala. Nature Reviews Neuroscience *10*, 423–433.

Rorty, A.O. (1980). Explaining emotions. In: Explaining Emotions, A.O. Rorty, ed. (Berkeley: University of California Press).

Rosenkranz, J.A., and Grace, A.A. (2002). Dopamine-mediated modulation of odour-evoked amygdala potentials during pavlovian conditioning. Nature *417*, 282–287.

Rudrauf, D., Lachaux, J.P., Damasio, A., Baillet, S., Hugueville, L., Martinerie, J., Damasio, H., and Renault, B. (2009). Enter feelings: Somatosensory responses following early stages of visual induction of emotion. International Journal of Psychophysiology *72*, 13–23.

Russell, J.A. (1980). A Circumplex model of affect. Journal of Personality and Social Psychology *39*, 1161–1178.

Russell, J.A. (1991). Natural language concepts of emotion. In: Perspectives in Personality, R. Hogan, W.H. Jones, A.J. Stewart, J.M. Healy, and D.J. Ozer, eds. (London: Jessica Kingsley).

Russell, J.A. (2003). Core affect and the psychological construction of emotion. Psychological Review *110*, 145–172.

Russell, J.A., and Barrett, L.F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. Journal of Personality and Social Psychology *76*, 805–819.

Salamone, J.D. (1994). The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. Behavioural Brain Research *61*, 117–133.

Sander, D., Grandjean, D., and Scherer, K.R. (2005). A systems approach to appraisal mechanisms in emotion. Neural Networks *18*, 317–352.

Saper, C.B. (2006). Staying awake for dinner: hypothalamic integration of sleep, feeding, and circadian rhythms. Progress in Brain Research *153*, 243–252.

Saper, C.B., Chou, T.C., and Elmquist, J.K. (2002). The need to feed: Homeostatic and hedonic control of eating. Neuron *36*, 199–211.

Sapolsky, R.M., Krey, L.C., and McEwen, B.S. (1986). The neuroendocrinology of stress and aging: the glucocorticoid cascade hypothesis. Endocrine Reviews *7*, 284–301.

Sarter, M., Hasselmo, M.E., Bruno, J.P., and Givens, B. (2005). Unraveling the attentional functions of cortical cholinergic inputs: interactions between signal-driven and cognitive modulation of signal detection. Brain Research Reviews *48*, 98–111.

Scarantino, A. (2009). Core affect and natural affective kinds. Philosophy of Science *76*, 940–957.

Schachter, S. (1975). Cognition and centralist-peripheralist controversies in motivation and emotion. In: Handbook of Psychobiology, M.S. Gazzaniga and C.B. Blakemore, eds. (New York: Academic Press), pp. 529–564.

Schachter, S., and Singer, J.E. (1962). Cognitive, social, and physiological determinants of emotional state. Psychological Review *69*, 379–399.

Scherer, K.R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. Review of Personality and Social Psychology *5*, 37–63.

Scherer, K.R. (1988). Criteria for emotion-antecedent appraisal: A review. In: Cognitive Perspectives on Emotion and Motivation, V. Hamilton, G.H. Bower, and N.H. Frijda, eds. (Norwell, MA: Kluwer Academic Publishers), pp. 89–126.

Scherer, K. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In: Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development, M. Lewis and I. Granic, eds. (New York: Cambridge University Press), pp. 70–99.

Scherer K.R., Schorr A., and Johnstone T., eds. (2001). Appraisal processes in emotion: Theory, Methods, Research (London: London University Press).

Schildkraut, J.J., and Kety, S.S. (1967). Biogenic amines and emotion. Science *156*, 21–37.

Schiller, D., and Delgado, M.R. (2010). Overlapping neural systems mediating extinction, reversal and regulation of fear. Trends in Cognitive Sciences (Regul. Ed.) *14*, 268–276.

Schimitel, F.G., de Almeida, G.M., Pitol, D.N., Armini, R.S., Tufik, S., and Schenberg, L.C. (2012). Evidence of a suffocation alarm system within the periaqueductal gray matter of the rat. Neuroscience *200*, 59–73.

Schneirla, T.C. (1959) An evolutionary theory of biphasic processes underlying approach and withdrawal. In: Nebraska Symposium on Motivation (Jones MR, ed), pp 1–42: University of Nebraska Press.

Schober, J., Weil, Z., and Pfaff, D. (2011). How generalized CNS arousal strengthens sexual arousal (and vice versa). Hormones and Behavior *59*, 689–695.

Schoenbaum, G., and Setlow, B. (2003). Lesions of nucleus accumbens disrupt learning about aversive outcomes. Journal of Neuroscience *23*, 9833–9841.

Schulkin, J. (2003). Rethinking Homeostasis: Allostatic Regulation in Physiology and Pathophysiology (Cambridge: MIT Press).

Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. Current Opinion in Neurobiology *7*, 191–197.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. Annual Review of Psychology *57*, 87–115.

Sears, R.M., Liu, R.J., Narayanan, N.S., Sharf, R., Yeckel, M.F., Laubach, M., Aghajanian, G.K., and DiLeone, R.J. (2010). Regulation of nucleus accumbens activity by the hypothalamic neuropeptide melanin-concentrating hormone. Journal of Neuroscience *30*, 8263–8273.

Seligman, M.E.P. (1971). Phobias and Preparedness. Behavior Therapy *2*, 307–320.

Selye, H. (1955). Stress and disease. Science *122*, 625–631.

Semendeferi, K., Teffer, K., Buxhoeveden, D.P., Park, M.S., Bludau, S., Amunts, K., Travis, K., and Buckwalter, J. (2011). Spatial organization of neurons in the frontal pole sets humans apart from great apes. Cerebral Cortex *21*, 1485–1497.

Sesack, S.R., and Grace, A.A. (2010). Cortico-Basal Ganglia reward network: Microcircuitry. Neuropsychopharmacology *35*, 27–47.

Shallice, T. (1988). Information processing models of consiousness. In: Consciousness in Contemporary Science, A. Marcel and E. Bisiach, eds. (Oxford: Oxford U. Press), pp. 305–333.

Shallice, T., Stuss, D.T., Alexander, M.P., Picton, T.W., and Derkzen, D. (2008). The multiple dimensions of sustained attention. Cortex *44*, 794–805.

Sheffield, F.D., and Roby, T.B. (1950). Reward value of a non-nutritive sweettaste. Journal of Comparative and Physiological Psychology *43*, 471–481.

Shepherd, G. (1988). Neurobiology, Second Edition (Oxford: Oxford University Press).

Sherrington, C.S. (1906). The Integrative Action of the Nervous System (New Haven: Yale U. Press).

Siegel, A., and Edinger, H. (1981). Neural control of aggression and rage behavior. In: Handbook of the Hypothalamus, Vol. 3, Behavioral Studies of the Hypothalamus, P.J. Morgane and J. Panksepp, eds. (New York: Marcel Dekker), pp. 203–240.

Smith, C.A., and Ellsworth, P.C. (1985). Patterns of cognitive appraisal in emotion. Journal of Personality and Social Psychology *48*, 813–838.

Sosulski, D.L., Bloom, M.L., Cutforth, T., Axel, R., and Datta, S.R. (2011). Distinct representations of olfactory information in different cortical centres. Nature *472*, 213–216.

Stellar, E. (1954). The physiology of motivation. Psychological Review *61*, 5–22.

Steriade, M. (1995). Brain activation, then (1949) and now: coherent fast rhythms in corticothalamic networks. Archives Italiennes de Biologie *134*, 5–20.

Steriade, M. (2004). Acetylcholine systems and rhythmic activities during the waking—sleep cycle. Progress in Brain Research *145*, 179–196.

Striedter, G.F. (2005). Principles of Brain Evolution (Sunderland: Sinauer Associates).

Sutton, R.S., and Barto, A.G. (1987) A temporal-difference model of classical conditioning. In: Proceedings of the Ninth Annual Conference of the Cognitive Science Society, pp 355–378.

Swanson, L.W. (2000). Cerebral hemisphere regulation of motivated behavior. Brain Research *886*, 113–164.

Swanson, L.W. (2002). Brain Architecture: Understanding the Basic Plan (Oxford: Oxford University Press).

Talarovicova, A., Krskova, L., and Kiss, A. (2007). Some assessments of the amygdala role in suprahypothalamic neuroendocrine regulation: A minireview. Endocrine Regulations *41*, 155–162.

Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., and de Gelder, B. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. Proceedings of the National Academy of Sciences USA *106*, 17661–17666.

Thompson, R., and LeDoux, J.E. (1974). Common brain regions essential for the expression of learned and instinctive visual habits in the albino rat. Bulletin of the Psychonomic Society *4*, 78–80.

Tinbergen, N. (1951). The Study of Instinct (New York: Oxford University Press).

Tolman, E.C. (1932). Purposive Behavior (New York: Appleton-Century-Crofts).

Tomkins, S.S. (1962). Affect, Imagery, Consciousness (New York: Springer).

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

Trowill, J.A., Panksepp, J., and Gandelman, R. (1969). An incentive model of rewarding brain stimulation. Psychological Review *76*, 264–281.

von Holst, E., and von Saint-Paul, U. (1962). Electrically controlled behavior. Scientific American *206*, 50–59.

Vyas, A., Kim, S.K., Giacomini, N., Boothroyd, J.C., and Sapolsky, R.M. (2007). Behavioral changes induced by Toxoplasma infection of rodents are highly specific to aversion of cat odors. Proceedings of the National Academy of Sciences USA *104*, 6442–6447.

Walker, D.L., and Davis, M. (2002). Light-enhanced startle: further pharmacological and behavioral characterization. Psychopharmacology (Berl.) *159*, 304–310.

Wallace, K.J., and Rosen, J.B. (2000). Predator odor as an unconditioned fear stimulus in rats: elicitation of freezing by trimethylthiazoline, a component of fox feces. Behavioral Neuroscience *114*, 912–922.

Wang, L., Han, X., Mehren, J., Hiroi, M., Billeter, J.C., Miyamoto, T., Amrein, H., Levine, J.D., and Anderson, D.J. (2011). Hierarchical chemosensory regulation of male-male social interactions in Drosophila. Nature Neuroscience *14*, 757–762.

Weinberger, N.M. (1995). Retuning the brain by fear conditioning. In: The Cognitive Neurosciences, M.S. Gazzaniga, ed. (Cambridge, MA: The MIT Press), pp. 1071–1090.

Whalen, P.J., and Phelps, E.A. (2009). The Human Amygdala (New York: Guilford Press).

Whalen, P.J., Kagan, J., Cook, R.G., Davis, F.C., Kim, H., Polis, S., McLaren, D.G., Somerville, L.H., McLean, A.A., Maxwell, J.S., and Johnstone, T. (2004). Human amygdala responsivity to masked fearful eye whites. Science *306*, 2061.

Whorf, B.L. (1956). Language, Thought, and Reality (Cambridge: Technology Press of MIT).

Wickens, J.R., Horvitz, J.C., Costa, R.M., and Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. Journal of Neuroscience *27*, 8181–8183.

Wierzbicka, A. (1994). Emotion, language, and cultural scripts. In: Emotion and Culture: Empirical Studies of Mutual Influence, S. Kitayama and H.R. Markus, eds. (Washington, DC: American Psychological Association), pp. 133–196.

Wise, R.A. (2006). Role of brain dopamine in food reward and reinforcement. Philosophical Transactions of the Royal Society B: Biological Sciences *361*, 1149–1158.

Wise, S.P. (2008). Forward frontal fields: phylogeny and fundamental function. Trends in Neurosciences. *31*, 599–608.

Yin, J.C., and Tully, T. (1996). CREB and the formation of long-term memory. Current Opinion in Neurobiology *6*, 264–268.

Yovell, Y., Kandel, E.R., Dudai, Y., and Abrams, T.W. (1992). A quantitative study of the Ca2+/calmodulin sensitivity of adenylyl cyclase in Aplysia, Drosophila, and rat. Journal of Neurochemistry *59*, 1736–1744.

Zachar, P. (2006). The classification of emotion and scientific realism. Journal of Theoretical and Philosophical Psychology *26*, 120–138.

Zajonc, R. (1980). Feeling and thinking: preferences need no inferences. American Psychologist *35*, 151–175.

Zhang, T.Y., Parent, C., Weaver, I., and Meaney, M.J. (2004). Maternal programming of individual differences in defensive responses in the rat. Annals of the New York Academy of Sciences *1032*, 85–103.

Regina M. Sullivan
Margo S. Landers
Emotional Brain Institute, Nathan Kline Institute, Orangeburg, NY, USA
Child and Adolescent Psychiatry, New York University School of Medicine, New York, NY, USA.

Development of the Emotional Brain and Infant Attachment: Lessons from an Animal Model[1]

Abstract

Survival of altricial infants depends on attachment to the caregiver, which requires infants to learn and maintain proximity to the caregiver to elicit caregiving. Using the literature on infant rats, we review the neurobiology of attachment, where learning about the caregiver is supported by a unique neural circuitry to ensure the infant-caregiver relationship is rapidly formed and maintained. Born deaf and blind, neonatal rat pups rely on the mother's odor, which is rapidly learned during both the prenatal and postnatal period. The brain's attachment circuit relies on a hyperfunctioning locus coeruleus and copious amounts of norepinephrine to produce learning-induced changes in the olfactory bulb and anterior piriform cortex. Infants also possess a reduced ability to acquire learned amygdala-dependent aversions or fear, which is facilitated through attenuated amygdala activity. This attachment circuitry constrains the infant to only learned preferences and prevents aversion to the caregiver, regardless of the

---

[1] This work was supported by grants NICHD-HD33402, NIMH-MH80603.

quality of care received. This attachment system is preserved in abusive attachment learning to maintain pup contact with the caregiver, although later in life, depressive-like behaviors and amygdala dysfunction emerge. These results suggest that the effects of early life maltreatment target the amygdala, disrupt infant social behavior, and produces adult depressive-like behavior.

1. Introduction

The infant and parent/caregiver form a reciprocal bond and the past 60 years of research have shown this earliest relationship has profound and enduring effects on emotion, cognition, and brain development. It was the combined insights of perceptive clinicians and researchers working with nonhuman animals that linked disturbed maternal care/separation and disturbed emotional and cognitive functioning that began in infancy and lasted into adulthood (Bowlby, 1969; Harlow and Harlow, 1965; Hofer, 2006). These pioneering studies provided a foundation for the field of early life experience and provided a unifying clinical and basic research theme that maternal behaviors were critical for the normal development of infants. Almost immediately the paradigm of infant separation from the mother, which is sometimes called maternal deprivation, became an important paradigm that permitted easily defined and measureable variables.

More recently, the approaches that define maternal care quality or mimic maternal care by providing/removing pups with specific sensory stimulation and/or handling have emerged (reviews Hofer and Sullivan, 2008; Romeo, Tang and Sullivan, 2009). Implicit in this approach is that the maternal-infant relationship is critical in programing infants' emotional and cognitive development. A more immediate role of maternal behavior is to meet the needs of the infant to ensure survival. Indeed, the infant's social behaviors are well-created to form and maintain the infant-caregiver social relationship. The infants of many altricial species must learn to recognize their caregiver as the target of their social behavior, and continue to express proximity-seeking behaviors toward their caregiver in order to receive food, protection, and warmth necessary for survival. This learning about the caregiver and the emergence of social behavior directed toward the caregiver is referred to as bonding and attachment, and this process has wide phylogenetic representation, including chicks, rodents, nonhuman primates, and humans. Here, we review the literature on infant attachment learning and the underlying neural circuitry that mediate early infant-caregiver social interactions, the transitioning role of this behavior and circuitry during development, and the enduring effects of early life experience on both the attachment circuitry and adult behavior.

2. Early-life social behavior: attachment learning

Attachment learning occurs in many species, including rats (Teicher et al., 1978; Blass and Teicher, 1980; Alberts and May, 1984; Risser and Slotnick, 1987; Polan and Hofer, 1998), rabbits (Hudson and Distel, 1983; Hudson, 1985), mice (Hennessy, Li, and Levine, 1980; Coppola, Coltrane, and Arsov, 1994; Moles, Kieffer, and D'Amato, 2004; Armstrong, DeVito, and Cleland, 2006; Roth et al., 2013), and nonhuman primates (Hennessy, Maken, and Graves, 1995, 2002; Carter and Keverne, 2002; DeVries, Glasper, and Detillion, 2003). During this time of dependency upon the mother, the infants' behaviors are focused on keeping physical contact with the mother. The sensory stimuli controlling this attachment varies with each species but in the infant rat, this behavior is controlled by maternal odor (Galef and Kaner, 1980; Leon, 1992). Once pups contact the mother's ventrum, perioral somatosensory cues combined with maternal odor become proximal maternal stimulus for nipple attachment (Hofer, Shair, and Singh, 1976; Teicher and Blass, 1977; Polan and Hofer, 1998).

When the importance of maternal odor was first discovered, it was assumed to be a pheromone since it functions to control pups' behavior at birth (Leon, 1992). The maternal odor elicits approach, physical contact with the mother, and nipple attachment. Without the maternal odor, pups show greatly diminished contact with the mother, failure to nipple attach and exhibit low survival rates (Blass and Teicher, 1980; Hofer et al., 1976; Raineki, Moriceau and Sullivan, 2010). However, it was quickly discovered that the maternal odor is learned (Leon, 1975; Rudy and Cheatle, 1977; Brunjes and Alberts, 1979; Galef and Kaner, 1980; Pedersen, Williams, and Blass, 1982; Campbell, 1984; Sullivan et al., 1986a; Miller, Jagielo, and Spear, 1989; Sullivan et al., 1990; Terry and Johanson, 1996). Any neutral odor appears to be able to become the maternal odor simply by placing a novel odor (i.e. peppermint or citral) on the mother during mother-infant interactions (Galef and Kaner, 1980; Alberts and May, 1984; Sullivan et al. 1990). Controlled classical conditioning experiments performed outside the nest without the mother also demonstrated that pups needed to learn the maternal odor (Spear, 1978; Haroutunian and Campbell, 1979; Sullivan et al., 1986b; Camp and Rudy, 1988; Sullivan et al., 2000; Roth and Sullivan, 2005; Moriceau et al., 2006; Raineki, Moriceau, and Sullivan, 2010). Indeed, pairing a novel odor with a reward is sufficient to produce both learned odor preferences (demonstrated by an approach to the odor) and nipple attachment. Many stimuli seem to be capable of functioning as a reward to support this early learning: pairing a novel odor with milk, warmth, nursing, and tactile stimulation or stroking to mimic grooming all support odor attachment learning (Galef and Sherry, 1973; Leon, 1975; Johanson and Hall, 1979; Johanson and Teicher, 1980; Brake, 1981; Pedersen, Williams, and Blass, 1982; Alberts and May, 1984; Sullivan, Hofer, and Brake, 1986; Weldon, Travis, and Kennedy, 1991; McLean et al., 1993; Wilson and Sullivan, 1994). Finally, the maternal odor continues to be learned repeatedly throughout the early life, perhaps because the maternal odor changes as the mother eats new foods (Pedersen, Williams, and Blass, 1982).

Somatosensory input is also important for pups' normal interactions with their mother as denervation of the whisker or whisker clipping greatly disrupts nipple attachment and interactions with the mother (Hofer et al., 1981). Learning also occurs within this sensory system: passive whisker stimulation paired with a reward results in stimulation-evoked movements of body elements such as increased body, head, and oromotor activity in pups as early as 1 day old, suggesting this whisker stimulation has functional significance (Landers and Sullivan, 1999a; Landers and Zeigler, 2006; Welker, 1964). Together, this research indicates that the maternal odor is used for early attachment but interacts with the somatosensory system.

3. Attachment learning circuitry

We have begun to document the neural circuit that supports attachment learning. The olfactory bulb, which is the first relay station for odor information within the brain, appears critical for this odor learning to occur. However, attachment learning is also associated with plasticity in the anterior piriform cortex, which receives direct information from the olfactory bulb and is part of the olfactory cortex. This learning occurs not only in response to odors experienced in the nest (Sullivan et al., 1990), but also in controlled learning experiments outside the nest (Sullivan et al. 1986a; Sullivan, Hofer, and Brake, 1986b; Woo, Coopersmith, and Leon, 1987; Johnson, Woo, Duong, Nguyen and Leon, 1995; Moriceau and Sullivan, 2004a,b; Roth and Sullivan, 2005; Yuan at al., 2002, 2003; Sullivan and Wilson, 1991; Wilson, Sullivan and Leon, 1987; Raineki, Moriceau, and Sullivan, 2010). These learning-induced olfactory bulb changes are attributable to the large influx of norepinephrine (NE) released from the LC (Shipley, Halloran, de la Torre, 1985), which prevents the mitral cells of the olfactory bulb from habituating to continual olfactory stimulation (Sullivan, Wilson, and Leon, 1989; Sullivan et al., 1992; Sullivan et al., 2000b; Wilson, Sullivan, and Leon, 1987; Okutani et al., 1998). In the infant, the abundant amount of NE released to the olfactory bulb is inducible by a wide range of sensory stimuli (Nakamura and Sakaguchi, 1990; Rangel and Leon, 1995). Furthermore, NE is both necessary and sufficient for the learning-induced behavioral and neural changes displayed in the infant. At the molecular level, NE also increases CREB phosphorylation (pCREB) via cAMP stimulation (McLean et al., 1999; Yuan et al., 2003; Zhang et al., 2003). This ultimately activates the transcription of immediate-early and late-response genes whose read-outs support synapse formation, neurogenesis, and learning (Bekinschtein et al., 2008; Tao et al., 1998). This is a common cellular cascade that supports learning in many species throughout development (Carew, 1996; Carew and Sutton, 2001; Rankin, 2002). However, what is unique to pups is the unique role of NE, which plays a more modulatory role in adult learned behavior (Ferry and McGaugh, 2000; McGaugh, 2006). In summary, the data we have reviewed thus far suggest that the contingent events of stimulus-induced NE release from the LC and NE-induced physiological and molecular changes in the olfactory bulb and anterior piriform cortex support the neural plasticity responsible for the acquisition of olfactory-based attachment behavior in the infant rat.

4. Fear and amygdala are attenuated in early life

Infant attachment learning is also characterized by limitations in aversive and fear learning. This was first demonstrated in an imprinting paradigm where shocking a chick actually enhances following of the surrogate caregiver, although shock supports avoidance just hours after the imprinting critical period closes (Hess, 1962; Rajecki, Lamb, and Obmascher, 1978b; Salzen, 1970). Similarly, shocking an infant dog or rat results in a strong attachment to the caregiver (Camp and Rudy, 1988; Roth and Sullivan, 2005; Sullivan, Hofer, and Brake, 1986; Sullivan et al., 2000a; Spear, 1978; Stanley, 1962). Infant primates and both human and nonhuman primates exhibit strong proximity-seeking behavior toward an abusive mother (Harlow and Harlow, 1965; Maestripieri, Tomaszycki, and Carroll, 1999; Sanchez, Ladd, and Plotsky, 2001; Suomi, 2003). This reduced fear/inhibition in infant rats has also been demonstrated in more controlled experiments in a laboratory setting: pups show suppressed fear to predators and attenuated inhibitory conditioning, and passive avoidance until after postnatal day 10-12 (Blozovski and Cudennec, 1980; Camp and Rudy, 1988; Collier et al., 1979; Goldman and Tobach, 1967; Haroutunian and Campbell, 1979; Stehouwer and Campbell, 1978; Sullivan et al., 2000a; Myslivecek, 1997). Indeed, aversive stimuli such as moderate shock and tailpinch, when paired with odor, support learned odor preferences in infant rats (Spear, 1978; Camp and Rudy, 1988; Haroutunian and Campbell, 1979; Moriceau et al., 2006; Moriceau and Sullivan, 2006; Roth and Sullivan, 2005; Sullivan, Hofer, and Brake, 1986; Sullivan et al., 2000a), despite an apparent pain response (Barr, 1995; Collier and Bolles, 1980; Emerich et al., 1985; Shair et al., 1997; Fitzgerald, 2005; Stehouwer and Campbell, 1978).

It appears that the lack of amygdala plasticity may play a leading role in pups' limited aversion learning since the functional emergence of the amygdala appears to coincide with the emergence of this learning (Blozovski and Cudennec, 1980; Collier et al., 1979; Myslivecek, 1997; Roth and Sullivan, 2005; Moriceau and Sullivan, 2006; Moriceau et al., 2006; Moriceau et al., 2004; Wiedenmayer and Barr, 2001). In older animals ranging from rodents to humans, amygdala activity is readily evoked by aversive stimuli in classical conditioning and natural fear paradigms (Davis, 1997; Fanselow and Gale, 2003; Fanselow and LeDoux, 1999; Herzog and Otto, 1997; Maren, 2003; McGaugh, Roozendaal, and Cahill, 1999; Pape and Stork, 2003; Pare, Quirk, and LeDoux, 2004; Rosenkranz and Grace, 2002; Sananes and Campbell, 1989; Schettino and Otto, 2001; Sevelinges et al., 2004; Blair et al., 2001; Sigurdsson et al., 2007).

Amygdala immaturity does not seem to underlie the lack of infant amygdala plasticity, although amygdala development is protracted (Bayer, 1980; Mizukawa, Tseng and Otsuka, 1989; Berdel, Morys, and Maciejewska, 1997; Morys et al., 1999; Nair and Gonzalez-Lima, 1999; Berdel and Morys, 2000b; Berdel and Morys, 2000a; Bouwmeester, Smits and Van Ree, 2002; Cunningham, Bhattacharyya, and Benes, 2002; Thompson, Sullivan, and Wilson, 2008). An evolutionary explanation may be this paradoxical attachment: perhaps it is better for an altricial infant to have a bad caretaker than no caretaker, since the infant is dependent upon the caregiver, warmth and caregiver protection for survival (Hofer and Sullivan, 2008). This notion is supported by data showing pups can learn aversions outside attachment. For example, infant rats are able to learn to avoid odors if paired with malaise, such as that produced by a LiCl injection or 1.2 mA shock (Hoffmann, Hunt, and Spear, 1990; Alleva and Calamandrei, 1986; Campbell, 1984; Coopersmith, Lee, and Leon, 1986; Gruest, Richer, and Hars, 2004; Haroutunian and Campbell, 1979; Hennessy, Smotherman, and Levine, 1976; Smotherman, Hennessy, and Levine, 1976; Hoffmann et al., 1987; Hunt, Spear, and Spear, 1991; Hunt et al., 1993; Miller, Molina, and Spear, 1990; Molina, Hoffmann, and Spear, 1986; Richardson and McNally, 2003; Rudy and Cheatle, 1983; Smotherman, 1982; Smotherman and Robinson, 1985; Smotherman and Robinson, 1990; Spear, 1978; Stickrod, Kimble, and Smotherman, 1982; Shionoya et al., 2006; Abate, Spear, and Molina, 2001; Spear and Rudy, 1991). Interestingly, while in adult and pre-weaning rats the amygdala responds to odor-malaise conditioning (Bermudez-Rattoni et al., 1986; LeDoux, 2000; Gale et al., 2004; Touzani and Sclafani, 2005), in infants, odor-malaise uses a non-amygdala neural circuit for odor aversion learning (Shionoya et al., 2006). Remarkably, even this malaise learning is constrained if the mother is present: if neonatal rats are nursing during odor-LiCl or taste-LiCl conditioning, aversion learning is prevented (Shionoya et al., 2006; Gubernick and Alberts, 1984; Martin and Alberts, 1979; Melcer, Alberts, and Gubernick, 1985). Together, these data indicate that infants, especially when related to attachment to the caregiver, do not readily learn aversions. We attribute this to a unique neural circuitry optimized to facilitate attachment to the caregiver, regardless of the quality of care provided.

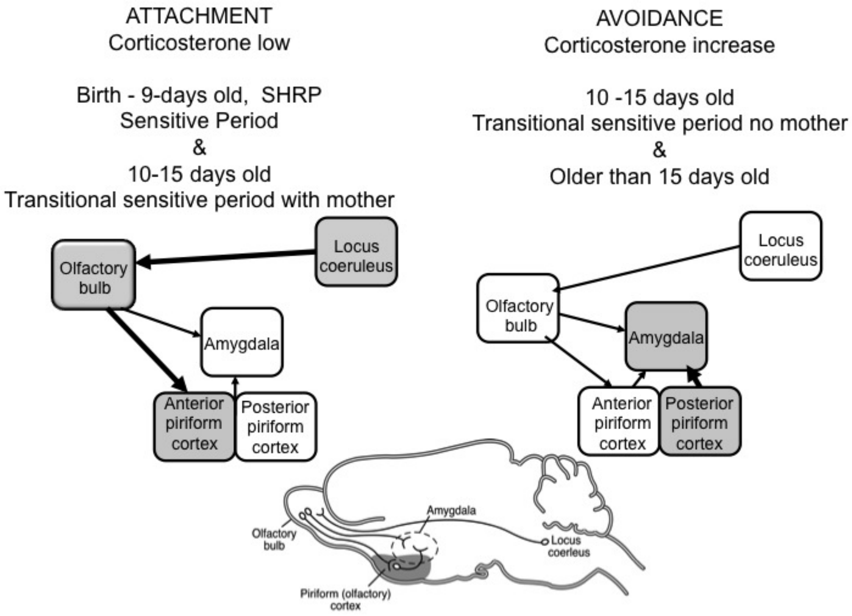


Figure 1. Here we illustrate attachment learning and transition to amygdala-dependent fear learning during the Sensitive Period and Transitional Sensitive Period using odor-0.5mA shock conditioning. As illustrated on the left, during the early life Sensitive Period for attachment, pups' olfactory learning results in the preference learning associated with attachment to the mother. Avoidance/fear learning is inhibited because pups' corticosterone level is low (stress hyporesponsive period, SHRP) and the amygdala does not exhibit the plasticity required for fear learning. As is illustrated on the right, this attachment learning system terminates around postnatal day 10, when pups endogenous corticosterone levels increases (SHRP ends) and pups learn amygdala-dependent fear. However, the attachment system can re-emerge in these older pups if the mother is present and social buffers (attenuates) pups' shock-induced corticosterone release to prevent amygdala plasticity. By postnatal day 16, the amygdala plasticity no longer requires corticosterone and pups learn amygdala-dependent fear regardless of corticosterone level, similarly to adults.

5. The critical role of corticosterone in emergence of fear learning

While the most parsimonious explanation for the delayed functional emergence of amygdala-dependent fear learning is delayed maturation of the amygdala, we have shown that pharmacological manipulations of corticosterone (CORT) levels control whether pups learn attachment or fear. In infant rats, CORT levels are relatively low (Walker et al., 1986; Henning, 1978), and the ability of most stressful stimuli (i.e., restraint, shock; Grino et al., 1994; Levine, 1962; 2001; Rosenfeld, Suchecki, and Levine, 1992) to evoke CORT secretion is greatly reduced compared to that in older animals (Guillet and Michaelson, 1978; Levine, 1967; Butte et al., 1973; Guillet, Saffran, and Michaelson, 1980; Gilles, Schultz, and Baram, 1996; Cate and Yasumura, 1975). This period of reduced hypothalamic-pituitary-adrenal (HPA) axis responsiveness during neonatal development has been termed the "stress hypo-responsive period" (SHRP). Sensory stimulation provided by the mother during nursing and grooming seems to control the pups' low CORT levels (Levine, 1962; van Oers et al., 1998). In fact, prolonged maternal separation (~24 hrs), which deprives pups of maternal sensory stimulation, increases pups' CORT levels (Levine, 2001), while the replacement of maternal sensory stimulation or maternal presence is able to reinstate the low level of CORT (Stanton, Wallstrom, and Levine, 1987; Stanton and Levine, 1990; Suchecki, Rosenfeld, and Levine, 1993). This reduced stress reactivity experienced by neonates is hypothesized to protect the developing organism from the negative influences of stress hormones (Sapolsky and Meaney, 1986).

CORT is able to switch whether infants learn an aversion or a preference. Specifically, increasing CORT by systemic injections or by intra-amygdala infusions during 0.5mA odor-shock conditioning or presentation of naturally aversive stimuli is sufficient to elicit both a fear response (learned or unlearned fear) and amygdala participation in the infant (Takahashi, 1994; Moriceau and Sullivan, 2004b, 2006; Moriceau et al., 2004). Maternal presence in older animals will lower CORT levels following stressful stimuli such as shock (Stanton, Wallstrom, and Levine, 1987; Suchecki, Rosenfeld, and Levine, 1993), block fear learning, reinstate the attachment learning (preference) and prevent the participation of the amygdala in learning (Moriceau and Sullivan, 2006). After PN15, only fear will be learned during odor-shock conditioning (Upton and Sullivan, 2010). Furthermore, we have verified the causal relationship between maternal presence and suppression of a shock-induced CORT release in pups' odor aversion learning by systemic and intra-amygdala CORT infusions, which then permit pups to learn odor aversions even in the presence of the mother. This is in sharp contrast to the role of CORT in adults, where it is considered to play a modulatory role in fear conditioning (Roozendaal, Quirarte, and McGaugh, 2002; Hui et al., 2004; Pugh et al., 1997; Thompson et al., 2004; Roozendaal, Carmi, and McGaugh, 1996; Corodimas et al., 1994). Together, these data indicate that during the attachment period, the mother maintains low infant CORT levels and attenuates amygdala activation, preventing infants from responding to fear/aversive stimuli. As pups mature, a developmental transition period emerges, increasing endogenous CORT levels and thereby enabling plasticity and fear learning. This suggests that maternal or genetically determined increases in CORT levels can have dramatic effects on infant attachment amygdala activity that are likely to have enduring effects on both behavior and amygdala.

6. Enduring effects of infant CORT and amygdala activity

The enduring effects of infant stress and maltreatment are well documented in both the clinical and basic literature. Adverse experiences within the context of attachment appear to have the most profound impact on adolescent and adult emotion and cognition in rodents, nonhuman primates, and humans (Harlow and Harlow, 1965; Levine, 1962; Rosenzweig et al., 1969; Bell and Denenberg, 1962; Denenberg, 1963; Kaufman et al., 2000; Schore, 2001; Bremner, 2003). Indeed, the effects appear ubiquitous and target many brain structures and functions in many species (Higley Hasert, Suomi and Linnoila, 1991; Plotsky and Meaney, 1993; Suomi, 1997; Caldji et al., 2003; Ladd et al., 2000; Liu, Caldji, Sharma, Plotsky and Meaney, 2000; Meaney, 2001; Meaney et al., 1996; Hall, Wilkinson, Humby and Robbins, 1999; Avishai-Eliner et al., 2001).

Work from our lab suggests that similar enduring, depressive-like effects are produced by repeated odor-shock conditioning but also by rearing pups with a maltreating mother. The first effect that emerges during development is disrupted social behavior and amygdala hyperactivity in infancy, although pups must be stressed to see the effect. As pups approach weaning, social behavior deficits emerge even without stress and by adolescence additional effects associated with depressive-like behavior emerge (Raineki et al., 2010). The depressive-like behaviors include decreased social behavior, decreased latency to immobility in a Forced Swim Test, and decreased sucrose consumption (Sevelinges et al., 2011). Additionally, depressive-like adult rats also showed disrupted amygdala activity, as indicated by reduced levels of amygdala paired-pulse inhibition (Sevelinges et al., 2011). Amazingly, both the depressive-like behaviors and amygdala dysfunction were brought to control levels simply by presenting the odor that was paired with shock during infant odor-shock conditioning. The ability of the odor to stop the depressive-like behavior is reminiscent of research on "safety signals" reducing conditioned fear responses, reducing depressive-like behaviors, and normalizing amygdala activity (Rogan, Leon, Perez and Kandel, 2005; Pollak, 2008; Schiller, Levy, Niv, LeDoux and Phelps, 2008). Finally, it should be noted that the unpaired odor-shock conditioning in infancy was not associated with depressive-like behavior, suggesting that shock alone is insufficient to produce these effects and further suggesting that perhaps activation of the attachment circuit with the trauma may be important (Sarro, Sullivan, and Barr, 2014). It is also noteworthy that only unpaired pups showed increased anxiety-like behaviors in the dark-light emergence test in adulthood. Together, these results suggest that odor in early life acquires powerful properties that can control infant attachment behaviors but also retains the power to control emotions in later life (Sullivan, 2012).

In conclusion, we have known since the 1950s that early life caregiving and trauma have profound influences on behavior of humans and other animals. More recent work has further demonstrated its importance and begun to highlight which brain areas are targeted, potential mechanisms for enduring effects, and the importance of genetics. Finally, the current work further suggests the specific role of trauma and attachment's powerful role for influencing adult emotional and cognitive well being.

References

Abate, P., Spear, N.E., and Molina, J.C. (2001). Fetal and infantile alcohol-mediated associative learning in the rat. Alcoholism: Clinical and Experimental Research *25*, 989–998.

Alberts, J.R., and May B. (1984). Nonnutritive, thermotactile induction of filial huddling in rat pups. Developmental Psychobiology *17*, 161–181.

Alleva, E., and Calamandrei, G. (1986). Odor-aversion learning and retention span in neonatal mouse pups. Behavioral and Neural Biology *46*, 348–357.

Armstrong, C.M.,L. DeVito, M., and Cleland, T.A. (2006). One-trial associative odor learning in neonatal mice. Chemical Senses *31*, 343–349.

Avishai-Eliner, S., Gilles, E., Eghbal-Ahmadi, M., Bar-El, Y., and Baram, T. (2001). Altered regulation of gene and protein expression of hypothalamic-pituitary-adrenal axis components in an immature rat model of chronic stress. Journal of Neuroendocrinology *13*, 799–807.

Barr, G.A. (1995). Ontogeny of nociception and antinociception. NIDA Research Monograph *158*, 172–201.

Bayer, S.A. (1980). Quantitative 3H-thymidine radiographic analyses of neurogenesis in the rat amygdala. Journal of Comparative Neurology *194*, 845–875.

Bekinschtein, P., Cammarota, M., Izquierdo, I., and Medina, J.H. (2008). BDNF and memory formation and storage. Neuroscientist *14*, 147–156.

Bell, R.W., and Denenberg, V.H. (1962). The interrelationships of shock and critical periods in infancy as they affect adult learning and activity. Animal Behaviour *11*, 21–27.

Berdel, B., and Morys, J. (2000a). Expression of calbindin-D28k and parvalbumin during development of rat's basolateral amygdaloid complex. International Journal of Developmental Neuroscience *18*, 501–513.

Berdel, B., and Morys, J. (2000b). Expression of calbindin-D28k and parvalbumin during development of rat's basolateral amygdaloid complex. International Journal of Developmental Neuroscience *18*, 501–513.

Berdel, B., Morys, J., and Maciejewska B. (1997). Neuronal changes in the basolateral complex during development of the amygdala of the rat. International Journal of Developmental Neuroscience *15*, 755–765.

Bermúdez-Rattoni, F., Grijalva, C.V., Kiefer, S.W., and Garcia J. (1986). Flavor-illness aversions: the role of the amygdala in the acquisition of taste-potentiated odor aversions. Physiology and Behavior *38 (4)*, 503–508.

Blair, H.T., Schafe, G.E., Bauer, E.P., Rodrigues, S.M., and LeDoux, J.E. (2001). Synaptic plasticity in the lateral amygdala: a cellular hypothesis of fear conditioning. Learning and Memory *8*, 229–242.

Blass, E.M., and Teicher, M.H. (1980). Suckling. Science *210*, 15–22.

Blozovski, D., and Cudennec, A. (1980). Passive avoidance learning in the young rat. Developmental Psychobiology *13*, 513–518.

Bouwmeester, H., Smits, K., and Van Ree, J.M. (2002). Neonatal development of projections to the basolateral amygdala from prefrontal and thalamic structures in rat. Journal of Comparative Neurology *450*, 241–255.

Bowlby, J. (1969). Attachment and Loss. Vol. 1 (New York: Basic Books).

Brake, S.C. (1981). Suckling infant rats learn a preference for a novel olfactory stimulus paired with milk delivery. Science *211*, 506–508.

Bremner, J.D. (2003). Long-term effects of childhood abuse on brain and neurobiology. Child and Adolescent Psychiatric Clinics of North America *12*, 271–292.

Brunjes, P.C., and Alberts, J.R. (1979). Olfactory stimulation induces filial preferences for huddling in rat pups. Journal of Comparative and Physiological Psychology *93*, 548–555.

Butte, J.C., Kakihana, R., Farnham, M.L., and Noble, E.P. (1973). The relationship between brain and plasma corticosterone stress response in developing rats. Endocrinology *92*, 1775–1779.

Caldji, C., Diorio, J., and Meaney, M.J. (2003). Variations in maternal care alter GABA(A) receptor subunit expression in brain regions associated with fear. Neuropsychopharmacology *28*, 1950–1959.

Camp, L.L., and Rudy, J.W. (1988). Changes in the categorization of appetitive and aversive events during postnatal development of the rat. Developmental Psychobiology *21*, 25–42.

Campbell, B.A. (1984). Reflections on the ontogeny of learning and memory. In: Comparative perspectives on the development of memory, R. Kail and N.E. Spear, eds. (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 23–35.

Carew, T.J. (1996). Molecular enhancement of memory formation. Neuron *16*, 5–8.

Carew, T.J., and Sutton, M.A. (2001). Molecular stepping stones in memory consolidation. Nature Neuroscience *4*, 769–771.

Carter, C.S., and Keverne, E.B. (2002). The neurobiology of social affiliation and pair bonding. In: Hormones, Brain, and Behavior, D.W. Pfaff, ed. (San Diego: Academic Press), pp. 299–337.

Cate, T.E., and Yasumura, S. (1975). Effects of ACTH and histamine stress on serum corticosterone and adrenal cyclic AMP in immature rats. Endocrinology *96*.

Collier, A., Mast, J., Meyer, D., and Jacobs, C. (1979). Approach-avoidance conflict in preweanling rats: A developmental study. Animal Learning & Behavior *7*, 514–520.

Collier, A.C., and Bolles, R.C. (1980). The ontogenesis of defensive reactions to shock in preweanling rats. Developmental Psychobiology *13*, 141–150.

Coopersmith, R., Lee, S., and Leon, M. (1986). Olfactory bulb responses after odor aversion learning by young rats. Brain Research *389*, 271–277.

Coopersmith, R., and Leon, M. (1986). Enhanced neural response by adult rats to odors experienced early in life. Brain Research *371*, 400–403.

Coppola, D.M., Coltrane, J.A., and Arsov, I. (1994). Retronasal or internasal olfaction can mediate odor-guided behaviors in newborn mice. Physiology & Behavior *56*, 729–736.

Corodimas, K.P., LeDoux, J.E., Gold, P.W., and Schulkin, J. (1994). Corticosterone potentiation of conditioned fear in rats. Annals of the New York Academy of Sciences *746*, 392–393.

Cunningham, M.G., Bhattacharyya, S., and Benes, F.M. (2002). Amygdalo-cortical sprouting continues into early adulthood: implications for the development of normal and abnormal function during adolescence. Journal of Comparative Neurology *453*, 116–130.

Davis, M. (1997). Neurobiology of fear responses: the role of the amygdala. Journal of Neuropsychiatry & Clinical Neurosciences *9*, 382–402.

Denenberg, V.H. (1963). Early experience and emotional development. Scientific American *208*, 138–146.

DeVries, A.C., Glasper, E.R., and Detillion, C.E. (2003). Social modulation of stress responses. Physiology & Behavior *79*, 399–407.

Emerich, D.F., Scalzo, F.M., Enters, E.K., Spear, N.E., and Spear, L.P. (1985). Effects of 6-hydroxydopamine-induced catecholamine depletion on shock-precipitated wall climbing of infant rat pups. Developmental Psychobiology *18*, 215–227.

Fanselow, M.S., and Gale, G.D. (2003). The amygdala, fear, and memory. Annals of the New York Academy of Sciences *985*, 125–134.

Fanselow, M.S., and LeDoux, J.E. (1999). Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. Neuron *23*, 229–232.

Ferry, B., and McGaugh, J.L. (2000). Role of amygdala norepinephrine in mediating stress hormone regulation of memory storage. Acta Pharmacologica Sinica *21*, 481–493.

Fitzgerald, M. (2005). The development of nociceptive circuits. Nature Reviews Neuroscience *6*, 507–520.

Gale, G.D., Anagnostaras, S.G., Godsil, B.P., Mitchell, S., Nozawa, T., Sage, J.R., Wiltgen, B., and Fanselow, M.S. (2004). Role of the basolateral amygdala in the storage of fear memories across the adult lifetime of rats. Journal of Neuroscience *24*, 3810–3815.

Galef, B.G., Jr., and Kaner, H.C. (1980). Establishment and maintenance of preference for natural and artificial olfactory stimuli in juvenile rats. Journal of Comparative and Physiological Psychology *94*, 588–595.

Galef, B.G., Jr., and Sherry, D.F. (1973). Mother's milk: a medium for transmission of cues reflecting the flavor of mother's diet. Journal of Comparative and Physiological Psychology *83*, 374–378.

Gilles, E., Schultz, L., and Baram, T. (1996). Abnormal corticosterone regulation in an immature rat model of continuous chronic stress. Pediatric Neurology *15*, 114–119.

Goldman, P.S., and Tobach, E. (1967). Behaviour modification in infant rats. Animal Behaviour *15*, 559–562.

Grino, M., Paulmyer-Lacroix, O., Faudon, M., Renard, M., and Anglade, G. (1994). Blockade of alpha 2-adrenoceptors stimulates basal and stress-induced adrenocorticotropin secretion in the developing rat through a central mechanism independent from corticotropin-releasing factor and arginine vasopressin. Endocrinology *135*, 2549–2557.

Gruest, N., Richer, P., and Hars, B. (2004). Emergence of long-term memory for conditioned aversion in the rat fetus. Developmental Psychobiology *44*, 189–198.

Gubernick, D.J., and Alberts, J.R. (1984). A specialization of taste aversion learning during suckling and its weaning-associated transformation. Developmental Psychobiology *17*, 613–628.

Guillet, R., and Michaelson, S.M. (1978). Corticotropin responsiveness in the neonatal rat. Neuroendocrinology *27*, 119–125.

Guillet, R., Saffran, M., and Michaelson, S.M. (1980). Pituitary-adrenal response in neonatal rats. Endocrinology *106*, 991–994.

Hall, F.S., Wilkinson, L.S., Humby, T., and Robbins, T.W. (1999). Maternal deprivation of neonatal rats produces enduring changes in dopamine function. Synapse *32*, 37–43.

Harlow, H., and Harlow, M. (1965). The affectional system. In: Behavior of Nonhuman Primates, A. Schrier, H. Harlow and F. Stollnitz, eds. (New York: Academic Press).

Haroutunian, V., and Campbell, B.A. (1979). Emergence of interoceptive and exteroceptive control of behavior in rats. Science *205*, 927–929.

Hennessy, J.W., Smotherman, W.P., and Levine, S. (1976). Conditioned taste aversion and the pituitary-adrenal system. Behavioral Biology *16*, 413–424.

Hennessy, M.B., Li, J., and Levine, S. (1980). Infant responsiveness to maternal cues in mice of 2 inbred lines. Developmental Psychobiology *13*, 77–84.

Hennessy, M.B., Maken, D.S., and Graves, F.C. (2002). Presence of mother and unfamiliar female alters levels of testosterone, progesterone, cortisol, adrenocorticotropin, and behavior in maturing guinea pigs. Hormones and Behavior *42*, 42–52.

Hennessy, M.B., Nigh, C.K., Sims, M.L., and Long, S.J. (1995). Plasma cortisol and vocalization responses of postweaning age guinea pigs to maternal and sibling separation: evidence for filial attachment after weaning. Developmental Psychobiology *28*, 103–115.

Henning, S.J. (1978). Plasma concentrations of total and free corticosterone during development in the rat. American Journal of Physiology *235*, E451–456.

Herzog, C., and Otto, T. (1997). Odor-guided fear conditioning in rats: 2. Lesions of the anterior perirhinal cortex disrupt fear conditioned to the explicit conditioned stimulus but not to the training context. Behavioral Neuroscience *111*, 1265–1272.

Hess, E (1962). Ethology: an approach to the complete analysis of behavior. In: New Directions in Psychology, R. Brown, E. Galanter, E. Hess, and G. Mendler, eds. (New York: Holt, Rinehart and Winston), pp. 159–199.

Higley, J.D., Hasert, M.F., Suomi, S.J., and Linnoila, M. (1991). Nonhuman primate model of alcohol abuse: effects of early experience, personality, and stress on alcohol consumption. Proceedings of the National Academy of Sciences USA *88*, 7261–7265.

Hofer, M. (2006). Psychobiological roots of early attachment. Current Directions in Psychological Sciences *15 (2)*. 84–88.

Hofer, M.A., Shair, H., and Singh, P. (1976). Evidence that maternal ventral skin substances promote suckling in infant rats. Physiology & Behavior *17*, 131–136.

Hofer, M.A., and Sullivan, R.M. (2008). Toward a neurobiology of attachment. In: Developmental Cognitive Neuroscience, C.A. Nelson and M. Lucian, eds. (Cambridge, MA: The MIT Press), pp. 599–616.

Hofer, M.A., Fisher A., and Shair H. (1981). Effects of infraorbital nerve section on survival, growth, and suckling behaviors of developing rats. Journal of Comparative and Physiological Psychology *95*, 123–133.

Hoffmann, H., Hunt, P., and Spear, N.E. (1990). Ontogenetic differences in the association of gustatory and tactile cues with lithium chloride and footshock. Behavioral and Neural Biology *53*, 441–450.

Hoffmann, H., Molina, J.C., Kucharski, D., and Spear, N.E. (1987). Further examination of ontogenetic limitations on conditioned taste aversion. Developmental Psychobiology *20*, 455–463.

Hudson, R. (1985). Do newborn rabbits learn the odor stimuli releasing nipple-search behavior? Developmental Psychobiology *18*, 575–585.

Hudson, R., and Distel, H. (1983). Nipple location by newborn rabbits: behavioral evidence for pheromonal guidance. Behaviour *85*, 260–275.

Hui, G.K., Figueroa, I.R., Poytress, B.S., Roozendaal, B., McGaugh, J.L., and Weinberger, N.M. (2004). Memory enhancement of classical fear conditioning by post-training injections of corticosterone in rats. Neurobiol. Learning and Memory *81*, 67–74.

Hunt, P.S., Molina, J.C., Rajachandran, L., Spear, L.P., and Spear, N.E. (1993). Chronic administration of alcohol in the developing rat: expression of functional tolerance and alcohol olfactory aversions. Behavioral and Neural Biology *59*, 87–99.

Hunt, P.S., Spear, L.P., and Spear, N.E. (1991). An ontogenetic comparison of ethanol-mediated taste aversion learning and ethanol-induced hypothermia in preweanling rats. Behavioral Neuroscience *105*, 971–983.

Johanson, I.B., and Hall, W.G. (1979). Appetitive learning in 1-day-old rat pups. Science *205*, 419–421.

Johanson, I.B., and Teicher, M.H. (1980). Classical conditioning of an odor preference in 3-day-old rats. Behavioral and Neural Biology *29*, 132–136.

Johnson, B.A., Woo, C.C., Duong, H., Nguyen, V., and Leon, M. (1995). A learned odor evokes an enhanced Fos-like glomerular response in the olfactory bulb of young rats. Brain Research *699*, 192–200.

Kaufman, J., Plotsky, P.M., Nemeroff, C.B., and Charney, D.S. (2000). Effects of early adverse experiences on brain structure and function: clinical implications. Biological Psychiatry *48*, 778–790.

Ladd, C.O., Huot, R.L., Thrivikraman, K.V., Nemeroff, C.B., Meaney, M.J., and Plotsky, P.M. (2000). Long-term behavioral and neuroendocrine adaptations to adverse early experience. Progress in Brain Research *122*, 81–103.

Landers, M.S., and Sullivan, R.M. (1999). Vibrissae evoked behavior and conditioning before functional ontogeny of somatosensory vibrissae cortex. Journal of Neuroscience *19*, 5131–5137.

Landers, M.S., and Zeigler, P.H. (2006). Development of rodent whisking: Trigeminal input and central pattern generation. Somatosens Mot Res *23*, 1–10.

LeDoux, J.E. (2000). Emotion circuits in the brain. Annual Review of Neuroscience *23*, 155–184.

Leon, M. (1975). Dietary control of maternal pheromone in the lactating rat. Physiology & Behavior *14*, 311–319.

Leon, M. (1992). The neurobiology of filial learning. Annual Review of Psychology *43*, 377–398.

Levine, S. (1962). Plasma-free corticosteroid response to electric shock in rats stimulated in infancy. Science *135*, 795–796.

Levine, S. (1967). Maternal and environmental influences on the adrenocortical response to stress in weanling rats. Science *156*, 258–260.

Levine, S. (1994). The ontogeny of the hypothalamic-pituitary-adrenal axis. The influence of maternal factors. Annals of the New York Academy of Sciences *746*, 275–288.

Levine, S. (2001). Primary social relationships influence the development of the hypothalamic-pituitary-adrenal axis in the rat. Physiology & Behavior *73*, 255–260.

Liu, D., Caldji, C., Sharma, S., Plotsky, P., and Meaney, M. (2000). Influence of neonatal rearing conditions on stress-induced adrenocorticotropin responses and norepinepherine release in the hypothalamic paraventricular nucleus. Journal of Neuroendocrinology *12*, 5–12.

Maestripieri, D., Tomaszycki, M., and Carroll, K.A. (1999). Consistency and change in the behavior of rhesus macaque abusive mothers with successive infants. Developmental Psychobiology *34*, 29–35.

Maren, S. (2003). The amygdala, synaptic plasticity, and fear memory. Annals of the New York Academy of Sciences *985*, 106–113.

Martin, L.T., and Alberts, J.R. (1979). Taste aversions to mother's milk: the age-related role of nursing in acquisition and expression of a learned association. Journal of Comparative and Physiological Psychology *93*, 430–445.

McGaugh, J.L. (2006). Make mild moments memorable: add a little arousal. Trends in Cognitive Sciences *10*, 345–347.

McGaugh, J.L., Roozendaal, B., and Cahill, L. (1999). Modulation of memory storage by stress hormones and the amygdaloid complex. In: Cognitive Neuroscience, M. Gazzaniga, ed. (Cambridge: MIT Press).

McLean, J.H., Darby-King, A., Sullivan, R.M., and King, S.R. (1993). Serotonergic influence on olfactory learning in the neonate rat. Behavioral and Neural Biology *60*, 152–162.

McLean, J.H., Harley, C.W., Darby-King, A., and Yuan, Q. (1999). pCREB in the neonate rat olfactory bulb is selectively and transiently increased by odor preference-conditioned training. Learning and Memory *6*, 608–618.

Meaney, M.J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. Annual Review of Neuroscience *24*, 1161–1192.

Meaney, M.J., Diorio, J., Francis, D., Widdwson, J., LaPlante, P., Caldji, C., Sharma, S., Seckl, J.R., and Plotsky, P.M. (1996). Early environmental regulation of forebrain glucocorticoid receptor gene expression: implications for adrenocortical responses to stress. Developmental Neuroscience *18*, 49–72.

Melcer, T., Alberts, J.R., and Gubernick, D.J. (1985). Early weaning does not accelerate the expression of nursing-related taste aversions. Developmental Psychobiology *18*, 375–381.

Miller, J.S., Jagielo, J.A., and Spear, N.E. (1989). Age-related differences in short-term retention of separable elements of an odor aversion. Journal of Experimental Psychology: Animal Behavior Processes *15*, 194–201.

Miller, J.S., Molina, J.C., and Spear, N.E. (1990). Ontogenetic differences in the expression of odor-aversion learning in 4- and 8-day-old rats. Developmental Psychobiology *23*, 319–330.

Mizukawa, K., Tseng, I.M., and Otsuka, N. (1989). Quantitative electron microscopic analysis of postnatal development of zinc-positive nerve endings in the rat amygdala using Timm's sulphide silver technique. Brain Research: Developmental Brain Research *50*, 197–203.

Moles, A., Kieffer, B.L., and D'Amato, F.R. (2004). Deficit in attachment behavior in mice lacking the mu-opioid receptor gene.[see comment]. Science *304*, 1983–1986.

Molina, J.C., Hoffmann, H., and Spear, N.E. (1986). Conditioning of aversion to alcohol orosensory cues in 5- and 10-day rats: subsequent reduction in alcohol ingestion. Developmental Psychobiology *19*, 175–183.

Moriceau, S., Roth, T.L., Okotoghaide, T., and Sullivan, R.M. (2004). Corticosterone controls the developmental emergence of fear and amygdala function to predator odors in infant rat pups. International Journal of Developmental Neuroscience *22*, 415–422.

Moriceau, S., and Sullivan, R.M. (2004a). Corticosterone influences on Mammalian neonatal sensitive-period learning. Behavioral Neuroscience *118*, 274–281.

Moriceau, S., and Sullivan, R.M. (2004b). Unique neural circuitry for neonatal olfactory learning. Journal of Neuroscience *24*, 1182–1189.

Moriceau, S., and Sullivan, R.M. (2006). Maternal presence serves as a switch between learning fear and attraction in infancy. Nature Neuroscience *9*, 1004–1006.

Moriceau, S., Wilson, D.A., Levine, S., and Sullivan, R.M. (2006). Dual circuitry for odor-shock conditioning during infancy: corticosterone switches between fear and attraction via amygdala. Journal of Neuroscience *26*, 6737–6748.

Morys, J., Berdel, B., Jagalska-Majewska, H., and Luczynska, A. (1999). The basolateral amygdaloid complex-its development, morphology and functions. Folia Morphologica (Warsz.) *58*, 29–46.

Myslivecek, J. (1997). Inhibitory learning and memory in newborn rats. Progress in Neurobiology *53*, 399–430.

Nair, H., and Gonzalez-Lima, F. (1999). Extinction of behavior in infant rats: development of functional coupling between septal, hippocampal, and ventral tegmental regions. Journal of Neuroscience *19*, 8646–8655.

Nakamura, S., and Sakaguchi, T. (1990). Development and plasticity of the locus coeruleus: a review of recent physiological and pharmacological experimentation. Progress in Neurobiology *34*, 505–526.

Okutani, F., Kaba, H., Takahashi, S., and Seto, K. (1998). The biphasic effects of locus coeruleus noradrenergic activation on dendrodendritic inhibition in the rat olfactory bulb. Brain Research *783*, 272–279.

Pape, H.C., and Stork, O. (2003). Genes and mechanisms in the amygdala involved in the formation of fear memory. Annals of the New York Academy of Sciences *985*, 92–105.

Pare, D., Quirk, G.J., and Ledoux, J.E. (2004). New vistas on amygdala networks in conditioned fear. Journal of Neurophysiology *92*, 1–9.

Pedersen, P.E., Williams, C.L., and Blass, E.M. (1982). Activation and odor conditioning of suckling behavior in 3-day-old albino rats. Journal of Experimental Psychology: Animal Behavior Processes *8*, 329–341.

Plotsky, P.M., and Meaney, M.J. (1993). Early postnatal experience alters hypothalamic corticotropin-releasing factor (CRF) mRNA, median eminence CRF content and stress-induced release in adult rats. Molecular Brain Research *18*, 195–200.

Polan, H.J., and Hofer, M.A. (1998). Olfactory preference for mother over home nest shavings by newborn rats. Developmental Psychobiology *33*, 5–20.

Pollak, S.D. (2003). Experience-dependent affective learning and risk for psychopathology in children. Annals of the New York Academy of Sciences *1008*, 102–111.

Pollak, D.S., Monje, F.J., Zuckerman, L., Denny, C.A., Drew, M.R. and Kandel, E.R. (2008). An animal model of a behavioral intervention for depression. Neuron *60*, 149–161.

Pugh, C.R., Tremblay, D., Fleshner, M., and Rudy, J.W. (1997). A selective role for corticosterone in contextual-fear conditioning. Behavioral Neuroscience *111*, 503–511.

Raineki, C., Moriceau, S., and Sullivan, R M. (2010) Developing a neurobehavioral animal model of infant attachment to an abusive caregiver. Biological Psychiatry *67(12)*, 1137–1145.

Rajecki, D., Lamb, M., and Obmascher, P. (1978a). Toward a general theory of infantile attachment: a comparative review of aspects of the social bond. Behavioral and Brain Sciences *3*, 417–464.

Rajecki, D., Lamb, M., and Obmascher, P. (1978b). Towards a general theory of infantile attachment: a comparative review of aspects of the social bond. Behavioral and Brain Sciences *3*, 417–464.

Rangel, S., and Leon, M. (1995). Early odor preference training increases olfactory bulb norepinephrine. Brain Research: Developmental Brain Research *85*, 187–191.

Rankin, C.H. (2002). Neuroscience. A bite to remember. Science *296*, 1624–1625.

Richardson, R., and McNally, G.P. (2003). Effects of an odor paired with illness on startle, freezing, and analgesia in rats. Physiology & Behavior *78*, 213–219.

Risser, J.M., and Slotnick, B.M. (1987). Nipple attachment and survival in neonatal olfactory bulbectomized rats. Physiology & Behavior *40*, 545–549.

Rogan, M., Leon, K.S. Perez, D.L., and Kandel, E.R. (2005). Distinct neural signatures for safety and danger in the amygdala and striatum of the mouse. Neuron *46*, 309–320.

Romeo, R.D., Tang ,A.C. and Sullivan, R.M, (2009). Early life experiences: Enduring behavioral and endocrinological consequences. In: Hormones, Brain and Behavior, D Pfaff, A. Arnold, and A.M. Etgen, eds. (Academic Press), pp. 1414–1415.

Roozendaal, B., Carmi, O., and McGaugh, J.L. (1996). Adrenocortical suppression blocks the memory-enhancing effects of amphetamine and epinephrine. Proceedings of the National Academy of Sciences USA *93*, 1429–1433.

Roozendaal, B., Quirarte, G.L., and McGaugh, J.L. (2002). Glucocorticoids interact with the basolateral amygdala beta-adrenoceptor-cAMP/cAMP/PKA system in influencing memory consolidation. European Journal of Neuroscience *15*, 553–560.

Rosenfeld, P., Suchecki, D., and Levine, S. (1992). Multifactorial regulation of the hypothalamic pituitary-adrenal axis during development. Neuroscience & Biobehavioral Reviews *16*, 553–568.

Rosenkranz, J.A., and Grace, A.A. (2002). Dopamine-mediated modulation of odour-evoked amygdala potentials during pavlovian conditioning. Nature *417*, 282–287.

Rosenzweig, M.R., Bennett, E.L., Diamond, M.C., Wu, S.Y., Slagle, R.W., and Saffran, E. (1969). Influences of environmental complexity and visual stimulation on development of occipital cortex in rat. Brain Research *14*, 427–445.

Roth, T., and Sullivan, R. (2005). Memory of early maltreatment: neonatal behavioral and neural correlates of maternal maltreatment within the context of classical conditioning. Biological Psychiatry *57*, 823–831.

Rudy, J.W., and Cheatle, M.D. (1977). Odor-aversion learning in neonatal rats. Science *198*, 845–846.

Rudy, J.W., and Cheatle, M.D. (1983). Odor-aversion learning by rats following LiCl exposure: ontogenetic influences. Developmental Psychobiology *16*, 13–22.

Roth, T., Raineki, C., Salstein, L., Perry, R., Sullivan-Wilson, T., Sloan, A., Lalji, B., Hammock, E., Wilson, D.A., Levitt, P., Okatanim F., Kabam H., and Sullivan, R.M. (2013). Neurobiology of secure infant attachment and attachment despite adversity: A mouse model. Genes, Brain and Behavior *12 (7)*, 673–680.

Salzen, E. (1970). Imprinting and environmental learning. In: Development and Evolution of Behavior, E.T.L. Aronson, D. Lehrman, and J. Rosenblatt, eds. (San Francisco: W.H. Freeman).

Sananes, C., and Campbell, B. (1989). Role of the central nucleus of the amygdala in olfactory heart rate conditioning. Behavioral Neuroscience *103*, 519–525.

Sanchez, M.M., Ladd, C.O., and Plotsky, P.M. (2001). Early adverse experience as a developmental risk factor for later psychopathology: evidence from rodent and primate models. Development and Psychopathology *13*, 419–449.

Sapolsky, R.M., and Meaney, M.J. (1986). Maturation of the adrenocortical stress response: neuroendocrine control mechanisms and the stress hyporesponsive period. Brain Research *396*, 64–76.

Sarro, E., Sullivan, R.M., and Barr, G. (2014). Unpredictable neonatal stress enhances adult anxiety and alters amygdala gene expression related to serotonin and GABA. Neuroscience *31;258*, 147–161.

Schettino, L. F., and Otto, T. (2001). Patterns of Fos expression in the amygdala and ventral perirhinal cortex induced by training in an olfactory fear conditioning paradigm. Behavioral Neuroscience *115*, 1257–1272.

Schiller D, Levy, I., Niv, Y.J., LeDoux, E., Phelps, E.A. (2008). From fear to safety and back – reversal of fear in the human brain. The Journal of Neuroscience *28*, 11517–11525.

Schore, A.N. (2001). The effects of early relational trauma on right brain development, affect regulation, and infant mental health. Inf. Mental Health J. *22*, 201–269.

Sevelinges, Y., Gervais, R., Messaoudi, B., Granjon, L., and Mouly, A.M. (2004). Olfactory fear conditioning induces field potential potentiation in rat olfactory cortex and amygdala. Learning and Memory *11*, 761–769.

Sevelinges, Y., Moriceau, S., Raineki, C., Forest, C., and Sullivan, R.M. (2011). Adult depression-like behavior and amygdala dysfunction rescued by odor previously paired with shock in infancy. Developmental Cognitive Neuroscience *1*, 77–87.

Shair, H.N., Masmela, J.R., Brunelli, S.A., and Hofer, M.A. (1997). Potentiation and inhibition of ultrasonic vocalization of rat pups: regulation by social cues. Developmental Psychobiology *30*, 195–200.

Shionoya, K., Moriceau, S., Lunday, L., Miner, C., Roth, T.L., and Sullivan, R.M. (2006). Development switch in neural circuitry underlying odor-malaise learning. Learning and Memory *13*, 801–808.

Shipley, M.T., Halloran, F.J., de la Torre, J. (1985). Surprisingly rich projection from locus coeruleus to the olfactory bulb in the rat. Brain Research *329*, 294–299.

Sigurdsson, T., Doyere, V., Cain, C.K., and LeDoux, J.E. (2007). Long-term potentiation in the amygdala: a cellular mechanism of fear learning and memory. Neuropharmacology *52*, 215–227.

Smotherman, W.P. (1982). Odor aversion learning by the rat fetus. Physiology & Behavior *29*, 769–771.

Smotherman, W.P., Hennessy, J.W., and Levine, S. (1976). Plasma corticosterone levels during recovery from LiCl produced taste aversions. Behavioral Biology *16*, 401–412.

Smotherman, W.P., and Robinson, S.R. (1985). The rat fetus in its environment: behavioral adjustments to novel, familiar, aversive, and conditioned stimuli presented in utero. Behavioral Neuroscience *99*, 521–530.

Smotherman, W.P., and Robinson, S.R. (1990). Rat fetuses respond to chemical stimuli in gas phase. Physiology & Behavior *47*, 863–868.

Spear, N. (1978). Processing Memories: Forgetting and Retention (Hillsdale, NJ: Erlbaum).

Spear, N.E., and Rudy, J.W. (1991). Tests of the ontogeny of learning and memory: issues, methods, and results. In: Developmental Psychobiology: New Methods and Changing Concepts, H.N. Shair, G.A. Barr, and et al., eds. (New York: Oxford University Press), pp. 84–113.

Stanley, W. (1962). Differential human handling as reinforcing events and as treatments influencing later social behavior in basenji puppies. Psychological Reports *10*, 775–788.

Stanton, M., and Levine, S. (1990). Inhibition of infant glucocorticoid stress response: specific role of maternal cues. Developmental Psychobiology *23*, 411–426.

Stanton, M.E., Wallstrom, J., and Levine, S. (1987). Maternal contact inhibits pituitary-adrenal stress responses in preweanling rats. Developmental Psychobiology *20*, 131–145.

Stehouwer, D.J., and Campbell, B.A. (1978). Habituation of the forelimb-withdrawal response in neonatal rats. Journal of Experimental Psychology: Animal Behavior Processes *4*, 104–119.

Stickrod, G., Kimble, D.P., and Smotherman, W.P. (1982). In utero taste/odor aversion conditioning in the rat. Physiology & Behavior *28*, 5–7.

Suchecki, D., Rosenfeld, P., and Levine, S. (1993). Maternal regulation of the hypothalamic-pituitary-adrenal axis in the infant rat: the roles of feeding and stroking. Brain Research: Developmental Brain Research *75*, 185–192.

Sullivan, R.M. (2012). Neurobiology of attachment to nurturing and abusive caregivers. Hastings Law Journal *63*, 1553–1571.

Sullivan, R.M., Brake, S.C., Hofer, M.A., and Williams, C.L. (1986a). Huddling and independent feeding of neonatal rats can be facilitated by a conditioned change in behavioral state. Developmental Psychobiology *19*, 625–635.

Sullivan, R.M., Hofer, M.A., and Brake, S.C. (1986b). Olfactory-guided orientation in neonatal rats is enhanced by a conditioned change in behavioral state. Developmental Psychobiology *19*, 615–623.

Sullivan, R.M., Landers, M., Yeaman, B., and Wilson, D.A. (2000a). Good memories of bad events in infancy. Nature *407*, 38–39.

Sullivan, R.M., and Leon, M. (1986). Early olfactory learning induces an enhanced olfactory bulb response in young rats. Brain Research *392*, 278–282.

Sullivan, R.M., Stackenwalt, G., Nasr, F., Lemon, C., and Wilson, D.A. (2000b). Association of an odor with activation of olfactory bulb noradrenergic beta-receptors or locus coeruleus stimulation is sufficient to produce learned approach responses to that odor in neonatal rats. Behavioral Neuroscience *114*, 957–962.

Sullivan, R.M., and Wilson, D.A. (1991). Neural correlates of conditioned odor avoidance in infant rats. Behavioral Neuroscience *105*, 307–312.

Sullivan, R.M., Wilson, D.A., Wong, R., Correa, A., and Leon, M. (1990). Modified behavioral and olfactory bulb responses to maternal odors in preweanling rats. Brain Research: Developmental Brain Research *53*, 243–247.

Sullivan, R.M., Zyzak, D.R., Skierkowski, P., and Wilson, D.A. (1992). The role of olfactory bulb norepinephrine in early olfactory learning. Brain Research: Developmental Brain Research *70*, 279–282.

Suomi, S.J. (1997). Early determinants of behaviour: evidence from primate studies. British Medical Bulletin *53*, 170–184.

Suomi, S.J. (2003). Gene-environment interactions and the neurobiology of social conflict. Annals of the New York Academy of Sciences *1008*, 132–139.

Takahashi, L.K. (1994). Organizing action of corticosterone on the development of behavioral inhibition in the preweanling rat. Brain Research: Developmental Brain Research *81*, 121–127.

Tao, X., Finkbeiner, S., Arnold, D.B., Shaywitz, A.J., and Greenberg, M.E. (1998). Ca2+ influx regulates BDNF transcription by a CREB family transcription factor-dependent mechanism. Neuron *20*, 709–726.

Teicher, M.H., and Blass, E.M. (1977). First suckling response of the newborn albino rat: the roles of olfaction and amniotic fluid. Science *198*, 635–636.

Teicher, M.H., Flaum, L.E., Williams, M., Eckhert, S.J., and Lumia, A.R. (1978). Survival, growth and suckling behavior of neonatally bulbectomized rats. Physiology & Behavior *21*, 553–561.

Terry, L.M., and Johanson, I.B. (1996). Effects of altered olfactory experiences on the development of infant rats' responses to odors. Developmental Psychobiology *29*, 353–377.

Thompson, B.L., Erickson, K., Schulkin, J., and Rosen, J.B. (2004). Corticosterone facilitates retention of contextually conditioned fear and increases CRH mRNA expression in the amygdala. Behavioural Brain Research *149*, 209–215.

Thompson, J.V., Sullivan, R.M., and Wilson, D.A. (2008). Developmental emergence of fear learning corresponds with changes in amygdala synaptic plasticity. Brain Research *1200C*, 58–65.

Touzani, K., and Sclafani, A. (2005). Critical role of amygdala in flavor but not taste preference learning in rats. European Journal of Neuroscience *22*, 1767–1774.

Upton, K.J., and Sullivan, R.M. (2010). Defining age limits of the sensitive period for attachment learning in rat pups. Developmental Psychobiology *52(5)*, 453–464.

van Oers, H.J., de Kloet, E.R., Li, C., and Levine, S. (1998). The ontogeny of glucocorticoid negative feedback: influence of maternal deprivation. Endocrinology *139*, 2838–2846.

van Oers, H.J., Kloet, E.D., Whelan, T., and Levine, S. (1998). Maternal deprivation effect on the infant's neural stress markers is reversed by tactile stimulation and feeding but not by suppressing corticosterone. Neuroscience *18*, 10171–10179.

Walker, C., Sapolsky, R., Meaney, M., Vale, W., and Rivier, C. (1986). Increased pituitary sensitivity to glucocorticoid feedback during the stress nonresponsive period in the neonatal rat. Endocrinology *119*, 1816–1821.

Weldon, D.A., Travis, M.L., and Kennedy, D.A. (1991). Posttraining D1 receptor blockade impairs odor conditioning in neonatal rats. Behavioral Neuroscience *105*, 450–458.

Welker, W. (1964). Analysis of sniffing of the albino rat. Behaviour *12*, 223–244

Wiedenmayer, C.P., and Barr, G.A. (2001). Developmental changes in c-fos expression to an age-specific social stressor in infant rats. Behavioural Brain Research *126*, 147–157.

Wilson, D.A., and Sullivan, R.M. (1994). Neurobiology of associative learning in the neonate: early olfactory learning. Behavioral and Neural Biology *61*, 1–18.

Wilson, D.A., Sullivan, R.M., and Leon, M. (1987). Single-unit analysis of postnatal olfactory learning: modified olfactory bulb output response patterns to learned attractive odors. Journal of Neuroscience *7*, 3154–3162.

Woo, C.C., Coopersmith, R., and Leon, M. (1987). Localized changes in olfactory bulb morphology associated with early olfactory learning. Journal of Comparative Neurology *263*, 113–125.

Yuan, Q., Harley, C.W., Darby-King, A., Neve, R.L., and McLean, J.H. (2003). Early odor preference learning in the rat: bidirectional effects of cAMP response element-binding protein (CREB) and mutant CREB support a causal role for phosphorylated CREB. Journal of Neuroscience *23*, 4760–4765.

Yuan, Q., Harley, C.W., McLean, J.H., and Knopfel, T. (2002). Optical imaging of odor preference memory in the rat olfactory bulb. Journal of Neurophysiology *87*, 3156–3159.

Zhang, J.J., Okutani, F., Inoue, S., and Kaba, H. (2003). Activation of the cyclic AMP response element-binding protein signaling pathway in the olfactory bulb is required for the acquisition of olfactory aversive learning in young rats. Neurosci *117*, 707–713.

M. Justin Kim, Rebecca A. Loucks, Amy L. Palmer,
Annemarie C. Brown, Kimberly M. Solomon,
Ashley N. Marchante, Paul J. Whalen
Department of Psychological and Brain Sciences, Dartmouth College, 6207
Moore Hall, Hanover, NH 03755, USA

The Structural and Functional Connectivity of the Amygdala: From Normal Emotion to Pathological Anxiety

Abstract

The dynamic interactions between the amygdala and the medial prefrontal cortex (mPFC) are usefully conceptualized as a circuit that both allows us to react automatically to biologically relevant predictive stimuli as well as regulate these reactions when the situation calls for it. In this review, we will begin by discussing the role of this amygdala–mPFC circuitry in the conditioning and extinction of aversive learning in animals. We will then relate these data to emotional regulation paradigms in humans. Finally, we will consider how these processes are compromised in normal and pathological anxiety. We conclude that the capacity for efficient crosstalk between the amygdala and the mPFC, which is represented as the strength of the amygdala–mPFC circuitry, is crucial to beneficial outcomes in terms of reported anxiety.

1. Introduction

Accurate evaluation of and response to potentially life threatening or sustaining events are hallmarks of biologically relevant learning in animals and humans. In response to cues of threat, rodents exhibit a distinctive "freezing" or somatomotor arrest behavior. This behavior is critical in a natural environment in which movement may attract a predator to its location. Humans show a similar freezing response to a potentially threatening situation (Roelofs et al., 2010). Rather than having to avoid predators, humans might more ordinarily show such a response when having to speak in front of a large audience. In such threatening instances, performance would be facilitated if able to override the initial freezing behavior.

In psychological terms, instinctive reactions to threat and subsequent regulatory responses are often referred to as bottom-up and top-down processes, respectively. The interplay between these two processes is exemplified by the following example: upon encountering a snake at a zoo, an initial reaction is driven by its appearance (i.e., bottom-up saliency), but the response is then implicitly controlled by the determination that the snake presents no immediate danger because it is behind a sheet of Plexiglas (i.e., top-down control). Of course, the context is critical since the same snake encountered in a field would evoke an initial freezing response followed by a very different type of top-down control in the form of running (or screaming in some cases). Thus, interactions between bottom-up and top-down processes will determine the adaptiveness of behavior in a given situation.

This conceptualization may be directly applicable to clinical research, as the interaction between these bottom-up and topdown processes is hypothesized to be impaired in psychiatric illnesses – and here we will focus on the anxiety disorders. For example, in specific phobias, perhaps a failure to employ top-down control mechanisms allows initial bottom-up responses to intrude on normal cognitive functioning. Alternatively, it may be the case that the initial bottom-up reactions are so potent and exaggerated that even a normally functioning top-down regulatory system cannot keep these responses in check. Individual differences in the function and structure of this circuitry can also explain differences in normal levels of anxiety.

Numerous studies have highlighted the critical role of the amygdala and the mPFC in behavioral phenomena that involve competition between bottom-up and top-down processes, including fear conditioning and extinction (Bishop, 2007; Ochsner and Gross, 2005; Quirk and Beer, 2006). Critically, it is believed that the mPFC regulates and controls amygdala output and the accompanying behavioral phenomena (Bishop, 2007; Ochsner and Gross, 2005; Quirk and Beer, 2006). The reciprocal relationship between the amygdala and the mPFC strongly suggests the need to investigate these brain regions as one circuit, rather than studying them separately. That is, while numerous studies have assessed the separate contributions that the amygdala and mPFC make to reactivity and regulation, respectively (Bishop et al., 2004; Simmons at al., 2008; Simpson et al., 2001; Straube et al., 2009), more recent studies suggest that the structural and functional connectivity between these two regions is a better predictor of these outcomes than the activity of either region alone (Pezawas et al., 2005; Kim and Whalen, 2009; Kim et al., 2010a). The idea here is that the stronger the coupling between the amygdala and the mPFC, the better the behavioral outcome in terms of reported anxiety.

2. The structural and functional connectivity of the human amygdala and prefrontal cortex

2.1. Structural neuroanatomy of amygdala–mPFC circuitry

The amygdala is an almond-shaped brain structure that resides in the medial temporal lobe of the brain (Aggleton, 1992; Whalen and Phelps, 2009). Its structure is comprised of many subnuclei, including the basolateral nuclei (BLA) and the central nucleus (Ce), which have distinct anatomical connections with other brain regions that serve different functions. Comprehensive descriptions of the anatomical connections of the amygdala exist elsewhere (Amaral et al., 1992; Freese and Amaral, 2009). Here, we focus on the connectivity between the amygdala and the prefrontal cortex, especially the mPFC. The mPFC can be roughly divided into two subregions, relative to the genu of the corpus callosum–dorsal mPFC (dmPFC) and the ventral mPFC (vmPFC). Broadly defined, the dmPFC includes the supragenual anterior cingulate and the medial frontal gyrus, whereas the vmPFC includes the subgenual anterior cingulate, ventromedial prefrontal and medial orbitofrontal cortex (Fig. 1).

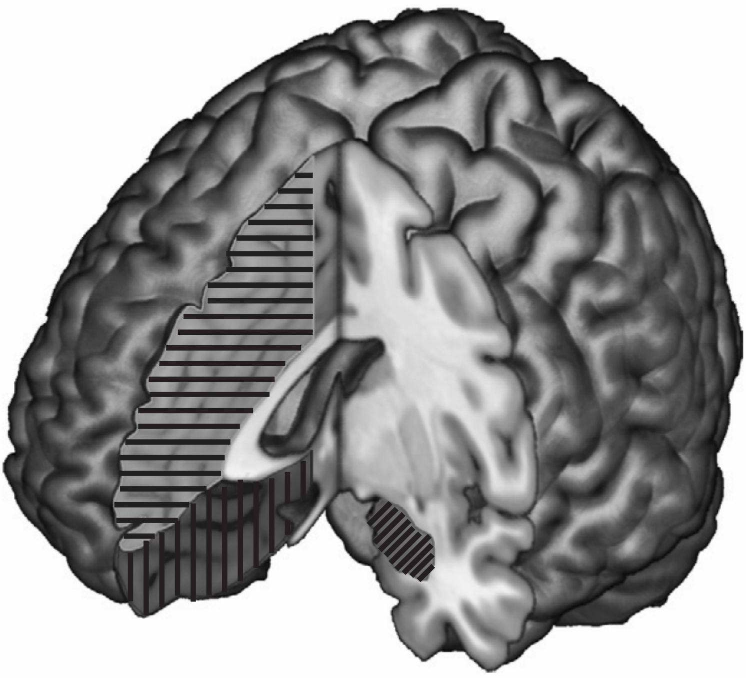


Fig. 1. Structural magnetic resonance image of the human brain highlighting the major components of the amygdala-prefrontal circuitry: amygdala (diagonal lines), ventromedial prefrontal cortex (vertical lines), and dorsomedial prefrontal cortex (horizontal lines) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

Most of the known facts about the anatomical connections of the amygdala–mPFC circuitry are derived from animal studies, especially non-human primates. This is because the invasive nature of the methods that are used to investigate brain connections – such as lesion and tracing studies – are difficult to employ in humans. Data from non-human primate brains show that the majority of the afferent fibers to the amygdala originate in the orbitofrontal cortex and the mPFC, and these projections are denser and heavier from the caudal compared to the rostral aspects of these prefrontal areas (Aggleton at al., 1980; Carmichael and Price, 1995; Ghashghaei and Barbas, 2002; Ghashghaei et al., 2007; Leichnetz and Astruc, 1976; Stefanacci and Amaral, 2002). In turn, the amygdala sends efferent projections to these orbitofrontal and mPFC regions (Carmichael and Price, 1995; Ghashghaei and Barbas, 2002; Ghashghaei et al., 2007; Amaral and Price, 1984; Barbas and De Olmos, 1990), and interestingly these projections are heavier than the reciprocal cortical afferents (Ghashghaei et al., 2007; LeDoux, 1996). Most of the amygdala-prefrontal connections are concentrated in the BLA, as opposed to Ce. Based on animal studies of fear conditioning and extinction, mPFC input to the BLA as well as the intercalated cells (adjacent to the BLA) is responsible for inhibiting amygdala output by regulating BLA inputs to the Ce (see Section 3, for details).

In humans, the structural connections of the amygdala have been investigated more recently using non-invasive neuroimaging methods such as diffusion tensor imaging (DTI). This imaging method takes advantage of the fact that the movement of water molecules in the brain differs in different types of brain tissue. To elaborate, in an unrestricted tissue environment, such as in the ventricles of the brain, water molecules show isotropic diffusion in all directions equally. Importantly, however, the movement of water molecules is greatly restricted in myelinated axons – that is, water molecules in white matter tend to move in a single direction along the myelinated axons. Thus DTI is a method optimized to assess white matter fiber tracts in the brain. There are two different types of information about which DTI can inform us – (1) the orientation of white matter fiber tracts, and (2) the strength or integrity of white matter fiber tracts. The former can be accomplished by using fiber tracking or tractography to measure the direction of water diffusion (Mori at al., 1999). The latter can be calculated by measuring the degree of anisotropic diffusion (Basser and Pierpaoli, 1996). Specifically, normalized measures such as fractional anisotropy can be computed for each brain voxel and used to index the structural integrity of the measured white matter fiber tracts (Basser and Pierpaoli, 1996). A number of studies have utilized these methods to identify an amygdala-prefrontal pathway in the human with a specific focus on connectivity with the dorsal and ventral aspects of the mPFC (Bracht at al., 2009; Croxson at al., 2005; Johansen-Berg at al., 2008).

2.2. Functional neuroanatomy of the amygdala–mPFC circuitry

The functional counterpart of structural connectivity – functional brain connectivity – can also be used to investigate amygdala-prefrontal interactions. This method is optimal for understanding the relationship between spatially remote brain regions by assessing brain activity across time. Analyses of functional brain connectivity can be defined in two ways: (1) functional connectivity and (2) effective connectivity (Friston, 1994). Functional connectivity is simply a measure of the temporal correlation of brain activity in two or more regions, whereas effective connectivity seeks to reveal the directional effect of one neuronal system exerted over another (Friston, 1994). By definition, functional connectivity is purely correlational in nature, and provides no information regarding the directionality of how one brain region affects another. In contrast, effective connectivity attempts to explain the causal relationship between the interactions of different brain regions, relying on more advanced statistical modeling methods such as structural equation modeling (McIntosh and Gonzalez-Lim, 1994), psychophysiological interactions (Friston et al., 1997), and dynamic causal modeling (Friston et al., 2003). Based on the extensive anatomical connections between the amygdala and mPFC shown in human and non-human primates, a number of investigations have used these functional and effective connectivity measures to assess the strength of amygdala–mPFC coupling and its relationship with behavioral outcomes (Pezawas et al., 2005; Banks et al., 2007; Kim et al., 2003).

2.3. Amygdala–mPFC circuitry at rest

The majority of the aforementioned functional connectivity studies were task-based, meaning that brain activity was measured in response to particular stimulus presentations and/or task instructions. For example, amygdala activity measured as subjects viewed surprised facial expressions was used to identify mPFC activity that predicted subjects' interpretations of these faces as either positively or negatively valenced (Kim et al., 2004). Higher mPFC activity predicted lesser amygdala activity and more positive ratings of these expressions. Recently, a growing body of functional neuroimaging studies has emerged investigating brain activity and connectivity at rest, in the absence of presented stimuli or task instructions. The "resting state" can be investigated using fMRI by measuring spontaneous, slow (<0.1 Hz) fluctuations in the brain that occur over time (Biswal et al., 1995). Identifying the brain's functional networks at rest can help investigators better understand how brain regions are coupled prior to a task. In fact, investigators have mapped highly detailed resting state functional networks associated with specific brain regions such as the anterior cingulate cortex (Margulies et al., 2007), the striatum (Di Martano et al., 2008), and the amygdala (Roy et al., 2009). Resting state functional connectivity analyses have also been used to identify distinct neural networks based on specific neurophysiological phenomena, such as repetition priming (Wig et al., 2009). Additionally, recent studies have linked the strength of resting state functional connectivity with individual differences in behavioral outcomes, such as behavioral performance on cognitive tasks (Kelly et al., 2008; Stevens et al., 2009), autistic traits (Di Martino et al., 2009) and reported anxiety (Kim et al., 2010a). Using this method, researchers can test whether the degree to which the amygdala is coupled with mPFC regions at rest influences how well a person regulates their emotional responses when challenged with stimulus presentations during a particular task.

In summary, there are a number of non-invasive methods to investigate the strength of amygdala–mPFC connectivity in vivo. Structural connectivity can be assessed by utilizing DTI, and functional brain connectivity can be evaluated through analyses of functional and effective connectivity. Furthermore, understanding how brain regions are connected with one another at rest may provide new insights that elucidate how the amygdala–mPFC circuitry is engaged during a task.

3. Amygdala-prefrontal circuitry and fear conditioning and extinction

Studies of the non-human animal amygdala have shown that sensory information received by the BLA is then passed to the Ce (Davis and Shi, 2000). Though outputs exist at the level of the BLA, a majority of outputs originate from the Ce (Kapp et al., 1994). The Ce projects directly to the hypothalamus and brain stem nuclei that drive autonomic and somatomotor responding (Holstege et al., 1996). The Ce also projects to all major neuromodulatory systems including dopaminergic, cholinergic, serotonergic and noradrenergic systems (Kapp et al., 1994). Thus, while direct projections can primarily affect physiological and motor responses, these neuromodulatory projections can serve to globally, nonspecifically and instantaneously effect neuronal excitability across the brain. Such changes could serve to induce a state of heightened vigilance rendering the organism a more efficient consumer of information in biologically relevant learning situations (Kapp et al., 1994). One such situation involves the acquisition and expression of learned responses through classical conditioning (Davis and Whalen, 2001; LeDoux, 2000). For example, in a typical aversive conditioning paradigm, subjects learn that a previously neutral stimulus (e.g., tone) predicts the occurrence of an unconditioned stimulus (US; e.g., electric shock), thereby acquiring the value of a conditioned stimulus (CS) which now elicits a conditioned response (CR; e.g., freezing) to the CS that was previously reserved for the US (Kim et al., 1992; Quirk et al., 1995). The generation of these CS–US associations and their behavioral expressions are known to be amygdala-dependent since manipulations of this structure block or retard such learning (Fanselow and LeDoux, 1999; Rogan et al., 1997).

A reversal of this classical conditioning procedure is known as extinction – suppressing previously learned CS–US associations (Quirk, 2002; Rescorla, 2001). The inhibition of CS–US associations can be achieved by top-down regulatory input from the mPFC to the BLA (Milad and Quirk, 2002). This process is supported by the existence of amygdala–mPFC connectivity that allows direct, reciprocal communication (Amaral et al., 1992; Ghashghaei et al., 2007; Milad and Quirk, 2002). For example, electric stimulation of the mPFC resulted in the inhibition conditioned responses, emulating the effects of extinction in the rat (Milad and Quirk, 2002). In humans, greater cortical thickness of the vmPFC was associated with better behavioral performance during extinction recall (Hartley et al.; Milad et al., 2005). Similar findings have also been demonstrated in humans using fMRI, highlighted by increased vmPFC activity during successful extinction of learned US-CS associations (Delgado et al., 2008; Milad et al., 2007; Phelps et al., 2004). Interestingly, a study using functional connectivity methods (Delgado et al., 2008) showed that the amygdala and the vmPFC were functionally coupled during the entire course of the experiment, which included a combination of fear extinction and emotion regulation tasks. Thus, these structural and functional findings highlight the importance of amygdala–mPFC interactions for the regulation and inhibition necessary for extinction learning and/or memory.

4. Amygdala-prefrontal circuitry and emotion regulation

The ability to regulate our emotions is essential in our everyday lives, and successful emotion regulation begets beneficial outcomes in many social situations. Emotion regulation is a classic example of how top-down and bottom-up processes compete and interact to produce optimal (or counterproductive) behavioral outcomes. For example, one's instinctive reaction to a frightening scene in a horror movie may include an urge to scream and/or run out of the room. Normally, this bottom-up reaction is controlled by a top-down intervention (e.g., reminding oneself that this is only a movie). Taking the scenario described above into account, it would not be too difficult to imagine that individuals may employ different strategies to achieve such emotion regulation.

To date, studies investigating the neural basis of emotion regulation have primarily examined two distinctive means of emotion regulation – by simply suppressing what one is feeling (i.e., suppression), or by cognitively reevaluating the stimulus that is evoking the emotion (i.e., reappraisal) (Ochsner and Gross, 2005;). Not surprisingly, emotion regulation and the extinction of fear conditioning are suggested to have overlapping underlying neural mechanisms, since the essence of both processes involves reevaluating biologically relevant stimuli (Quirk and Beer, 2006; Hartley and Phelps, 2010). Like extinction, it is useful to assume that during emotional regulation the prefrontal cortex exerts control over the amygdala in response to an emotional challenge (Ochsner and Gross, 2005; Davidson at al., 2000). Based on this framework, numerous functional neuroimaging studies have demonstrated increased prefrontal activity and concomitant decreased amygdala activity during successful emotion regulation (Delgado et al., 2008; Hariri et al., 2000; Hariri et al., 2003; Lieberman et al., 2007; Ochsner et al., 2002; Phan et al., 2005; Erk et al., 2010; Urry et al., 2006; Wager et al., 2008). Unlike extinction, many emotion regulation studies point to the ventral and dorsal lateral prefrontal cortex (vlPFC and dlPFC, respectively), in addition to the ventral medial prefrontal cortex, as critical for regulating amygdala activity (Delgado et al., 2008; Lieberman et al., 2007; Ochsner et al., 2002; Erk et al., 2010; Wager et al., 2008). In general, studies have shown largely overlapping prefrontal-amygdala activity to suppression and reappraisal strategies, which was characterized by decreased activity of the amygdala and increased activity of the prefrontal cortex – usually including both medial and lateral PFC (Ochsner and Gross, 2005; Ochsner et al., 2002). Furthermore, the frequency of using reappraisal to regulate emotion in everyday life has been shown to be related to decreased amygdala activity, increased prefrontal and parietal activity (Drabant et al., 2009), and greater vmPFC volume (Welborn et al., 2009). In addition, during an affect labeling task (i.e., putting emotions into words), which can be regarded as a specific form of emotion regulation (Lieberman et al., 2007; Foland-Ross et al., 2010), diminished amygdala activity was again associated with greater activity of the mPFC (Hariri et al., 2000; Hariri et al., 2003), vlPFC (Lieberman et al., 2007) and also with increased cortical thickness of the vmPFC (Foland-Ross et al., 2010). These findings provide functional and structural evidence for shared neural mechanisms during different types of emotion regulation strategies.

These neuroimaging findings lead us to an interesting question – does emotion regulation share similar underlying neural mechanisms with more classic forms of cognitive control? Or are there unique brain circuitries recruited by these distinct emotion regulation processes? According to a cognitive control model of emotion regulation (Ochsner and Gross, 2005), the neural representation of emotion regulation can be summarized as interactions between prefrontal (both dlPFC and mPFC) and ACC systems and their influence on subcortical systems, including the amygdala. The two major types of emotion regulation–suppression and reappraisal – have yielded similar results in terms of brain activations, and how the prefrontal and anterior cingulate cortices interact is strikingly similar to other top-down control mechanisms that do not involve emotional processing, such as cognitive control (Botvinick et al., 1999; Botvinick et al., 2004; Miller and

Cohen, 2001). However, another study has shown that emotion regulation through moodincongruent autobiographical recall recruits the ventral mPFC and vlPFC, but not dlPFC, implying that activations of neural circuitry depend on the type of emotion regulation being used (Cooney et al., 2007). Taken together, we can tentatively conclude that while emotion regulation does to some extent share similar neural circuitry with cognitive control, it also recruits unique brain regions, such as the vlPFC. Other forms of top-down control processes, such as regulation of appetitive behaviors, attitudes or prejudice, have also shown to use an overlapping amygdala-prefrontal circuitry (Heatherton and Wagner, 2011).

Recent findings raise the possibility that a more efficient cross-talk between the amygdala and the prefrontal cortex begets a better ability to regulate one's emotions. Supporting this idea, the strength of amygdala–mPFC coupling was quantified by computing the functional connectivity between these two areas and comparing this connectivity with how effectively participants regulated their emotions (Banks et al., 2007). It was found that the functional coupling between the amygdala–mPFC was strengthened during reappraisal, and that the degree of this functional coupling was positively correlated with the self-reported effectiveness of emotion regulation (Banks et al., 2007). A selective increase in the functional coupling of the amygdala with the vmPFC and dlPFC during emotion regulation has also been reported (Erk et al., 2010), highlighting the importance of efficient communication between the amygdala and the prefrontal cortex in successful top-down control of emotion. Further, functional coupling of the amygdala and vmPFC at rest, predicts beneficial outcomes in terms of reported anxiety (Kim et al., 2010a). Future studies linking the effective success of emotion regulation strategies and the structural and functional connectivity of the amygdala–mPFC circuitry might provide a better understanding of the neural correlates of these emotion regulatory processes.

5. Amygdala-prefrontal circuitry and the interpretation of emotionally ambiguous facial expressions

In humans, patients with selective amygdala lesions have displayed deficits in processing the facial expressions of fear (Adolphs et al., 1995), leading to numerous functional neuroimaging studies using presentations of fearful faces to probe amygdala activity (Kim and Whalen, 2009; Breiter et al., 1996; Hariri et al., 2002; Kim et al., 2010b; Morris et al., 1996; Whalen et al., 1998; Whalen et al., 2001). These studies have shown that the amygdala is particularly responsive to fearful faces compared to other expressions (Fusar-Poli et al., 2009), including angry, happy, and neutral (Kim and Whalen, 2009; Breiter et al., 1996; Hariri et al., 2002; Morris et al., 1996; Whalen et al., 1998; Whalen et al., 2001), except for one report (Fitzgerald et al., 2006). The affinity of the human amygdala for fearful faces compared to these other expressions, provides insights into amygdala function. For example, since the amygdala is more responsive to fearful faces compared to angry faces, which embody a direct threat, it has been suggested that one function of the amygdala is to augment cortical function through the major neuromodulatory centers to assist in the resolution of predictive uncertainty (Whalen et al., 2001; Whalen and Fear, 1998). That is, the inherent ambiguity of fearful faces in that they predict the increased probability of threat without providing information about its nature or location – leads to selective activation of the amygdala (Whalen et al., 2001; Whalen and Fear, 1998).

Given that the amygdala plays a major role in the resolution of predictive uncertainty associated with fearful faces, surprised facial expressions provide a particularly important comparison expression. Indeed, there is evidence that surprise may be the second-most compromised expression in patients with selective amygdala damage, following fear (Adolps et al., 1994). Fearful and surprised faces have common facial features (e.g., eye-widening), and both expressions indicate the detection of a significant, but unknown, eliciting event (Kim et al., 2004). Surprised faces are particularly interesting because, unlike fear, they do not predict the valence of the unknown eliciting event. Indeed, research has shown that surprised faces can be interpreted as either positive or negative in nature (Kim et al., 2003; Kim et al., 2004; Neta et al., 2009). Previous research has shown that when individuals make valence judgments of surprised faces, ratings reflect individual differences in one's positivity/ negativity bias (Kim et al., 2003; Neta et al., 2009) and these differences are mirrored by a distinct pattern of brain activity, which critically involves the vmPFC as well as the amygdala (Kim et al., 2003). Specifically, decreased amygdala activity accompanied by increased vmPFC activity was observed in people who interpreted surprised faces as positive, with the reverse brain pattern seen in those who interpreted surprised faces as negative (Kim et al., 2003). The role of the vmPFC in resolving the emotional ambiguity (i.e., the valence of a given surprised face) could be understood as a top-down regulatory input to the amygdala, much akin to the neural mechanism of fear extinction or emotion regulation (Quirk and Beer, 2006). Indeed, greater vmPFC activity predicts both (a) more positive ratings of surprise and (b) more positive interpretations of an extinguished tone (i.e., tone now predicts no shock) (Oler et al., 2009). In a subsequent study (Kim et al., 2004), positive and negative sentences (e.g., "He just lost $500" or "He just found $500") were used to provide contextual information for the presented surprised faces, in order to see how brain activity was influenced by information that provided clear resolution to the source ambiguity problem associated with surprised faces. Again, data from this study showed that greater amygdala response to negative versus positive faces was accompanied by diminished vmPFC activity, and interestingly greater vlPFC activity (Kim et al., 2004). Thus, similar medial prefrontal-amygdala regions were activated when a context was provided (i.e., valence of the surprised faces were determined by the experimental condition), compared to when the subjects had to judge the valence of the surprised faces themselves – but additional lateral prefrontal regions were recruited in the contextually mediated condition (Kim et al., 2003).

In summary, data from these experiments collectively suggest that using emotionally ambiguous stimuli such as surprised faces instigates a competition between top-down and bottom-up processes. This engages the amygdala–mPFC circuitry and the balance of activity within this circuit reflects the resolution of the inherent ambiguity of the perceived surprised faces.

6. Amygdala-prefrontal circuitry and anxiety within the normal range

Anxiety is characterized by chronic, nonspecific apprehension and arousal related to the potential occurrence of future threat (Eysenck, 1992; Rosen and Schulkin, 1998). Neurobiological theories of anxiety have highlighted the central role of the amygdala in the generation and experience of the fear that can give rise to anxiety (Davis and Whalen, 2001; LeDoux, 2000), and fear extinction investigations in animals support such theories (Davis and Whalen, 2001; LeDoux, 2000). Similar to the inhibition of previously conditioned fear responses during fear extinction, reduced anxiety is associated with the top-down regulation of amygdala activity by the mPFC (Pezawas et al., 2005; Hariri et al., 2003; Hare et al., 2008). Findings from anatomical investigations of amygdala connectivity (Amaral et al., 1992; Ghashghaei et al., 2007) and fear extinction studies in animals (Milad and Quirk, 2002) emphasize the top-down and bottom-up interactions between the amygdala and mPFC regions in anxiety. To put it another way, efficient crosstalk between the amygdala and the mPFC produces a better outcome in terms of controlling anxiety.

Consistent with this framework, a number of functional neuroimaging studies in humans have shown elevated amygdala activity in highly anxious but otherwise healthy individuals (Bishop et al., 2004b; Dickie and Armony, 2008; Etkin et al., 2004; Somerville et al., 2004; Stein et al., 2007). For example, increased amygdala activity to unattended fearful faces was associated with higher levels of self-reported anxiety (Bishop et al., 2004), although this effect may be more prominent in women than men (Dickie and Armony, 2008). Using backward masking, an experimental paradigm that has been shown to reliably evoke human amygdala activity and mitigate subjective awareness of fearful face stimuli (e.g., (Kim et al., 2010b; Whalen et al., 1998)), it has been reported that increased amygdala activity was linked to elevated anxiety levels (Etkin et al., 2004). It is worth noting that in these studies, the relationship between increased amygdala activity and anxiety was evident when the subjects were not attending to or were unaware of the stimuli, not when they were attending to or aware of them. This raises the possibility that attention or awareness may be an important factor that interacts with amygdala activation and subsequent reported anxiety. Anxiety was not only associated with elevated amygdala activity to threat-related stimuli (e.g., fearful faces, emotionally negative pictures), but was also associated with increased activity to non threat-related stimuli (neutral faces; (Somerville et al., 2004)), suggesting that amygdala activity may reflect greater anxiety levels even in the absence of clear threat.

Other evidence from the human neuroimaging literature shows that altered mPFC activity is associated with anxiety (Bishop et al., 2004; Simmons at al., 2008; Simpson et al., 2001; Straube et al., 2009; Hare et al., 2008). Although changes in mPFC activity has been consistently reported in anxiety research, the spatial location of that activity (i.e., whether it is dorsal or ventral) varies across studies. Depending on the experimental task, different studies have reported divergent results (for review, see (Bishop, 2007)) – for example, anxiety reduced activity of the vmPFC in one study (Simpson et al., 2001), and dmPFC in another (Bishop et al., 2004a). More recently, a number of studies have shown that higher levels of anxiety are associated with both decreased vmPFC activity and increased dmPFC activity (Simmons at al., 2008; Straube et al., 2009), suggesting differential roles for these mPFC subregions in anxiety.

Based on the findings highlighting the importance of both the amygdala and mPFC regions in anxiety, a number of studies have investigated the amygdala–mPFC circuitry in conjunction with anxiety using functional and structural connectivity measures (Pezawas et al., 2005; Pezawas et al., 2005; Kim and Whalen, 2009; Kim et al., 2010). For example, individuals with anxious temperaments had weaker functional coupling between the amygdala and the vmPFC during a task that involved matching fearful and angry faces (Pezawas et al., 2005). Using DTI, it was demonstrated that the structural integrity of an amygdala–vmPFC pathway was compromised in the participants who exhibited high trait anxiety (Kim and Whalen, 2009). Furthermore, studies employing resting state functional connectivity methods have shown that the strength of the coupling between the amygdala and mPFC at rest predicted self-reported levels of anxiety (Kim et al., 2010a; Seeley et al., 2007), where a positive correlation between the amygdala and vmPFC predicted beneficial outcomes in terms of reported anxiety (Kim et al., 2010a)). In this study the dmPFC showed an opposite relationship to that observed between the amygdala and vmPFC–dmPFC activity at rest that was negatively correlated with amygdala activity predicted lower levels of anxiety. This latter finding is complemented by a recent task-based fMRI study in which amygdala–dmPFC functional connectivity strength was positively correlated with neuroticism (an anxiety-related personality trait characterized by a bias to interpret normal situations as harmful and threatening; (Costa and McCrae, 1997) during viewing emotionally negative faces (Cremers et al., 2010). Taken together, these data suggest that the strength of amygdala–mPFC functional connectivity during rest may represent efficient crosstalk between the two brain regions, which may be responsible for abolishing the generation of anxious states (Kim et al., 2010a). This idea is consistent with findings from task-based functional connectivity (Pezawas et al., 2005). Findings from these studies all fit well with the idea that efficient crosstalk between the amygdala and the mPFC, perhaps particularly the vmPFC, is critically involved in lowering anxiety levels.

7. Amygdala-prefrontal circuitry and pathological anxiety

Taking individual differences in normal fluctuations in anxiety as our starting point, disrupted bottom-up and top-down emotional and cognitive processes are thought to be a crucial component of symptomology in pathological anxiety. This model suggests an imbalance between the amygdala and the prefrontal cortex, which is typically characterized by hyperactivity of the amygdala and hypoactivity of the prefrontal cortex (Rauch et al., 2006; Shin and Handwerger, 2009).

7.1. Social anxiety disorder

A prevalent subtype of the anxiety disorders (Kessler et al., 1994), social anxiety disorder (SAD) is characterized by intense anxiety during social situations in which the person is exposed to unfamiliar people (APA, 1994). To this end, emotional facial expressions provide a particularly useful paradigm for studying SAD, which is thought to involve exaggerated emotional reactivity to social stimuli and the inability to regulate these responses (Clark and McManus, 2002; Rapee and Heimberg, 1997). Individuals with SAD reliably showed elevated amygdala reactivity when viewing “harsh” faces (facial expressions displaying anger, contempt, or a combination of both) (Goldin et al., 2009; Stein et al., 2002; Yoon et al., 2007), and even neutral faces (113), compared to healthy individuals in fMRI studies. Amygdala reactivity was positively correlated with symptom severity and/or trait anxiety in SAD patients, further demonstrating the neurobiological significance of the amygdala in SAD (Stein et al., 2002; Cooney et al., 2006; Shah et al., 2009). These individuals also had exaggerated amygdala reactivity to pictures of emotionally negative scenes (i.e., unpleasant and/or aversive) suggesting abnormal neural activity during general emotional, not just social, processing in SAD (Shah et al., 2009). Direct examination of neural activity during emotion regulation demonstrated that SAD patients fail to recruit the mPFC (Goldin et al., 2009), implying that the connectivity of the amygdala–mPFC circuitry is disrupted in SAD. A resting state fMRI study showed that SAD patients had markedly reduced functional connectivity between the left amygdala and the medial orbitofrontal cortex (Hahn et al., 2011), corroborating the previous findings assessing a normal range of anxiety (Kim et al., 2010a). In addition, state anxiety levels in SAD subjects was inversely correlated with the functional connectivity strength between the amygdala and the medial orbitofrontal cortex, further validating the central role of the amygdala–mPFC circuitry in SAD (115). In addition to these functional abnormalities, SAD patients exhibited compromised structural integrity of the uncinate fasciculus (Phan et al., 2009), a major white matter fiber tract that is known to connect the amygdala and the orbitofrontal cortex (Ebeling and von Cramon, 1992). Each of these studies provides examples of SAD patients' failure to recruit the proper cognitive regulatory circuits in the brain, and that the functional abnormalities in these circuits may be attributable, in part, to white matter microstructural problems caused by the pathophysiology of SAD.

7.2. Posttraumatic stress disorder

Posttraumatic stress disorder (PTSD) is a stress-induced anxiety disorder characterized by re-experiencing the traumatic event, avoidance of stimuli associated with the trauma, and more generalized symptoms of hyperarousal (APA, 1994). PTSD patients show a diminished ability to extinguish this conditioned fear, which may be evidence for prefrontal cortex dysfunction and reduced amygdala inhibition (Milad et al., 2009; Kolassa et al., 2007). Shin et al. (Shin et al., 2005) have demonstrated that PTSD is marked by heightened amygdala activation and reduced anterior cingulate and prefrontal cortex activity when viewing fearful faces. Consistent with this finding, PTSD patients exhibited diminished activity in the mPFC to unattended fearful faces (Kim et al., 2008). Both studies reported that PTSD symptom severity was associated with decreased mPFC activity, demonstrating the neurobiological importance of this brain region in the pathophysiology of PTSD (Shin et al., 2005; Kim et al., 2008). Likewise, compared to healthy individuals, PTSD patients failed to recruit vmPFC activity when viewing pictures that were threatening, but unrelated to trauma (Phan et al., 2006). In patients with PTSD, the default mode network – brain regions that include the mPFC and the posterior cingulate cortex that are believed to be more "active" during rest (Gusnard and Rachle, 2001) – has been affected by the pathophysiology of the disorder as well. Specifically, resting state functional connectivity of the posterior cingulate cortex with the perigenual anterior cingulate and the right amygdala is associated with current PTSD symptoms, and that correlation with the right amygdala predicts future PTSD symptoms (Lanius et al., 2010). Furthermore, supporting these functional studies, there is DTI evidence that the white matter structural integrity of the cingulum bundle is compromised in PTSD patients compared to healthy individuals (Kim et al., 2005; Kim et al., 2006). Therefore, it is clear that not only the functionality of the amygdala and the mPFC are impaired in PTSD, but also their connectivity is disrupted as well.

Future research exploring the similarities and differences between non-anxious, normal anxious, and pathologically anxious individuals is needed. Based on numerous findings highlighting the relationship between the amygdala–mPFC circuitry and anxiety, developing treatments – whether they involve medication or psychotherapy – for anxiety disorders that target these brain regions will prove to be useful.

8. Conclusions

From normal emotion to pathological anxiety, an organism's reaction to biologically relevant stimuli and the regulation of these responses can be usefully conceived as a constant struggle between bottom-up and top-down brain processes. A wealth of animal and human neuroimaging studies has shown that the amygdala and the prefrontal cortex, particularly the medial regions, are central to these processes. Investigating the connectivity between the amygdala and the prefrontal cortex has provided a deeper understanding of the role of the amygdala–mPFC circuitry in anxiety. Efficient crosstalk between the amygdala and the prefrontal cortex – represented as stronger structural and functional connectivity – predicts beneficial behavioral outcomes in terms of emotion regulation and anxiety.

Acknowledgements

Supported by the National Institute of Mental Health grants to MJK (F31 MH090672) and PJW (R01 MH080716).

References

Adolphs, R., Tranel, D., Damasio, A.R., and Damasio, H. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Lett Nat *372*, 669–672.

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A.R. (1995). Fear and the human amygdala. Journal of Neuroscience *15*, 5879–5891.

Aggleton, J.P., ed. (1992). The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction (New York: Wiley-Liss, Inc).

Aggleton, J.P., Burton, M.J., and Passingham, R.E. (1980). Cortical and subcortical afferents to the amygdala of the rhesus monkey (Macaca mulatta). Brain Research *190*, 347–368.

Amaral, D.G., and Price, J.L. (1984). Amygdalo-cortical projections in the monkey (Macaca fascicularis). Journal of Comparative Neurology *230*, 465–496.

Amaral, D.G., Price, J.L., Pitkanen, A., and Carmichael, S.T. (1992). Anatomical organization of the primate amygdaloid complex. In: The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction, J.P. Aggleton, ed. (New York, NY: Wiley-Liss, Inc.).

APA (1994). Diagnostic and Statistical Manual of Mental Disorders. 4th Ed. (Washington DC: American Psychiatric Press).

Banks, S.J., Eddy, K.T., Angstadt, M., Nathan, P.J., and Phan, K.L. (2007). Amygdala-frontal connectivity during emotion regulation. Social Cognitive and Affective Neuroscience *2*, 303–312.

Barbas, H., and De Olmos, J. (1990). Projections from the amygdala to basoventral and mediodorsal prefrontal regions in the rhesus monkey. Journal of Comparative Neurology *300*, 549–571.

Basser, P.J., and Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. Journal of Magnetic Resonance B *111*, 209–219.

Bishop, S.J. (2007). Neurocognitive mechanisms of anxiety: an integrative account. Trends in Cognitive Sciences *11*, 307–316.

Bishop, S., Duncan, J., Brett, M., and Lawrence, A.D. (2004a). Prefrontal cortical function and anxiety: Controlling attention to threat-related stimuli. Nature Neuroscience *7*, 184–188.

Bishop, S.J., Duncan, J., and Lawrence, A.D. (2004b). State anxiety modulation of the amygdala response to unattended threat-related stimuli. Journal of Neuroscience *24*, 10364–10368.

Biswal, B., Yetkin, F.Z., Haughton, V.M., and Hyde, J.S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine *34*, 537–541.

Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., and Cohen, J.D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature *402*, 179–181.

Botvinick, M.M., Cohen, J.D., and Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: An update. Trends in Cognitive Sciences *8*, 539–546.

Bracht, T., Tuscher, O., Schnell, S., Kreher, B., Rusch, N., Glauche, V., et al. (2009). Extraction of prefronto-amygdalar pathways by combining probability maps. Psychiatry Research *174*, 217–222.

Breiter, H., Etcoff, N., Whalen, P.J., Kennedy, W., Rauch, S.L., Buckner, R., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. Neuron *17*, 875–887.

Carmichael, S.T., and Price, J.L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. Journal of Comparative Neurology *363*, s. 615–641.

Clark, D.M., and McManus, F. (2002). Information processing in social phobia. Biological Psychiatry *51*, 92–100.

Cooney, R.E., Atlas, L.Y., Joormann, J., Eugene, F., and Gotlib, I.H. (2006). Amygdala activation in the processing of neutral faces in social anxiety disorder: Is neutral really neutral? Psychiatry Research *148*, 55–59.

Cooney, R.E., Joormann, J., Atlas, L.Y., Eugene, F., and Gotlib, I.H. (2007). Remembering the good times: Neural correlates of affect regulation. Neuroreport *18*, 1771–1774.

Costa Jr, P.T., McCrae, R.R. (1997). Stability and change in personality assessment: The revised NEO Personality Inventory in the year 2000. J Pers Assess *68*, 86–94.

Cremers, H.R., Demenescu, L.R., Aleman, A., Renken, R., van Tol, M.J., van der Wee, N.J., et al. (2010). Neuroticism modulates amygdala-prefrontal connectivity in response to negative emotional facial expressions. Neuroimage *49*, 963–970.

Croxson, P.L., Johansen-Berg, H., Behrens, T.E., Robson, M.D., Pinsk, M.A., Gross, C.G., et al. (2005). Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. Journal of Neuroscience *25*, 8854–8866.

Davidson, R.J., Putnam, K.M., and Larson, C.L. (2000). Dysfunction in the neural circuitry of emotion regulation – A possible prelude to violence. Science *289*, 591–594.

Davis, M., and Shi, C. (2000). The amygdala. Current Biology *10*, R131.

Davis, M., and Whalen, P.J. (2001). The amygdala: Vigilance and emotion. Molecular Psychiatry *6*, 13–34.

Delgado, M.R., Nearing, K.I., LeDoux, J.E., and Phelps, E.A. (2008). Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. Neuron *59*, 829–838.

Di Martino, A., Scheres, A., Margulies, D.S., Kelly, A.M., Uddin, L.Q., Shehzad, Z., et al. (2008). Functional connectivity of human striatum: A resting state FMRI study. Cerebral Cortex *18*, 2735–2747.

Di Martino, A., Shehzad, Z., Kelly, C., Roy, A.K., Gee, D.G., Uddin, L.Q., et al. (2009). Relationship between cingulo-insular functional connectivity and autistic traits in neurotypical adults. American Journal of Psychiatry *166*, 891–899.

Dickie, E.W., and Armony, J.L. (2008). Amygdala responses to unattended fearful faces: Interaction between sex and trait anxiety. Psychiatry Research *162*, 51–57.

Drabant, E.M., McRae, K., Manuck, S.B., Hariri, A.R., and Gross, J.J. (2009). Individual differences in typical reappraisal use predict amygdala and prefrontal responses. Biological Psychiatry *65*, 367–373.

Ebeling, U., and von Cramon, D. (1992). Topography of the uncinate fascicle and adjacent temporal fiber tracts. Acta Neurochirurgica (Wien) *115*, 143–148.

Erk, S., Mikschl, A., Stier, S., Ciaramidaro, A., Gapp, V., Weber, B., et al. (2010). Acute and sustained effects of cognitive emotion regulation in major depression. Journal of Neuroscience *30*, 15726–15734.

Etkin, A., Klemenhagen, K.C., Dudman, J.T., Rogan, M.T., Hen, R., Kandel, E.R., et al. (2004). Individual differences in trait anxiety predict the response of the basolateral amygdala to unconsciously processed fearful faces. Neuron *44*, 1043–1055.

Eysenck, M. (1992). Anxiety: The Cognitive Perspective (Hillsdale, NJ: Erlbaum).

Fanselow, M.S., and LeDoux, J.E. (1999). Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. Neuron *23*, 229–232.

Fitzgerald, D.A., Angstadt, M., Jelsone, L.M., Nathan, P.J., and Phan, K.L. (2006). Beyond threat: amygdala reactivity across multiple expressions of facial affect. Neuroimage *30*, 1441–1448.

Foland-Ross, L.C., Altshuler, L.L., Bookheimer, S.Y., Lieberman, M.D., Townsend, J., Penfold, C., et al. (2010). Amygdala reactivity in healthy adults is correlated with prefrontal cortical thickness. Journal of Neuroscience *30*, 16673–8.

Freese, J.L., and Amaral, D.G. (2009). Neuroanatomy of the primate amygdala. In: The Human Amygdala, P.J. Whalen, E.A. Phelps, eds. (New York, NY: The Guilford Press).

Friston, K.J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. Human Brain Mapping *2*, 56–78.

Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., and Dolan, R.J. (1997). Psychophysiological and modulatory interactions in neuroimaging. Neuroimage *6*, 218–229.

Friston, K.J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. Neuroimage *19*, 1273–1302.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., et al. (2009). Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. Journal of Psychiatry & Neuroscience *34*, 418–432.

Ghashghaei, H.T., and Barbas, H. (2002). Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. Neuroscience *115*, 1261–1279.

Ghashghaei, H.T., Hilgetag, C.C., and Barbas, H. (2007). Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. Neuroimage *34*, 905–923.

Goldin, P.R., Manber, T., Hakimi, S., Canli, T., and Gross, J.J. (2009). Neural bases of social anxiety disorder: emotional reactivity and cognitive regulation during social and physical threat. Archives of General Psychiatry *66*, 170–180.

Gusnard, D.A., and Raichle, M.E. (2001). Searching for a baseline: functional imaging and the resting human brain. Nature Reviews Neuroscience *2*, 685–694.

Hahn, A., Stein, P., Windischberger, C., Weissenbacher, A., Spindelegger, C., Moser, E., et al. (2011). Reduced resting-state functional connectivity between amygdala and orbitofrontal cortex in social anxiety disorder. Neuroimage *56*, 881–889.

Hare, T.A., Tottenham, N., Galvan, A., Voss, H.U., Glover, G.H., and Casey, B.J. (2008). Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. Biological Psychiatry *63*, 927–934.

Hariri, A.R., Bookheimer, S.Y., and Mazziotta, J.C. (2000). Modulating emotional responses: Effects of a neocortical network on the limbic system. Neuroreport *11*, 43–48.

Hariri, A.R., Mattay, V.S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., et al. (2002). Serotonin transporter genetic variation and the response of the human amygdala. Science *297*, 400–403.

Hariri, A.R., Mattay, V.S., Tessitore, A., Fera, F., and Weinberger, D.R. (2003). Neocortical modulation of the amygdala response to fearful stimuli. Biological Psychiatry *53*, 494–501.

Hartley, C.A., and Phelps, E.A. (2010). Changing fear: the neurocircuitry of emotion regulation. Neuropsychopharmacology *35*, 136–146.

Hartley, C.A., Fischl, B., and Phelps, E.A. Brain structure correlates of individual differences in the acquisition and inhibition of conditioned fear. Cerebral Cortex. doi:10.1093/cercor/bhq253.

Heatherton, T.F., and Wagner, D.D. (2011). Cognitive neuroscience of self-regulation failure. Trends in Cognitive Sciences *15*, 132–139.

Holstege, G., Bandler, R., and Saper, C.B. (1996). The emotional motor system. Progress in Brain Research *107*, 3–6.

Johansen-Berg, H., Gutman, D.A., Behrens, T.E., Matthews, P.M., Rushworth, M.F., Katz, E., et al. (2008). Anatomical connectivity of the subgenual cingulate region targeted with deep brain stimulation for treatment-resistant depression. Cerebral Cortex *18*, 1374–1383.

Kapp, B.S., Supple, W.F., and Whalen, P.J. (1994). Stimulation of the amygdaloid central nucleus produces EEG arousal. Behavioral Neuroscience *108*, 81–93.

Kelly, A.M., Uddin, L.Q., Biswal, B.B., Castellanos, F.X., and Milham, M.P. (2008). Competition between functional brain networks mediates behavioral variability. Neuroimage *39*, 527–537.

Kessler, R.C., McGonagle, K.A., Zhao, S., Nelson, C.B., Hughes, M., Eshleman, S., et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Archives of General Psychiatry *51*, 8–19.

Kim, H., Somerville, L.H., Johnstone, T., Alexander, A.L., and Whalen, P.J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. Neuroreport *14*, 2317–2322.

Kim, H., Somerville, L.H., Johnstone, T., Polis, S., Alexander, A.L., Shin, L.M., et al. (2004). Contextual modulation of amygdala responsivity to surprised faces. Journal of Cognitive Neuroscience *16*, 1730–1745.

Kim, J.J., Fanselow, M.S., DeCola, J.P., and Landeira-Fernandez, J. (1992). Selective impairment of long-term but not short-term conditional fear by the N-methyl-d-aspartate antagonist APV. Behavioral Neuroscience *106*, 591–596.

Kim, M.J., and Whalen, P.J. (2009). The structural integrity of an amygdala-prefrontal pathway predicts trait anxiety. Journal of Neuroscience *29*, 11614–11618.

Kim, M.J., Chey, J., Chung, A., Bae, S., Khang, H., Ham, B., et al. (2008). Diminished rostral anterior cingulate activity in response to threat-related events in posttraumatic stress disorder. Journal of Psychiatric Research *42*, 268–277.

Kim, M.J., Gee, D.G., Loucks, R.A., Davis, F.C., Whalen, P.J. (2010a). Anxiety dissociates dorsal and ventral medial prefrontal cortex functional connectivity with the amygdala at rest. Cerebral Cortex. doi:10.1093/cercor/bhq237.

Kim, M.J., Loucks, R.A., Neta, M., Davis, F.C., Oler, J.A., Mazzulla, E.C., et al. (2010b). Behind the mask: the influence of mask-type on amygdala response to fearful faces. Social Cognitive and Affective Neuroscience *5*, 363–368.

Kim, M.J., Lyoo, I.K., Kim, S.J., Sim, M., Kim, N., Choi, N., et al. (2005). Disrupted white matter tract integrity of anterior cingulate in trauma survivors. Neuroreport *16*, 1049–1053.

Kim, S.J., Jeong, D.U., Sim, M.E., Bae, S.C., Chung, A., Kim, M.J., et al. (2006). Asymmetrically altered integrity of cingulum bundle in post-traumatic stress disorder. Neuropsychobiology *54*, 120–125. Emotion is a major research growth.

Kolassa, I.T., Wienbruch, C., Neuner, F., Schauer, M., Ruf, M., Odenwald, M., et al. (2007). Altered oscillatory brain dynamics after repeated traumatic stress. BMC Psychiatry *7*, 56.

Lanius, R.A., Bluhm, R.L., Coupland, N.J., Hegadoren, K.M., Rowe, B., Theberge, J., et al. (2010). Default mode network connectivity as a predictor of post-traumatic stress disorder symptom severity in acutely traumatized subjects. Acta Psychiatrica Scandinavica *121*, 33–40.

LeDoux, J.E. (1996). The Emotional Brain (New York: Simon and Shuster).

LeDoux, J.E. (2000). Emotion circuits in the brain. Annual Review of Neuroscience *23*, 155–184.

Leichnetz, G.R., and Astruc, J. (1976). The efferent projections of the medial prefrontal cortex in the squirrel monkey (Saimiri sciureus). Brain Research *109*, 455–472.

Lieberman, M.D., Eisenberger, N.I., Crockett, M.J., Tom, S.M., Pfeifer, J.H., and Way, B.M. (2007). Putting feelings into words: affect labeling disrupts amygdala activity in response to affective stimuli. Psychological Science *18*, 421–428.

Margulies, D.S., Kelly, A.M., Uddin, L.Q., Biswal, B.B., Castellanos, F.X., and Milham, M.P. (2007). Mapping the functional connectivity of anterior cingulate cortex. Neuroimage *37*, 579–588.

McIntosh, A.R., and Gonzalez-Lim, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. Human Brain Mapping *2*, 2–22.

Milad, M.R., Pitman, R.K., Ellis, C.B., Gold, A.L., Shin, L.M., Lasko, N.B., et al. (2009). Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. Biological Psychiatry *66*, 1075–1082.

Milad, M.R., Quinn, B.T., Pitman, R.K., Orr, S.P., Fischl, B., and Rauch, S.L. (2005). Thickness of ventromedial prefrontal cortex in humans is correlated with extinction memory. Proceedings of the National Academy of Sciences of the USA *102*, 10706–10711.

Milad, M.R. and Quirk, G.J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. Nature *420*, 70–74.

Milad, M.R., Wright, C.I., Orr, S.P., Pitman, R.K., Quirk, G.J., and Rauch, S.L. (2007). Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. Biological Psychiatry *62*, 446–454.

Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annual Review of Neuroscience *24*, 167–202.

Mori, S., Crain, B.J., Chacko, V.P., and van Zijl, P.C. (1999).Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Annals of Neurology *45*, 265–269.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D., Young, A.W., Calder, A.J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. Nature *383*, 812–815.

Neta, M., Norris, C.J., and Whalen, P.J. (2009). Corrugator muscle responses are associated with individual differences in positivity–negativity bias. Emotion *9*, 640–648.

Ochsner, K.N., Bunge, S.A., Gross, J.J., and Gabrieli, J.D. (2002). Rethinking feelings: an FMRI study of the cognitive regulation of emotion. Journal of Cognitive Neuroscience *14*, 1215–1229.

Ochsner, K.N., and Gross, J.J. (2005). The cognitive control of emotion. Trends in Cognitive Sciences *9*, 242–249.

Oler, J.A., Quirk, G.J., and Whalen, P.J. (2009). Cinguloamygdala interactions in surprise and extinction: interpreting associative ambiguity. In: Cingulate Neurobiology and Disease, B.A. Vogt, ed. (USA: Oxford University Press).

Pezawas, L., Meyer-Lindenberg, A., Drabant, E.M., Verchinski, B.A., Munoz, K.E., Kolachana, B.S., et al. (2005). 5-HTTLPR polymorphism impacts human cingulateamygdala interactions: a genetic susceptibility mechanism for depression. Nature Neuroscience *8*, 828–834.

Phan, K.L., Fitzgerald, D.A., Nathan, P.J., Moore, G.J., Uhde, T.W., and Tancer, M.E. (2005). Neural substrates for voluntary suppression of negative affect: A functional magnetic resonance imaging study. Biological Psychiatry *57*, 210–219.

Phan, K.L., Britton, J.C., Taylor, S.F., Fig, L.M., and Liberzon, I. (2006). Corticolimbic blood flow during nontraumatic emotional processing in posttraumatic stress disorder. Archives of General Psychiatry *63*, 184–192.

Phan, K.L., Orlichenko, A., Boyd, E., Angstadt, M., Coccaro, E.F., Liberzon, I., et al. (2009). Preliminary evidence of white matter abnormality in the uncinate fasciculus in generalized social anxiety disorder. Biological Psychiatry *66*, 691–694.

Phelps, E.A., Delgado, M.R., Nearing, K.I., and LeDoux, J.E. (2004). Extinction learning in humans: role of the amygdala and vmPFC. Neuron *43*, 897–905.

Quirk, G.J., and Beer, J.S. (2006). Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. Current Opinion in Neurobiology *16*, 723–727.

Quirk, G.J., Repa, C., and LeDoux, J.E. (1995). Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. Neuron *15*, 1029–1039.

Quirk, G.J. (2002). Memory for extinction of conditioned fear is long-lasting and persists following spontaneous recovery. Learning and Memory *9*, 402–407.

Rapee, R.M., and Heimberg, R.G. (1997). A cognitive-behavioral model of anxiety in social phobia. Behaviour Research and Therapy *35*, 741–756.

Rauch, S.L., Shin, L.M., and Phelps, E.A. (2006). Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research – past, present, and future. Biological Psychiatry *60*, 376–382.

Rescorla, R.A. (2001). Retraining of extinguished Pavlovian stimuli. Journal of Experimental Psychology: Animal Behavior Processes 27, 115–124.

Roelofs, K., Hagenaars, M.A., and Stins, J. (2010). Facing freeze: social threat induces bodily freeze in humans. Psychological Science *21*, 1575–1581.

Rogan, M.T., Staubli, U.V., and LeDoux, J.E. (1997). Fear conditioning induces associative longterm potentiation in the amygdala. Nature *390*, 604–607.

Rosen, J.B., and Schulkin, J. (1998). From normal fear to pathological anxiety. Psychological Review *105*, 325–350.

Roy, A.K., Shehzad, Z., Margulies, D.S., Kelly, A.M., Uddin, L.Q., Gotimer, K., et al. (2009). Functional connectivity of the human amygdala using resting state fMRI. Neuroimage *45*, 614–626.

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. Journal of Neuroscience *27*, 2349–2356.

Shah, S.G., Klumpp, H., Angstadt, M., Nathan, P.J., and Phan, K.L. (2009). Amygdala and insula response to emotional images in patients with generalized social anxiety disorder. Journal of Psychiatry & Neuroscience *34*, 296–302.

Shin, L.M., and Handwerger, K. (2009). Is posttraumatic stress disorder a stress-induced fear circuitry disorder? Journal of Traumatic Stress.

Shin, L.M., Wright, C.I., Cannistraro, P.A., Wedig, M.M., McMullin, K., Martis, B., et al. (2005). A functional magnetic resonance imaging study of amygdala and medial prefrontal cortex responses to overtly presented fearful faces in posttraumatic stress disorder. Archives of General Psychiatry *62*, 273–281.

Simmons, A., Matthews, S.C., Feinstein, J.S., Hitchcock, C., Paulus, M.P., and Stein, M.B. (2008). Anxiety vulnerability is associated with altered anterior cingulate response to an affective appraisal task. Neuroreport *19*, 1033–1037.

Simpson Jr, J.R., Drevets, W.C., Snyder, A.Z., Gusnard, D.A., and Raichle, M.E. (2001). Emotioninduced changes in human medial prefrontal cortex: II. During anticipatory anxiety. Proceedings of the National Academy of Sciences of the USA *98*, 688–693.

Somerville, L.H., Kim, H., Johnstone, T., Alexander, A.L., and Whalen, P.J. (2004). Human amygdala responses during presentation of happy and neutral faces: correlations with state anxiety. Biological Psychiatry *55*, 897–903.

Stefanacci, L., and Amaral, D.G. (2002). Some observations on cortical inputs to the macaque monkey amygdala: an anterograde tracing study. Journal of Comparative Neurology *451*.

Stein, M.B., Goldin, P.R., Sareen, J., Zorrilla, L.T., and Brown, G.G. (2002). Increased amygdala activation to angry and contemptuous faces in generalized social phobia. Archives of General Psychiatry *59*, 1027–34.

Stein, M.B., Simmons, A.N., Feinstein, J.S., and Paulus, M.P. (2007). Increased amygdala and insula activation during emotion processing in anxiety-prone subjects. American Journal of Psychiatry *164*, 318–327.

Stevens, W.D., Buckner, R.L., and Schacter, D.L. (2009). Correlated low-frequency BOLD fluctuations in the resting human brain are modulated by recent experience in category-preferential visual regions. Cerebral Cortex.

Straube, T., Schmidt, S., Weiss, T., Mentzel, H.J., and Miltner, W.H. (2009). Dynamic activation of the anterior cingulate cortex during anticipatory anxiety. Neuroimage *44*, 975–981.

Urry, H.L., van Reekum, C.M., Johnstone, T., Kalin, N.H., Thurow, M.E., Schaefer, H.S., et al. (2006). Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. Journal of Neuroscience *26*, 4415–4425.

Wager TD., Davidson ML., Hughes BL., Lindquist MA., and Ochsner K.N. (2008). Prefrontalsubcortical pathways mediating successful emotion regulation. Neuron *59*, 1037–1050.

Welborn, B.L., Papademetris, X., Reis, D.L., Rajeevan, N., Bloise S.M., and Gray, J.R. (2009). Variation in orbitofrontal cortex volume: relation to sex, emotion regulation and affect. Social Cognitive and Affective Neuroscience *4*, 328–339.

Whalen, P.J. (1998). Fear, vigilance, and ambiguity: initial neuroimaging studies of the human amygdala. Current Directions in Psychological Science *7*, 177–188.

Whalen, P.J., and Phelps, E.A. (2009). The Human Amygdala (New York, NY: The Guilford Press).

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.

Whalen, P.J., Shin, L.M., McInerney, S.C., Fischer, H., Wright, C.I., and Rauch, S.L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. Emotion *1*, 70–83.

Wig, G.S., Buckner, R.L., and Schacter, D.L. (2009). Repetition priming influences distinct brain systems: evidence from task-evoked data and resting-state correlations. Journal of Neurophysiology *101*, 2632–2648.

Yoon, K.L., Fitzgerald, D.A., Angstadt M., McCarron R.A., and Phan K.L. (2007). Amygdala reactivity to emotional faces at high and low intensity in generalized social phobia: a 4-Tesla functional MRI study. Psychiatry Research *154*, 93–98.

Jacek Dębiec
Molecular & Behavioral Neuroscience Institute
and Department of Psychiatry
University of Michigan, Ann Arbor, MI 48109, USA

The Matter of Emotions: Toward the Brain-Based Theory of Emotions

A formulation of the brain-based theory of emotions poses challenges to neural sciences and even more so to our traditional views on emotions and affective processes. Such methodological challenges and tensions between sciences and traditional views can best be understood from the historical perspective. However, while philosophical attempts to understand emotions have a long tradition and are deeply rooted in culture and society, the neuroscientific pursuit of emotions is a relatively recent endeavor. This essay will discuss selected historical and contemporary issues characterizing the neuroscientific search for emotions.

Early Accounts and Conceptualizations of Emotions

Emotions have been traditionally viewed as internal 'stirrings' or subjective mental 'movements' associated with certain bodily responses, behaviors and thoughts (Debiec and LeDoux, 2013). Primary intuitions and ideas about affective aspects of human life expressed in ancient Greek works of literary arts and conceptualized by early philosophers have had a lasting impact on Western thought. Fear, love, anger, jealousy and hate are universally present in the works of

Homer, Sophocles, Aeschylus and others. One of the most remarkable and striking descriptions of an emotional state can be found in Homer's *Iliad*. Andromache, Hector's wife, awaiting her husband's return from a duel with Achilles, hears from a distance her step-mother's lamenting cry:

> (…) She heard the cry coming as from the wall, and trembled in every limb;
> the shuttle fell from her hands.....she said…I heard the voice of my husband's honoured mother; my own heart beats as though it would come into my mouth and my limbs refuse to carry me; some great misfortune for Priam's children must be at hand. May I never live to hear it, but I greatly fear that Achilles has cut off the retreat of brave Hector and has chased him on to the plain where he was singlehanded; I fear he may have put an end to the reckless daring which possessed my husband… (XXII. 437; Homer, 1898)

In the above quoted paragraph, Homer describes most of what the Ancients observed about emotions: 1) Emotions occur in the context that co-defines them (Andromache's fright is preceded by her waiting for the return of Hector from a duel with the dangerous enemy); 2) Emotions often have identifiable triggers (hearing a sudden cry of Hecuba, Hector's mother); 3) Emotions are associated with bodily changes (such as Andromache's accelerated heartbeat, muscle contractions followed be a lowered muscle tone and loosening of the handgrip); 4) Emotions are associated with certain behaviors (Hecuba's cry or Andromache's dropping an object); 5) Emotions are associated with relevant thoughts (Andromache's thoughts of Hector's death) and 6) Bodily states and behaviors triggered by emotional cues occur before thoughts associated with emotions (Andromache's thoughts of Hector's death follow her bodily responses elicited by Hecuba's lamenting cry).

Emotions, often referred to as 'passions', 'desires', 'appetites' or 'feelings' were frequently addressed by Ancient Greek philosophers (Finger, 2001). Plato recognized a universal character of emotions and their fundamental role in constituting human nature. In *Gorgias*, Plato indicated that sharing emotions enabled people's understanding of each other's emotional states:

> …if mankind did not share one common emotion which was the same though varying in its different manifestations, but some of us experienced peculiar feelings unshared by the rest, it would not be easy for one of us to reveal his feelings to another (*Gorgias*; 481c, p. 265; transl. by W.D. Woodhead; in Plato 2002).

In another of his books, *Phaedo,* Plato emphasized that emotions were responsible for linking the soul (which represented an intellect or reason) to the body:

> …every pleasure or pain has a sort of rivet with which it fastens the soul to the body and pins it down and makes it corporeal, accepting as true whatever the body certifies (*Phaedo*; 83d, p. 66; transl. by H. Tredennick; Plato, 2002).

Plato, thus, recognized a unique function of emotions in connecting the body with the soul and in revealing bodily states to the soul. If we apply this model to the above quoted example from *Iliad*, a key to understanding Andromache's sudden changes in behavior and bodily states, as well as associated thoughts would be her fright and despair. However, despite his insights into emotions, Plato diminished their role. The concept of the soul was central to Plato's philosophical anthropology, whereas emotions were perceived as forces compromising the soul and reasoning. In *Laws*, Plato explained this limiting and dangerous potential of emotions:

> Passion is an ill-favored thing, and the speaker who does his wrath the favor to feast it on the poison it craves turns all the humanity education has fashioned within him into brutishness once more; persistence in his morose rancor makes him a wild beast, and that sorry return is all the return passion makes him for his favors (*Laws*; 935a, p. 1485; transl. by A.E. Taylor; in Plato, 2002).

Like Plato, Aristotle recognized that certain affective states were shared by humans and non-human animals. The fact of sharing emotions with animals, however, was not compromising to the soul; e.g. in the *Nicomachean Ethics*, while discussing parental affections for their offspring (referred to as friendship) Aristotle notes that these affections: *are not only in men, but also in birds, and in most animals* (*Nicomachean Ethics;* VIII.1; Aristotle, 1898). Aristotle distinguished between passions that were instinctive and emotions that required an involvement of the intellect (Finger, 2001). Aristotle emphasized that passions were driven by seeking pleasure or avoiding pain, whereas emotional actions, such as acts of altruism or courage, could reveal the true virtue of a person (Finger, 2001). Although, Aristotle's views on emotions were much more nuanced than those of Plato, the author of the *Nicomachean Ethics* continued Plato's legacy of perceiving affective states as something inferior to reason.

Early literary accounts show that emotions were viewed as complex multidimensional phenomena associated with distinct triggers and contexts, recognizable behaviors, different bodily states and thoughts. Despite these insights into the complexity of emotions, philosophers since Antiquity have often defined emotions through their opposition to the intellect and thus considered them irrational. Furthermore, by many, like Plato, emotions were not only contrasted with reason but were also suspected of distorting thoughts and the intellect. Plato overlooked multidimensional accounts of emotions available in his times, such as Homer's description in *The Iliad*. Assuming that the intellect was the superior human capacity, Plato did not recognize the uniqueness of emotions. For Plato, reason was a measure of emotions. Emotions were significant in so far as they were affecting the intellect. The tradition of looking at emotions from the perspective of the workings of the intellect has continued until this day. This simplified view of emotions slowly started changing after the publication of Charles Darwin's book *The Expression of the Emotions in Man and Animals* in 1872 and William James' article *What is an Emotion?* in 1884. Darwin's work provided a biological framework for studying emotions, whereas James' theory set up a direction for discussion and research on emotions in psychology.

Emotions as Biological Phenomena

Although the occurrence of emotions in animals was previously recognized by Ancient Greek thinkers, it was Charles Darwin who provided the theoretical framework for biological studies of emotions. Darwin presented his views on emotions in *The Expression of the Emotions in Man and Animals* in which he drew upon his own animal and human observations, anthropological work, clinical observations and experimental research on anatomical and physiological basis of facial expressions (Darwin, 1886; Finger, 2001). Darwin emphasized that affective displays and behaviors, such as expressions of joy, pain, fear or anger, have been formed by adaptive pressures as much as anatomical features. These behaviors served in the past (and in many instances continue to serve to some degree) the survival of an organism. For example, sneering or snarling, common expressions of anger in humans, may remind of our evolutionary past when displaying a readiness to bite could chase away a potential enemy:

> The expression here considered, whether that of a playful sneer or ferocious snarl, is one of the most curious which occurs in man. It reveals his animal descent; for no one, even if rolling on the ground in a deadly grapple with an enemy, and attempting to bite him, would try to use his canine teeth more than his other teeth (Darwin, 1886; p. 253).

While discussing displays of fear and horror Darwin wrote:

> Men, during numberless generations, have endeavored to escape from their enemies or danger by headlong flight, or by violently struggling with them; and such great exertions will have caused the heart to beat rapidly, the breathing to be hurried, the chest to heave, and the nostrils to be dilated. As these exertions have often been prolonged to the last extremity, the final result will have been utter prostration, pallor, perspiration, trembling of all the muscles, or their complete relaxation (Darwin, 1886; 308).

In the light of Darwin's theory, bodily states and behaviors experienced by Andromache, after she had learned about Hector's death, gain a new meaning. The evolutionary context is as important for the understanding of emotional states as the context of their immediate occurrence. A sudden cry of horror alerts the motor system (trembling of Andromache's limbs) and increases blood circulation (accelerated heart rate) preparing the body for typical defense responses, even if in Andromache's case neither fighting, fleeing or staying immobile (to avoid being spotted by a predator) were practical options. As emphasized by Ronald De Sousa in his philosophical analysis of the evolutionary approach to emotions, an adaptive character of emotional disposition is always historical and does not necessarily represent the best solution (De Sousa, 1990). Darwin's observations provided tools for the shift of focus from a discussion on emotions in general to the study of particular emotions. It paved the way for the biological and social studies of emotions.

Twelve years after the publication of Darwin's *The Expression of the Emotions*, William James wrote his remarkable paper *What is an Emotion?* (James, 1884). James noted the lack of brain research of emotions despite the dynamic progress in the discipline of brain anatomy and physiology, and proposed to study the neural mechanisms of emotions based upon the model emerging from studies of sensory and motor processes. James proposed the existence of neural mechanisms predisposed to responding to specific environmental triggers:

> Every living creature is in fact a sort of lock, whose wards and springs presuppose special forms of key, - which keys however are not born attached to the locks, but are sure to be found in the world near by as life goes on. And the locks are indifferent to any but their own keys (James, 1884, p. 191).

James suggested the existence of emotion circuits that included: 1) afferent pathways carrying sensory information about environmental triggers to the central nervous system, 2) efferent projections controlling motor and visceral responses associated with affective states and 3) afferent pathways carrying information about bodily states to the brain. Emotions, according to James, were arising from the commotion in inner organs and in behavior produced by the environmental triggers:

> the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur is the emotion (James, 1884, pp. 189–190).

James' model provides a possible explanation of the sequence of events associated with an emotional state, e.g. hearing Hecuba's cry triggers Andromache's behaviors and bodily responses which, in turn, are detected and identified by brain areas responsible for conscious thinking as fear. A similar view that feelings arise from the bodily changes was independently formulated in 1885 by a Danish anatomist, Carl G. Lange, who studied physiological responses associated with emotions. The James-Lange "peripheral" theory of emotions, although not much supported by experimental data, inspired theorists and experimentalists searching for neural circuits of emotions. Discussions and attempts to refine the James-Lange theory continue to this day (Damasio, 2000; Prinz, 2004).

Darwin's evolutionary approach to emotions and the James-Lange theory laid grounds for empirical studies of emotions. They showed that emotions are in themselves worth studying (as opposed to being perceived as merely distortions of thought processes) and started a process of restoring a view of emotions as multidimensional phenomena.

Early Studies of Emotion Circuits in the Brain

Without a sound empirical support, the James-Lange theory soon encountered criticism (LeDoux, 1996; Prinz, 2004; Finger, 2001). Clinical studies of patients with spinal cord injuries and experimental disruptions of major sensory pathways in animals showed that apparent disconnection between the periphery and the brain did not abolish emotional responses. Among the main critics of the James-Lange theory were Charles L. Dana and Walter B. Cannon.

Dana, who was a neurologist, observed that damage to the cortex alone may disrupt conscious feelings (Dana, 1921). He emphasized that the brain, especially the cortex, which was at that time known to be associated with conscious processes, may generate feelings without the sensory input from the body as posited by the "peripheral' theory of emotions. Another line of argument came from Cannon who was an experimentalist and who conducted animal research (Cannon, 1927). Cannon argued that the autonomic nervous system, which controls visceral reactions, would lack the specificity to distinguish between affective states and be too slow to generate emotions. In contrast to Dana, who associated emotions with the cortex, Cannon emphasized a key role of the hypothalamus, which receives afferents carrying information about environmental triggers, and sends projections to the body to regulate emotional responses, and to the cortex to generate conscious feelings. The works of Cannon and others (especially Philip Bard) suggested the role of the hypothalamus in behaviors and bodily responses associated with rage, fear and sexual excitement (Cannon, 1927; Bard, 1934a, 1934b).

The first brain model of the plausible circuitry of emotions was proposed by James W. Papez (Papez, 1937; Debiec and LeDoux, 2007). Papez never conducted research on emotions and his model was purely speculative. He synthesized experimental data from Cannon, Bard and others with clinical observations of patients with brain damage. Papez's model appealed to his contemporaries as he envisioned the neural circuitry involved in emotional processing, starting at the point of sensory input up to the level of subjective feeling. According to Papez, the plausible circuit of emotional processing included sensory areas of the thalamus and cortex, the hypothalamus, anterior thalamus, cingulate cortex, hippocampus, and mammilary bodies. Papez proposed that the stream of processing that starts with the receptor organs reaches the thalamus where it splits into the "stream of movement", the "stream of thought" and the "stream of feeling". The "stream of movement" projects to the subcortical areas involved in movement control, the "stream of thought" carries information to the lateral cerebral cortex where sensations are transformed into perceptions, thoughts and memories. The "stream of feeling" conveys sensory information to the hypothalamus which, as shown by the work of Bard and Canon, is responsible for emotional responses. Papez posited that mamillary bodies, a part of the hypothalamus, project to the cingulate cortex which interconnects with the lateral cortex. According to Papez, the cingulate cortex influences emotional processing in the hypothalamus through its outputs to the hippocampus, which, via the fornix sends messages to the posterior region of the hypothalamus. The cingulate cortex in Papez' model served as a receptive region for the experiencing of emotion. The interconnections between the cortical regions and the anterior cingulate would: 1) "add emotional coloring to psychic processes" in the cortex and 2) explain how emotion may arise from psychic (cortical) activity. Emotions in Papez' model may thus result either from hypothalamic or cortical activity. However, Papez attributed the key role in emotions to the hippocampus:

> The central emotive process of cortical origin may then be conceived as being built up in the hippocampal formation and as being transferred to the mamillary body and thence through the anterior thalamic nuclei to the cortex of the gyrus cinguli (Papez, 1937, p. 728).

His hypothesis on the role of the hippocampus was based on its location within the proposed circuitry, as well as clinical observations of patients with rabies who displayed profound emotional disturbances and whose hippocampi, as revealed by postmortem studies, were characterized by typical for rabies changes. The circuitry proposed by Papez is often referred to as limbic circuit, since the element of this circuitry, the cingulate cortex is also named the limbic cortex (this name was introduced by the French anatomist Pierre Paul Broca who observed that a significant part of the cingulate cortex is situated at the edge or *limbus* dividing the medial and the dorsal parts of each hemisphere).

Interestingly, Papez who himself was an anatomist, never specified what he meant by "emotion." He used emotion for modeling the organization of brain pathways. However, his localization of emotion processes in distinct brain areas inspired new research. Papez' model did not withstand the scrutiny of empirical verification. Interestingly, Papez seemed to believe that there is one central circuitry for processing various emotional responses, an assumption not supported by further research. However, his basic ideas regarding the interrelationships between brain sites, as well as suggestions that changes of activity in one area may affect the entire circuit, are still commonly applied (Neylan, 1995).

Current Brain Research of Emotions and its Philosophical Criticism

After a lengthy period of time during which the study of emotions had been marginal in neuroscience, partly due to the complexity of the subject and also due to the enduring perception of emotions as being inferior to, and therefore not such an interesting topic of studies as cognitive processes, emotions have become one of the mainstream themes of brain research. Significant progress has been achieved in unveiling molecular and circuit mechanisms associated with affective processes. Lessons from the history and philosophy of science teach us that a historical perspective is necessary to understand the value of this progress and its impact on science and society (Heller, 2011). In addition, the abundance of studies makes it simply impossible to review the current state of research in affective neuroscience. There have been, however, some successful attempts by affective neuroscientists to comprehensibly communicate to the lay audience perspectives emerging from the current research on emotion (LeDoux, 1996, Panskepp, 1998; Damasio, 1996, 2000, 2003). These popularized accounts of emotions presented by neuroscientists have been shaped by the authors' own research and expertise and, as often admitted by the authors themselves, are not meant to be comprehensive descriptions of all emotion processes or general theories of emotions. For example, Joseph E. LeDoux who experimentally studies circuits controlling threat responses in rats, proposed a model explaining how harmful stimuli are processed in the brain to produce defense reactions (LeDoux, 1996). In his research LeDoux uses a threat (or fear) conditioning, an experimental paradigm in which a neutral stimulus is paired with a noxious event. As a result of threat conditioning procedure, a previously neutral stimulus acquires the ability to trigger inherent behavioral, autonomic, and endocrine responses that are expressed automatically in the presence of danger. Threat conditioning occurs in a variety of species ranging from fruits flies to humans. LeDoux and others have shown that the key structure for threat conditioning is the amygdala. The activation of the amygdala neurons by threat signaling stimuli excites the pathways controlling threat responses. If the flow of information through the amygdala is disrupted, threat signaling stimuli do not trigger defense responses anymore. LeDoux has shown in his "two-road model" that sensory information about threat is transmitted to the amygdala through two independent pathways: the "low road", or thalamic pathway which provides the amygdala with a rapid but imprecise representation of the sensory input and the "high road", or cortical path, which conveys a more complex representation based on cortical computations. Since conscious processes are associated with cortical areas, LeDoux's model explains how environmental stimuli trigger automatic threat responses before we are aware of the actual danger:

> The direct pathway allows us to begin to respond to potentially dangerous stimuli before we fully know what the stimulus is (LeDoux, 1996, p. 164).

Although, LeDoux's model is mostly based on animal research, recent human brain imaging studies show that the amygdala may be excited by the stimuli that are not consciously perceived and that the subcortical pathway appears to be mainly involved in rapid amygdala responses (Vuilleumier et al., 2001; Ohman, 2002; Phelps, 2006; Luo et al., 2010; Cecere et al., 2013). The LeDoux model, thus, explains at the neural circuitry level the mechanisms triggering Andromache's bodily reactions and behavior (in response to Hecuba's dreadful cry) before she became aware of her fear (Homer, 1898; XXII.437).

However, like the ancient thinkers, who ignored rich literary descriptions of emotional states and presented simplified views of emotions, some contemporary philosophers also appear to show a lack of understanding of the empirical studies of emotions and their significance. For example, Peter Hacker (a philosopher) together with his co-author Maxwell Bennett (a neuroscientist) present critical views on neuroscientific contributions to the understanding of emotions (Bennett and Hacker, 2003). Hacker and Bennett reject animal research as a source of insight into human emotions. Discussing LeDoux's model, Hacker and Bennett state:

> Conditioning rats to fear an electric shock (cf. LeDoux's important research) is a poor foundation for insight into human emotion and its neuropsychological conditions and accompaniments (Bennett and Hacker, 2003, p. 205).

Hacker and Bennett base their criticism on their assumption that emotions are by definition conscious states:

> Emotions, we have noted, are not brain states or bodily responses. They are neither objects of perception nor sensations that are felt. To feel frightened just is to be frightened, just as feeling angry is not distinct from being angry (Bennett and Hacker, 2003, p. 210).

One can imagine Hacker's and Bennett's resistance to accepting animal research as a source of insight into human emotion consciousness processes, however, the claim that animal threat conditioning research is useless for understanding neuropsychological conditions of human fear is not only unjustified but also contrary to empirical evidence. Animal research for over two decades has informed studies of human fear. Ample data demonstrate that the same anatomical structures, such as the amygdala, are associated with threat responses in humans and non-human animals (Whalen and Phelps, 2009). Like their Ancient predecessors, Hacker and Bennett appear to ignore available sources of insight into emotion processes and perceive emotions through the prism of the intellect and conscious thinking (*to feel frightened just is to be frightened*). As a consequence of such simplified understanding of emotions, infants or patients with alexithymia, a condition characterized by difficulties with identifying one's emotions (Moriguchi and Komaki, 2013; Gu et al., 2013), would be considered as being void of emotions. Despite his distrust in emotions, Plato was able to acknowledge that humans can recognize each other emotions because they share them (*Gorgias*; 481c, p. 265; transl. by W.D. Woodhead; in Plato, 2002). We have to share something with a crying infant or an anxious alexithymic patient that allows us to name their emotional states even if they cannot do it themselves. In their reductionist (reducing emotions to what is studied through introspection) approach to emotions the authors of the *Philosophical Foundations of Neuroscience* appear to miss the social context of emotions. Hacker and Bennett also reject the LeDoux "two-road model":

> It makes no sense, save as misleading figure of speech, to say, as LeDoux does, that it is "possible for your brain to know that something is good or bad before it knows exactly what it is (Bennett and Hacker, 2003, p. 152).

Hacker's and Bennett's criticism of the neuroscientific accounts of emotions encounter its own philosophical criticism. Daniel Dennett replies:

> But who is misled? Not LeDoux and not LeDoux's readers, if they read carefully, for they can see that he has actually found a very good way to make the surprising point that a specialist circuit in the brain can discriminate something as dangerous, say, or as desirable, on the basis of a swift sort of 'triage' that is accomplished before the information is passed on those networks that complete the identification of the stimulus (Dennett, 2007, p. 94–95).

Hacker and Bennett assume a rigid notion of "knowing", which is limited to conscious processes. Therefore, they are unable to accept the fact that the brain can detect, compute and respond to stimuli without the involvement of conscious processes. Contrary to expectations of some of the philosophers, the role of philosophy in conceiving models or concepts, which could be used in the neuroscientific study of emotions, is very limited. Dennett continues:

> Do the authors offer anything else that might be of value to the neurosciences? They offer no positive theories or models or suggestions

> about how such theories or models might be constructed, of course, since that would be not the province of philosophy (Dennett, 2007, p. 95).

By making the introspective intellect a measure for all other mental processes, philosophy early-on rejected emotions and their multiple dimensions. As a consequence, philosophy has not established any adequate conceptual framework to understand emotions in their complexity. Indeed, although the neuroscientific view on emotions is incomplete and as long as neuroscience continues it will remain incomplete, it is emerging as a major source of knowledge about emotions. In as much as 'understanding' of Andromache's horror and despair at the very basic human level requires empathy, grasping the complexity and multidimensionality of emotions necessitates opening to perspectives brought by empirical research.

Acknowledgments

The author acknowledges the support of the University of Michigan and the NARSAD Young Investigator Award from the Brain & Behavior Research Foundation.

References

Aristotle (1889). The Nicomachean Ethics, R.W. Browne, transl. (London: George Bell and Sons).

Bard, P. (1934a). On emotional expression after decortication with some remarks on certain theoretical views: Part I. Psychological Review *41* (4), 309–329.

Bard, P. (1934b). On emotional expression after decortication with some remarks on certain theoretical views: Part II. Psychological Review *41* (5), 424–449.

Bennett, M.R., and Hacker, P.M.S. (2003). Philosophical Foundations of Neuroscience (Malden, MA: Blackwell Publishing).

Cannon, W.B. (1927). The James-Lange theory of emotions: a critical examination and an alternative theory. American Journal of Psychology *39*, 106–124.

Cecere, R., Bertini, C., and Ladavas, E. (2013). Differential Contribution of Cortical and Subcortical Visual Pathways to the Implicit Processing of Emotional Faces: A tDCS Study. The Journal of Neuroscience *33* (15), 6469–6475.

Damasio, A. (1996). Descartes' Error: Emotion, Reason and the Human Brain (London: Papermac).

Damasio, A. (2000). The Feeling of What Happens. Body and Emotion in the Making of Consciousness (San Diego: A Harvest Book Harcourt, Inc.).

Damasio, A. (2003). Looking for Spinoza. Joy, Sorrow and the Feeling Brain, (Orlando: Harcourt, Inc.).

Dana, C.L. (1921). The anatomic seat of the emotions: a discussion of the James-Lange theory. Archives of Neurology and Psychiatry *6*, 634–639.

Darwin, C. (1886). The Expression of the Emotions in Man and Animals (New York: D. Appleton and Company).

Debiec, J., and LeDoux, J.E. (2013). Emotion. In: Encyclopedia of Sciences and Religions, A.L.C. Runnehov and L. Oviedo, eds. (Dodrecht: Springer), pp. 721–723.

Debiec, J., and LeDoux, J.E. (2007). James W. Papez. In: New Dictionary of Scientific Biography, Vol. 6, N. Koerte, ed. (Detroit: Thompson – Gale), pp. 10–14.

Dennet, D. (2007). Philosophy as naïve anthropology. Comment on Bennett and Hacker. In: Neuroscience and Philosophy. Brain, Mind, and Language, M. Bennett, D. Dennett, P. Hacker and J. Searle. (New York: Columbia University Press).

De Sousa, R. (1990). The Rationality of Emotion (Cambrdige, MA: The MIT Press).

Finger, S. (2001). Origins of Neuroscience. A History of Explorations into Brain Function (Oxford: Oxford University Press).

Gu, X., Hof, P.R., Friston, K.J., and Fan, J. (2013). Anterior insular cortex and emotional awareness. The Journal of Comparative Neurology *521*, 3371–3388.

Heller, M. (2011). Philosophy in Science. An Historical Introduction (Heidelberg: Springer).

Homer (1898). The Iliad, S. Butler, transl. (London – New York: Longmans, Green, & Co.).

James, W. (1884). What is an Emotion? Mind *9*, 188–205.

LeDoux, J.E. (1996). The Emotional Brain (New York: Simon and Schuster).

LeDoux, J.E. (2014). Coming to terms with fear. Proceedings of the National Academy of Sciences of the United States of America *111* (8), 2871–2878.

Luo, Q., Holroyd, T., Majestic, C., Cheng, X., Schechter, J., and Blair, R.J. (2010). Emotional Automaticity Is a Matter of Timing. The Journal of Neuroscience *30* (17), 5825–5829.

Moriguchi, Y., and Komaki, G. (2013). Neuroimaging studies of alexithymia: physical, affective, and social perspectives. Biopsychosocial Medicine *7*(1):8. doi: 10.1186/1751-0759-7-8.

Neylan, T.C. (1995). Classical articles in neuropsychiatry: introduction to the series. Journal of Neuropsychiatry and Clinical Neurosciences *7*, 102–103.

Ohman, A. (2002). Automaticity and the amygdala: Nonconscious responses to emotional faces. Current Directions in Psychological Science *11* (2), 62–66.

Panskepp, J. (1998). Affective Neuroscience. The Foundations of Human and Animal Emotions (New York / Oxford: Oxford University Press).

Papez, J.W. (1937). A proposed mechanism of emotion. Archives of Neurology and Psychiatry *38*, 725–743.

Phelps, E.A. (2006). Emotion and cognition: insights from studies of the human amygdala. Annual Review of Psychology *57*, 27–53.

Plato (2002). The Collected Dialogues, E. Hamilton and H. Cairns, eds. (Princeton, New Jersey: Princeton University Press).

Prinz, J. (2004). Embodied emotions. In: Thinking About Feeling, R.C. Solomon, ed. (Oxford: Oxford University Press).

Vuilleumier, P, Armony, J.L., Driver, J., and Dolan, R.J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron *30* (3), 829–841.

Whalen, P.J., and Phelps, E.A. (edts.) (2009). The Human Amygdala (New York: The Guilford Press).

Bram T. Heerebout
R. Hans Phaf
Psychonomics Department, University of Amsterdam

Good Vibrations Switch Attention

An Affective Function for Network Oscillations in Evolutionary Simulations

Abstract

A new hypothesis on the neural mechanisms linking affect to attention is brought forward by evolutionary simulations on agents navigating a virtual environment while collecting food and avoiding predation. The connection strengths between nodes in the networks controlling the agents were subjected to random variation and the fittest agents were selected for reproduction. Unexpectedly, oscillations of node activations emerged, which drastically enhanced the agent's fitness. Here we analyzed the mechanisms involved in the modulation of attention and found that oscillations acted on competitive networks. Response selection depended on the connection structure, but the speed and efficacy of switching between selections was modulated by oscillation frequency. The main focus of this study was the differential emergence of stimulus-specific oscillation frequencies. Oscillations had a higher frequency in an appetitive motivational state than in an aversive state. We suggest that oscillations in biological networks also mediate the affective modulation of attention.

1. Good vibrations switch attention

The method of computationally simulating evolutionary processes provides a unique opportunity for the automated development of models and hypotheses on cognitive and affective processes and their underlying neural mechanisms. The role of the modeler is limited to setting up the evolutionary selection procedure (e.g., a genetic algorithm; Holland, 1975), the initial state, and the environmental conditions. This has several advantages of which the opportunity for innovation is not the least. Novel models with mechanisms and functions that had not been previously considered may emerge from these undirected optimization procedures.

In previous simulation work (Heerebout and Phaf, in press) we serendipitously discovered 'artificial neural oscillations' in agents that inhabited a virtual environment (see Figure 1). The oscillations were an emergent property, not intentionally built in, and not even thought of beforehand, which proved to be highly adaptive. These simulations extended our investigation of LeDoux's (1996) evolutionary justification of his dual-pathway model for the processing of emotional stimuli (den Dulk, Heerebout and Phaf, 2003). The agents, which were controlled by artificial neural networks, increased their chances of reproduction and survival by collecting food while avoiding predators. Analyses of the oscillating agents showed that the evolutionary advantage was conferred by an enhanced capacity for attentional switching when in an oscillatory mode.

To test the switching efficacy of an oscillating agent, it was compared to a non-oscillating agent from a control simulation with a simpler network. Both agents responded to a plant, placed in front of them at an angle of 45 degrees to its left, which would, as the agent approached the plant, suddenly be replaced by a predator. Their phenotypic behavior was described in terms of speed of movement and rotational speed, and revealed distinct approach and avoidance behaviors. The oscillating agent would first 'cautiously' approach the plant following a swerving, 'zigzag', trajectory and then, as it detected the

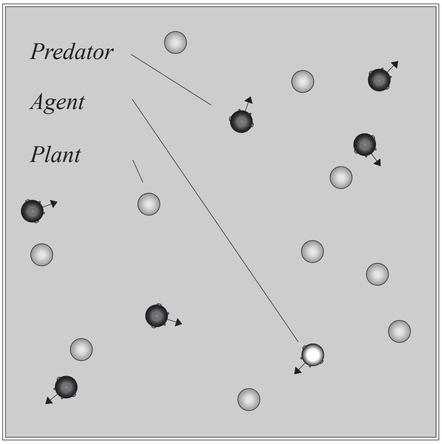


Figure 1. Graphic depiction of the virtual environment with one agent, six predators and ten plants. The arrows pointing away from the agent and the predators indicate the direction of movement. The environment is 'torus' shaped (i.e., if a predator or agent leaves the environment on one side of the square, it immediately re-appears on the opposite side).

predator, make a sharp turn (at 0.042 degrees per time step) and accelerate strongly(the speed increased 87% in the first ten time steps). The non-oscillatory agent kept a more constant speed. When it would detect a predator, it accelerated only slightly (its speed increased only 13% in the first ten time steps) while turning away (with 0.034 degrees per time step). In addition, this appeared to be a robust finding, because fast switching oscillating agents turned up in five out of the seven replications.

Although the behavioral consequences of the oscillations were evident, the underlying mechanism demanded further analysis. Moreover, the networks from the replicated simulations exhibited oscillations at different frequencies. If the oscillations were adaptive because they enhanced the ability to quickly switch between competing behaviors, then the evolutionary demands imposed by the two types of stimuli might induce different oscillation frequencies. After all, increased flexibility may be advantageous when searching for food, but when fleeing from a predator foraging constitutes a distraction with potentially fatal consequences.

The present study aimed to address these two issues. First, we will briefly reiterate the method of evolutionary simulations. This is followed by the discussion of a simplified connection scheme, newly abstracted from previously evolved agents with oscillations in their networks. This simple connection scheme operated according to a winner-take-all competitive process, of which the switching speed was modulated by oscillation frequency. Similar competitive processes are often assumed to be involved with attentional selection in biological neural networks. More specifically, in our evolved networks the competitive mechanism seems to be responsible for selection-for-action (cf. Allport, 1989). We further investigated how the oscillations are able to modulate selection speed without influencing the specific outcome of this selection process. In addition, the simulation was replicated another 25 times to test whether positive stimuli indeed yielded faster oscillations than negative stimuli. Finally, the results from the evolutionary simulations are combined with empirical data on attentional selection and affect into the hypothesis that also in biological neural networks positive affect is associated with higher neural oscillation frequencies and more efficient attentional switching than negative affect, which conversely fosters the maintenance of attentional focus.

2. Evolutionary simulations

Computer simulations allow researchers to perform experiments that would otherwise be impossible because of temporary, spatial, monetary or ethical constraints. Our evolutionary history is not directly observable and, therefore, empirically investigating evolutionary explanations is a difficult task. To address this problem with computational means, we applied a Genetic Algorithm (Holland, 1975) to agents navigating a virtual environment while collecting food and avoiding predators. Genetic Algorithms (GA's) are traditionally used as an optimization method, inspired by natural selection. Starting from an initial population that consists of a random combination of gene values, the search for optimal solutions proceeds through selection and reproduction. Fitness is defined as some measure of how well the parameter set solves the problem. Thus, better solutions get a higher chance of survival and reproduction. When a specific solution gets to be reproduced, its parameters are subject to mutations and crossovers. Mutation may improve the performance of the population by occasionally suggesting a new partial solution. With ‘crossover’, two individual solutions selected for reproduction are recombined. The new solutions replace the less fit solutions in a population. From generation to generation, this leads to a higher overall fitness.

The evolutionary inspired optimization technique has been fruitfully applied to real-world problems. For example, the Aircraft Aerodynamics and Design Group at Stanford employed GA’s for optimization of aircraft design (Gage, 1995). A novel wing design was sought that minimized drag. The initial population resembled simple solutions, encoding the wing as a collection of tapered and twisted elements. Next, the solutions were slightly mutated, adding or removing a few random elements. The drag for each solution was then calculated using a vortex lattice analysis. This ‘fitness function’ determined which designs were selected for reproduction (i.e., which were copied with slight alterations). Multiple repetitions of the process quickly led to the discovery of winglets, small vertical additions at the wingtips (a standard CO_2 emission reducing measure on passenger jets nowadays). Later on, the algorithm found even more efficient “C”-shaped wingtips. This design concept was patented and its application to future aircraft models is under investigation. Although the fitness function guided the evolutionary process, the random variations allowed for the emergence of a fundamental, innovative design concept.

Due to their innovative power, GAs stand in stark contrast with more conventional cognitive modeling approaches. Connectionist models, but also models constructed with a symbolic formalism like ACT-R or SOAR, are usually analytical models constructed to explain empirical data, formalizing some cognitive theory. The evolutionary simulations, consisting of a GA applied to the agents' neural networks, on the other hand, automatically synthesize models that produce the goal-behavior, as specified in the fitness function. It should be emphasized that the modeler does not construct a model, but merely sets up conditions conducive to the evolution of the model that performs a particular function. When considering different models that explain behavioral data equally well, the model that has been shown to be more likely to develop in evolutionary simulations should be preferred above the others. Moreover, due to the innovative power of the GA even new models and hypotheses, not previously considered by neuroscientists, may emerge.

The conventional neural-network modeling of affective processes (Armony et al., 1995; Den Dulk et al., 1997; Phaf et al., 2001) has been strongly inspired by the neurobiological work of LeDoux (1986, 1996). According to LeDoux, the mammalian brain processes fear stimuli via two parallel pathways, running from the sensory thalamus to the amygdala. One pathway projects directly from the sensory thalamus to the lateral nucleus of the amygdala, the other indirectly, first passing through the cortex before reaching the amygdala. LeDoux suggested that these pathways might have evolved to functionally supplement each other. The direct pathway can produce a fast fear response, consisting of autonomic, endocrine and motor (e.g., avoidance) reactions, to a potentially threatening stimulus. The slower indirect pathway processes the stimulus more extensively and may strengthen the initial response if appropriate in case of an actual threat, or inhibit the direct reaction in case of a false alarm. LeDoux (1996) argued that the different functions of the two pathways are adaptive, because the evolutionary cost of a miss exceeds the total costs of the many false alarms produced by the direct pathway.

The straightforward formulation in terms of evolutionary benefits and costs by LeDoux (1996) facilitated the translation by den Dulk et al. (2003) into evolutionary simulations, which indeed resulted in the emergence of dual-processing networks. The setup largely followed the simulations of Beer (1990), in which Beer used Braitenberg

(1984) vehicles (i.e., the agents) to collect plants through chemotaxis. For the present study, the same software for the simulations was used as in the study den Dulk et al. (2003) but the networks of the agents were extended with an extra recurrent layer (see Figure 2) to investigate the adaptive value of such recurrent connections.

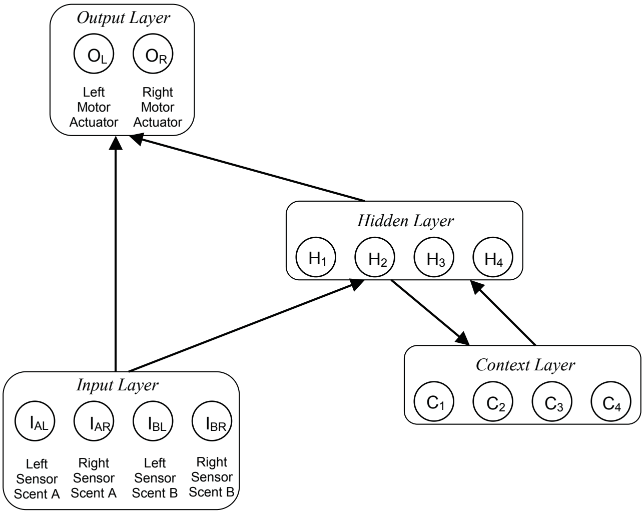
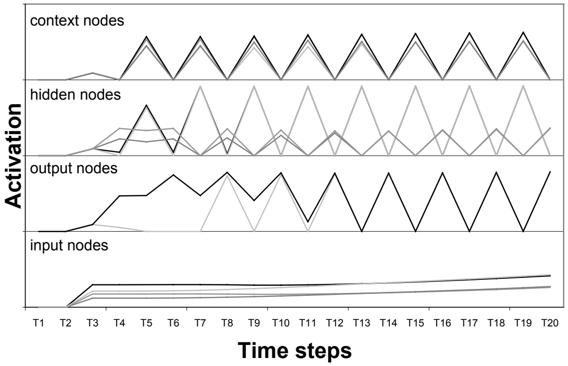


Figure 2. The four layers of the network architecture used in the simulations. The layers were potentially fully connected (i.e., all connections in the indicated direction could evolve to non-zero values in the course of the simulations). The arrows depict the direction of the connections. The context layer was added by Heerebout and Phaf (in press) to allow for recurrent processing.

The present study completely followed the original design choices made by den Dulk et al. (2003). The oscillations, thus, resulted from a 'design' process by the GA, but not from the part of the modelers. Relative to this previous study no parameter adjustments were made and the space of possible fitness functions was not explored in any manner to arrive at these results. Insofar as the range of parameter values was explored by den Dulk et al., it mostly proved of little influence when the parameters remained within reasonable bounds. For instance, whether the number of plants that were simultaneously present in the environment was four, six, or eight, the agents would always evolve to approach plants. The choice for the number of six plants, which was also used in our simulations, was therefore arbitrary. After initialization the simulations followed a simple three-stage procedure, the gist of which will be described here in brief, but more details can be found in Appendices A, B and C.

(A) Food stimulus



(B) Predator stimulus

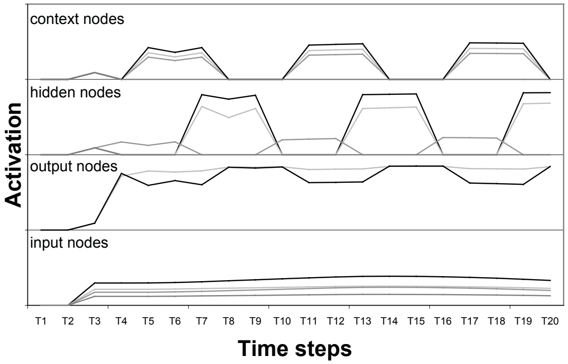


Figure 3. An example of oscillating activations in the nodes of the network of a last-generation network. The nodes' activations are shown over time, separated per layer. The activations shown in panel (A) were recorded when the agent was confronted with a plant and in panel (B) when the agent was confronted with a predator. In this example, the period of the oscillation was two time steps with the plant stimulus and six time steps with the predator stimulus.

During initialization a population of agents was created with all the agents' network connection weights set to zero. The potential network connections, which could evolve to nonzero values, specified a simple multi-layered architecture with an input layer of four nodes projecting to two output nodes and to four hidden nodes. The hidden nodes projected to the output nodes and also to another layer of four hidden nodes, which recurrently projected back to the former hidden nodes. The connections between the two layers of hidden nodes could develop independently (i.e., they did not have symmetrical weight values). One time step was defined as the duration of the calculation of a node's activation. Therefore, the signals a node received at time t determined its activation at time t, but this activation would be propagated over its sending connection at t+1 (see also Appendix D).

In the first stage the algorithm would select agents to create offspring until the population size had tripled. With the initial population the selection is arbitrary, but with subsequent generations the agent's fitness (measured in the second stage) determined the agent's probability of being selected (see Appendix A for details on the selection and reproduction process). In the second stage, the fitness of each agent would be determined individually in twelve different random virtual environments, which always contained ten plants and six predators. The movement of the predators was governed by networks similar to those of the agents. These networks, however, were pre-configured to let the predators approach the agents. This configuration would not change (i.e., there was no 'arms race' between prey and predator).

The agents had olfactory sensors, located on the front left and right of their round bodies, which picked up the smells emitted by the plants and predators. The agents could utilize the lateral difference in smell intensity to estimate the direction of the source. The sensors relayed the detected intensities to the input nodes of the agent's network. After processing, the output nodes' activation would drive the agents' left and right motor actuators, giving the agents tank-like propulsion. Appendix B describes the details of the smell distribution.

If an agent collided with a plant, the food was considered eaten, and one unit would be added to the agent's energy level. The plant would subsequently reappear at a randomly selected location. Movement by the agent led to a reduction of energy (the exact energy cost was calculated from the force an agent exerted on the environment; also see Appendix B for details). An agent would be removed from the environment if it collided with a predator, if its energy was depleted or if the maximum of 10,000 time steps had passed.

When the agent was removed, its fitness was calculated as its energy level multiplied by the total number of time steps spent in all environments divided by the number of times it had been tested. In the third stage agents are removed from the population, whereby the lowest scoring agents have the highest probability of being removed. Appendix C explains how the agents' fitness is used to calculate this probability.

After cycling through the stages described above 10,000 times, den Dulk et al. (2003) obtained weight configurations that had a functionality similar to LeDoux's dual pathway model. This architecture was found only in simulations, when (a) the plants and predators were hard to distinguish, (b) processing via a hidden layer took more time than direct processing, and (c) both plant and predator were relevant to the agent's fitness. The qualitatively different types of processing in the two pathways were further supported by lesion studies of the separate pathways (i.e., setting all the weights in one path to zero) in the artificial neural networks. These findings support LeDoux's hypothesis that the dual pathway model is adaptive and that it can evolve over time through small mutations, when there is an exchange between time pressures and need for accuracy.

3. Oscillations in competitive networks

Heerebout and Phaf (in press) subsequently extended the simulation setup to investigate the adaptive value of recurrent connections via a new hidden layer in the indirect pathway. The recurrent processing could hypothetically lead to a working memory capacity for previously processed stimuli by allowing activations from previous time steps to contribute to the present network state. Unexpectedly, the added layer caused the networks to oscillate (e.g., see Figure 3). These oscillations clearly had an adaptive function, more than doubling the agents' performance relative to agents from the control simulation, without recurrent connections.

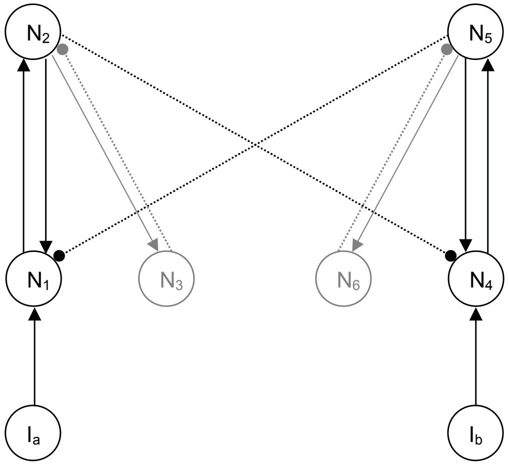


Figure 4. The network configuration distilled from the evolved agents that generates the oscillations. De 'I' nodes provide inputs of scent 'a' and scent 'b'. The solid lines with the arrowheads depict excitatory connections (with weight = 5) and the dashed lines with the globule heads depict inhibiting connections (with weight = -10). Without the grey nodes N_3 and N_6, the nodes N_1 and N_2 simply compete with nodes N_4 and N_5, resulting in the attentional selection of one of these pairs. With the indicated connections to the nodes N_3 and N_6 added, the network oscillates which modulates the competitive process.

In comparison to the LeDoux's dual-pathway model, an alternative type of dual-processing dynamics emerged in these simulations. The direct route was no longer biased towards avoidance, but had a general energizing function. After lesioning the direct route, approach and avoidance tendencies could be distinguished in the indirect route, but the actions were not actually performed. For the action tendency to actually develop into an action, 'arousal' from the direct route was needed. The nodes in the indirect route entered into a competitive process whenever input was delivered to them. The selection of specific approach or avoidance action tendencies was determined by the evolved weight configuration. The oscillations added speed to this selection to a similar extent as was added by preparation through the direct route in the classical dual-pathway model.

In classical competitive network models (e.g., see McClelland and Rumelhart, 1981; Rumelhart and Zipser, 1985) input causes the nodes in a network to compete for activation and a steady state can mostly be reached through a process of multiple constraint satisfaction, in which the connection weights constitute the constraints. Although networks of this kind have been applied to model visual attention (e.g., Phaf et al., 1990), the inability of competitive networks to swiftly leave a winning state remains a major problem. Switching in these networks usually requires a period of activation decay or even manual intervention by the modeler. The oscillations might suggest a solution to this switching problem by providing a supplementary mechanism to reset the competition. With oscillating competitive networks the winning state could easily shift to another set of nodes in the troughs of the periodical activations. To examine how the oscillations influenced response switching, we distilled an abstract connection structure (see Figure 4) from the oscillating agents. Because the competitive behavior was preserved in the abstract network, even when the oscillation-generating connections were removed, the impact of the oscillations on the competitive process could be investigated.

The abstract network was constructed by identifying the connection sets that produced the oscillations in the networks. To this end, negligible and non-effective weights (without which the oscillatory properties of the remaining network did not change) were set to zero and nodes that consequently did not receive any activation were omitted from the abstract network. With these steps the idealized symmetrical connection scheme of Figure 4 (for reasons of computational parsimony all networks were symmetrical with respect to the left and right nodes) was formed. Activations were mirrored with respect to input from node I_a and I_b. Activation of N_1, by input node I_a, for instance, triggered activation of N_2. Excitatory feedback from N_2 to N_1 helped to sustain the activation in both nodes. The excitatory connection to the inhibitory node N_3, however, implemented a "flip-flop" mechanism and caused these activations to oscillate. Interestingly, this type of recurrent inhibition has also been identified as a neural oscillation generator by Ritz and Sejnowski (1997; see also Dupret et al., 2008). In addition, the strong inhibitory crossed connections from N_2 to N_4 and from N_5 to N_1 caused N_1 and N_4 to compete for activation. For the inhibitory mechanism to successfully overrule the excitatory input, all inhibitory connections were, in this idealized model, exactly twice the strength of the excitatory connections (-10 and 5 respectively). When context nodes N_3 and N_6 are removed from the connection scheme, a classical competitive network remains. The N_1 and N_4 nodes compete for activation when input activation is applied, and only the most strongly activated node (i.e., the 'winner') can preserve a non-zero activation.

The switching efficacy was examined by comparing performances of oscillatory and non-oscillatory competitive networks in a switching task (see Figure 5). The non-oscillatory network had the same connection weights as the oscillatory network, but the connections involved in the generation of oscillations were lesioned (i.e., set to zero). So, both types of networks were derived from the evolved agents that showed oscillatory activations. Non-oscillatory agents evolved in the simulations may have developed other switching mechanisms, presumably involving less mutual inhibition in the first place, but from their fitness scores we know that these mechanisms were less effective than oscillations (see Heerebout and Phaf, in press). For the switching task, input consisted of a monotonically increasing signal to the input node I_a, which would be replaced instantaneously by a strong signal to the I_b input node (see upper panel A of Figure 5). Panels B and C of Figure 5 show the ensuing activations of the nodes N_1, N_2, N_3, and N_4 of the non-oscillatory and oscillatory competitive networks, respectively. It is clear that the non-oscillatory network was unable

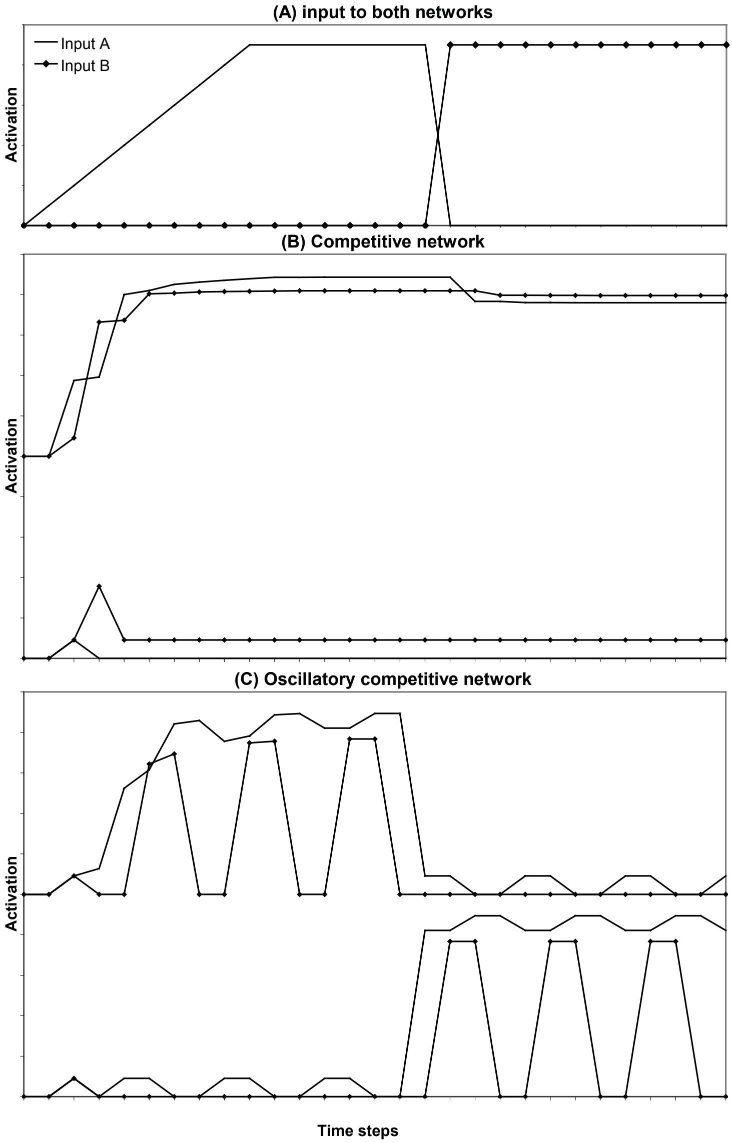


Figure 5. The input activations to both networks (A), activations in the non-oscillatory network (B), and in the oscillatory network (C). The scale of the x-axis (time steps) and the y-axis (activation) are the same for the three panels.

to overcome the winner-take-all mechanism after the switch and remained stuck. Because the oscillatory network never completely settled, it did not take a large change in input to tip over to another winner. Due to the inhibitory pulses from node $N_{3,}$ which caused the intermittent deactivation of node N_2 and provided an opportunity for N_4 to become activated, the oscillatory network smoothly switched in a few time steps. The abstract model demonstrates that because its winning 'steady' state consists of oscillations, it is able to switch to another 'steady' state in the periodically occurring troughs of near-zero activation.

The different abilities of oscillatory and non-oscillatory competitive networks to switch winners clearly support our hypothesis that oscillations increase behavioral flexibility. Inhibition may have been exaggerated in the non-oscillatory networks relative to non-oscillatory networks evolved in actual simulations, but still in the latter networks switching ability is inferior to that of evolved oscillatory networks (Heerebout and Phaf, in press). In addition, the abstract network distinguishes the specific underlying mechanisms for selection and switching. In both oscillatory and non-oscillatory agents action tendencies (i.e., approach or avoidance) compete with each other, which according to the 'selection-for-action' framework (e.g., Allport, 1989) can be characterized as an attentional process (cf. Desimone and Duncan, 1995; Duncan, 1996; Phaf et al., 1990). In the abstract network the modulation of competition by oscillations is superimposed on the process of attentional selection. The oscillations, thus, enable the switching between winning representations without affecting the nature of the response that is being selected.

In a similar evolutionary-computation study by Ward and Ward (2008) agents evolved to actively direct attention to successive falling targets in order to 'catch' them. They found that to be able to focus on the first falling target, the activations from the second falling target, which was simultaneously present, needed to be inhibited. This inhibition subsequently caused difficulty in redirecting attention to the second target after the first one was caught. As a result the agent showed a moment of hesitation after the first target was caught. If the targets were closer together, the inhibition required to stay focused on the first target was more powerful, and the cost in terms of hesitation was larger. Switching cost thus increased with level of competition. These findings are remarkably consistent with our own (Heerebout and Phaf, in press) in showing that evolution in most real-world situations not only needs to solve the problem of attentional selection but also of how attention should be subsequently redirected. The artificial neural networks in Ward and Ward's setup had bilaterally symmetric weights, which prevented the emergence of the flip-flop mechanism (i.e., requiring asymmetrical excitatory and inhibitory connections) responsible for oscillations. To be sure the actual simulations should of course be done, but we would guess that letting loose the constraint of bidirectional symmetry in their simulations would lead to a strengthening of mutual inhibition and competition and to a reduction in switching costs due to the emergence of oscillations.

4. Oscillations in biological neural networks

The oscillations emerging in the evolutionary simulations can probably be associated with, relatively high-frequency, gamma oscillations (30–80 Hz) in biological neural networks (for an overview, see Buzsáki and Draghun, 2004). From the different frequency bands gamma oscillations have most consistenly been associated with attention (Womelsdorf and Fries, 2007; Bauer et al., 2009). The relevance of these oscillations, if any (Pareti and DePalma, 2004), remains a matter of debate. Because evolutionary simulations optimize the genotypic neural networks given particular environmental conditions, they may be in a unique position to shed light on the adaptive function of the emerging oscillations from a computational angle. Our previous work, at least, strongly argues in favor of an adaptive (i.e., fitness-enhancing) function of oscillations (Heerebout and Phaf, in press).

Neurobiological research has suggested a large variety of roles for the oscillations in different areas of the mammalian brain. Possible functions include the binding of cell assemblies (Gray, König, Engel and Singer, 1989), biasing input selection through neuronal resonance (Hutcheon and Yarom, 2000), providing a sense of time (Buhusi and Meck, 2005), selective amplification (Lengyel et al., 2005), and sequence learning (Ulanovsky and Moss, 2007). The enhanced ability of the agents to reorganize behavior in response to specific stimuli in our simulations may correspond to the suggestion of Schaefer, Angelo, Spors, and Margrie (2006) that oscillations serve to discriminate between stimuli (see also Brody and Hopfield, 2003, who showed that simple oscillating network models implemented sensory segmentation). In ground-breaking research, Sohal et al. (2009) even demonstrated in genetically modified mice that light-induced gamma oscillations enhanced information transmission in neocortex by reducing circuit noise and amplifying signals through the circuit. Increased flexibility and switching ability would indeed imply that the information carried by the neural output is more strongly related to the input signal. The mice with induced gamma oscillations, thus, seemed to be able to switch attention more easily than the mice without gamma oscillations.

According to Fries, Nikolić, and Singer (2007) stimuli can be distinguished more easily because the interaction between excitatory pyramidal neurons and inhibitory interneurons results in a time-critical competition during the gamma cycle. The few pyramidal cells that are able to spike first, are the only ones to spike at all. Subsequently, the activity of these pyramidal cells is suppressed by strong inhibitory interneurons which are themselves activated by the initial activation of these cells. The mechanism closely resembles the flip-flop mechanism we encountered in the agents' networks from our simulations. They suggest this mechanism enables fast processing and flexible routing of activity, supporting fast selection of responses. The correspondence between the neurobiological findings of Fries et al. (2007) and our results further strengthens the association of the oscillations emerging in our simulations with gamma band oscillations in biological neural networks.

In contrast to our hypothesis, Börgers, Epstein, and Kopell (2008) argued on the basis of their (non-evolutionary) simulations, that oscillations increase stimulus competition, help to suppress distracters, and more narrowly focus selective attention. Their oscillations resulted from the interactions between fast-spiking interneurons, which were periodically subjected to a 'bath of inhibition' due to cholinergic activation by a second class of inhibitory interneurons. During this 'bath of inhibition' all neural activation was suppressed. Particularly activation from potential distracters could not reach threshold values. Subsequently, only the strongest representation in the network could peak before inhibition started all over again. In their view, only the selection resulting from competitive processes is strengthened by the oscillations. The increased focusing of attention due to oscillations in the Börgers et al. (2008) model would decrease, rather than increase, behavioral flexibility and reduce the amount of information transferred by a network (e.g., see Sohal et al., 2009). In contrast to Börgers et al., in our simulations a distinction between attentional selection and attentional switching emerges, with oscillations only affecting the latter. The further finding of affect-specific oscillation frequencies, which will be discussed hereafter, provides additional support for the distinction and this specific role of oscillations. The evolutionary motivation, which can be distilled quite easily from the simulations, and the experimental evidence supporting the hypothesized link between affect and attentional switching, makes it likely that a similar setup can also be found in biological neural networks.

5. New simulations reveal stimulus-specific frequency

According to LeDoux (1996) the faster, more coarsely grained, direct pathway is biased to evoke an aversive response, because the costs of hesitation are higher when confronted with a threat than with food. This evolutionary reasoning suggests that threatening stimuli should quickly capture and hold attention (cf. Öhman et al.,, 2001). We wondered whether a similar bias towards focusing attention on negative stimuli could be found in the oscillating networks, even when they did not have the classical LeDoux type of dual-route architecture. While moving away from an approaching predator, there is little need to heed to other stimuli. A high degree of distractibility caused by high-frequency oscillations seems a disadvantage in this situation.

A plant stimulus, on the other hand, should enable rapid behavioral switching, if a predator would appear suddenly. This increased behavioral flexibility, moreover, is supported by ample empirical research in which positive affect was found to increase cognitive and behavioral flexibility (Baumann and Kuhl, 2005; Das and Fennis, 2008; Dreisbach and Goschke, 2004; Fenske and Eastwood, 2003; Isen, 1999; Luu et al., 1998; Tan et al., 2009). The broaden-and-build theory of Fredrickson (1998), for instance, generalizes these ideas even further by assuming that positive emotions serve to broaden an individual's momentary thought-action repertoire by expanding the attentional focus. If indeed attentional flexibility is modulated by oscillations, it is only a logical step to expect higher oscillation frequencies for foraging agents than for fleeing agents in the simulations.

To investigate stimulus-specific oscillation frequencies, we replicated the simulations with the recurrent connections of Heerebout and Phaf (in press) an additional 25 times. This resulted in a total of 32 simulations. The new simulations were exact replications, but new pseudo-random numbers were used in the stochastic decisions (e.g., the weight mutations) for each simulation. We found 430 agents with oscillating networks in the last generation of 27 out of 32 simulations (totaling 547 agents). The oscillation frequencies, both with food and predator stimuli, ranged from zero to 0.5 time step^{-1} (frequency was measured as the inverse of the number of time steps in a complete activation cycle; periods ranged from infinite, 6, 5, 4, 3 to 2 time steps). Some (*n* = 347) networks oscillated both with plant and predator, sometimes with different frequencies, whereas 54 networks showed only plant oscillations and 29 networks showed only predator oscillations.

The presence of oscillations had a clear effect on the agents' fitness. The oscillating agents (*n* = 430; *average fitness* = 20,112; *SD*

= 17,833 energy x time step) reached significantly higher levels of fitness ($F(1, 546) = 7.67$; $p < 0.01$, $\eta_p^2 = 0.014$) than non-oscillating agents ($n = 110$; *average fitness* = 15,220; *SD* = 12,493 energy x time step) demonstrating that the oscillations have clear functional advantages in these simulations. Over the total of 547 agents the average frequency was significantly higher ($F(1, 547) = 25.27$, $p < 0.0001$, $\eta_p^2 = 0.044$) when an agent detected a plant ($f = 0.22$, $SD = 0.19$ time step^{-1}) than when an agent detected a predator ($f = 0.18$; $SD = 0.17$ time step^{-1}). The extrapolation of these results to biological neural networks, of course, remains hazardous, but if we assume that a time step (i.e., the time needed for activation transfer by a neuron) is about 5 ms, this would result on average in 44 Hz oscillations with the food and 36 Hz oscillations with the predator (c.f. the two frequency bands in Dupret et al., 2008). The frequency difference between stimulus types supports the conclusion that oscillations increase the agents' ability to switch behaviors, and that the difference arises from the differential environmental demands posed by food and predators.

An anonymous reviewer to a previous version of the manuscript suggested that the dynamic properties of the plants and predators could be confounded with their valence. The plants were always stationary, but the predators actively pursued the agents. To test whether this dynamic property could cause the difference in oscillation frequency we performed a further set of 22 simulations, in which the plants also moved through the environment, albeit in random directions. The speed (0.3 length units per time step) approximated the average speed of the predators. For the remainder, all conditions and parameters remained the same. If the speed difference is responsible for the differential processing of plants and predators, no frequency difference should emerge in these simulations.

The total number of agents in the last generations of all simulations equaled 377. Oscillations were again recorded in response to plants and to predators. Of the 377 agents 113 showed oscillations only in response to plants and 17 only to predators. The majority, 209 agents, oscillated both with plants and with predators. A small group of 38 agents did not show oscillations with either stimulus. Similar to the previous simulations, the oscillating agents ($n = 339$; *average fitness* = 16,977; *SD* = 14,752 energy x time step) reached significantly higher levels of fitness ($F(1, 375) = 10.9$; $p < 0.01$, $\eta_p^2 = 0.026$) than the non-oscillating agents ($n = 38$; *average fitness* = 9,282; *SD* = 6,738 energy x time step). The difference in oscillation frequency was significant ($F(1,376) = 67.98$, $p < 0.0001$, $\eta_p^2 = 0.153$) and even larger than with the previous simulations. When an agent detected a plant the average frequency was 0.25 (SD = 0.18) cycle per time step and when an agent detected a predator the average frequency was 0.16 (SD = 0.17) cycle per time step. If we again assume that a time step lasts 5 ms, the plant frequency corresponds to 50 Hz and the predator frequency to 32 Hz. These results do not seem to support alternative interpretations in terms of stimulus dynamics.

6. Frequency reflects affective and motivational state

Specific stimuli have acquired their positive affective value, because they generally raised fitness levels throughout evolutionary history. Inversely, negative stimuli are characterized by their recurrent potential to lower fitness. Positive and negative affect indeed appear to be generated by the nervous system as a neural code “to those aspects of the environment that were a consistent benefit or threat to gene survival in ancestral environments” (p. 173, Johnston, 2003). Affective states in the networks are thus inferred from the consequences specific stimuli may have for fitness level. During evolution the network ‘learns’ to predict these consequences and to prepare for appropriate actions. Appetitive-approach tendencies towards positive stimuli and aversive-avoidance tendencies away from negative stimuli have likely developed in evolution to maximize fitness benefits and to minimize fitness costs, respectively. Such action tendencies, at least, emerged from the undifferentiated initial networks in our simulations as a response to fitness-relevant stimuli. Fitness costs are minimized by a focusing of attention on the negative stimulus and low distractibility. Fitness benefits are maximized by an approach to all positive stimuli and a high level of distractibility by other stimuli. The level of distractibility in our simulations was positively related to oscillation frequency. On average, oscillations had a higher frequency when the agent was foraging than when it was trying to escape from a predator. We suggest that a specific oscillation frequency sets the network in a suitable attentional mode to deal with affectively valenced stimuli.

The aforementioned, highly comparable simulation study by Ward and Ward (2008) also addressed attentional switching, but did not investigate affective influences on switching. Interestingly, Ward and Ward linked the reactive inhibition of the second target to the attentional blink (AB) phenomenon, and referred to recent neurocognitive studies (Hommel et al., 2006; Kessler et al., 2006) that also relate attentional dwell time to reactive inhibition. The inhibition of the second target when it is sufficiently close to the first target in the Ward and Ward simulations can of course easily be transposed to interference in the attentional blink task. The comparison can even be taken one step further, because also affective influences have been found on the attentional blink (Olivers and Nieuwenhuis; 2006; see also Most et al., 2006; Todorović, 2009, recently obtained similar affective modulation of the attentional blink in our laboratory). Not only for the attentional blink task, but also for other experimental tasks there is substantial evidence for the hypothesized link between affect and attentional switching (Baumann and Kuhl, 2005; Das and Fennis, 2008; Dreisbach and Goschke, 2004; Fenske and Eastwood, 2003; Isen, 1999; Luu et al., 1998, Tan et al., 2009). In comparison, the empirical evidence for a relationship between affect and oscillation frequency is much scarcer.

The extrapolation of the simulation results to biological neural networks remains a big step, and surely needs to be supported by empirical evidence. There is some research to suggest that the frequency differences emerging from the evolutionary simulations may have a similar biological function. A motivational role of gamma oscillations is, for instance, suggested by a study of Rougeul-Buser and Buser (1997). They observed 40 Hz oscillations in a cat's motor, parietal, and visual cortices when it was waiting in front of a hole in the wall from which at times a mouse could pop out and then quickly disappear. When the cat was simply watching the mouse in a perspex box, however, only lower frequencies of 10–15 Hz showed up. We would argue that the cat by the hole is in a positive affective state, full of expectations of a nice meal, and is prepared to quickly switch from immobility to vigorous attack. With the box, however, where the cat cannot reach the mouse, both the need for this preparation and the corresponding oscillation frequency is lower. In the mouse, on the other hand, we would expect the oscillation frequency to be much lower when it is being hunted by the cat than when it, for instance, detects a piece of cheese.

A causal relationship between gamma oscillations and positive affect has to our knowledge only been demonstrated very recently by Tsai et al. (2009). They showed with their highly innovative optogenetic method (also see Sohal et al., 2009) that trains of high-frequency light flashes (50Hz) applied to the ventral tegmental area (i.e., containing many dopaminergic neurons) of optogenetically manipulated mice served as a strong reward signal in place conditioning. In the trial, mice were exposed to both high-frequency light trains in one room and low-frequency light trains in another. The total amount of light the mice were exposed to was equal in both rooms. Afterwards, when the mice were given the choice, they all preferred the room were they had been exposed to the high-frequency light trains. If the high, but not the low, frequency light trains can substitute a reward signal, normally produced by a positive stimulus, then this shows that, at least in some cases, high frequency oscillations correspond to positive affect.

The other optogenetic study we mentioned (Sohal et al., 2009), which was performed by the same group and published simultaneously with the study of Cardin et al. (2009), showed the induced gamma oscillations enhanced information transmission. Our evolutionary simulations suggest that the two observations are connected and that high neural oscillation frequencies facilitate efficient attentional switching and high levels of information transmission, whereas low frequencies foster the maintenance of attentional focus to negative stimuli. Although at the outset of their research these authors have not formulated specific hypotheses about the relation between oscillation frequency, attention, and motivational state, their results provide converging evidence for our hypothesis.

Two different hypotheses were brought forward by the evolutionary simulations in this study. First, oscillations have a modulatory role which is superimposed on attentional selection by means of competition. Second, oscillations have a higher frequency in an appetitive state, associated with positive affect, than in an aversive state associated with negative affect. Particularly the latter hypothesis appears to be new and may provide the missing neurobiological link between affect and attention, which has often been investigated in psychological research. There are only few ideas about the neurobiological mechanisms underlying affective influences on information processing and in view of these simulation results the oscillations seem a likely candidate. Computational simulations can only show that a particular hypothesis is more probable than another (i.e., the likelihood of its emergence is higher), but not that it necessarily holds true in biological neural networks. Evolutionary computation takes a special position within the modeling approaches, because it can create new models and thus form the starting point of the empirical cycle. Of course, novel hypotheses can only be successful if the good vibrations generated by the computer simulations resonate in empirical work and eventually lead to a switching in neurobiological research perspective.

References

Allport, A. (1989). Visual attention. In: Foundations of Cognitive Science, M.I. Posner, ed. (Cambridge, Mass: MIT Press), pp. 631–682.

Armony, J.L., Servan-Schreiber, D., Cohen, J.D., and LeDoux, J.E. (1995). An anatomically constrained neural network model of fear conditioning. Behavioral Neuroscience *109*, 246–257. doi:10.1037/0735-7044.109.2.246

Bauer, F., Cheadle, S.W., Parton, A., Müller, H.J., and Usher, M. (2009). Gamma flicker triggers attentional selection without awareness. Proceedings of the National Academy of Sciences *106*, 1666–1671. doi:10.1073/pnas.0810496106

Baumann, N., and Kuhl, J. (2005). The positive affect and flexibility: Overcoming the precedence of global over local processing of visual information. Motivation and Emotion *29*, 123–134. doi:10.1007/s11031-005-7957-1

Beer, R.D. (1990). Intelligence as adaptive behavior: An experiment in computational neuroethology. Artificial Intelligence *72*, 173–215. doi:10.1016/0004-3702(94)00005-L

Börgers, C., Epstein, S., and Kopell, N.J. (2008). Gamma oscillations mediate stimulus competition and attentional selection in a cortical network model. Proceedings of the National Academy of Sciences *105*, 18023–18028. doi:10.1073/pnas.0809511105

Brody, C.D., and Hopfield, J.J. (2003). Simple Networks for Spike-Timing-Based Computation, with Application to Olfactory Processing. Neuron *37*, 843–852. doi:10.1016/S0896-6273(03)00120-X

Braitenberg, V. (1984). Vehicles: Experiments in Synthetic Psychology (Cambridge, MA: MIT Press).

Buhusi, C.V., and Meck, W.H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. Nature Reviews Neuroscience *6*, 755–765. doi:10.1038/nrn1764

Buzsáki, G., and Draghun, A. (2004). Neuronal oscillations in cortical networks. Science *304*, 1926–1929. doi:10.1126/science.1099745

Cardin, J.A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L., and Moore, C.I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. Nature *459*, 663. doi:10.1038/nature08002

Das, E., and Fennis, B.M. (2008). In the mood to face the facts: When a positive mood promotes systematic processing of self-threatening information. Motivation and Emotion *32*, 221–230. doi:10.1007/s11031-008-9093-1

den Dulk, P., Heerebout, B.T., and Phaf, R.H. (2003). A computational study into the evolution of dual-route dynamics for affective processing. Journal of Cognitive Neuroscience *15*, 194–208. doi:10.1162/089892903321208132

den Dulk, P., Rokers, B., and Phaf, R.H. (1997). Connectionist simulations with a dual route model of fear conditioning. In: Perspective on Cognitive Science, Vol. 4, B. Kokinov, ed. (Sofia: New Bulgarian University Press), pp. 102–112.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual Review of Neuroscience *18*, 193–222. doi:10.1146/annurev.ne.18.030195.001205

Dreisbach, G., and Goschke, T. (2004). How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. Journal of Experimental Psychology *30*, 343–353.

Duncan, J. (1996). Cooperating brain systems in selective perception and action. In: Attention and Performance, Vol. XVI. Information integration in perception and communication, T. Inui and J.L. McClelland, eds. (Cambridge, MA: MIT Press), pp. 549–578.

Dupret, D., Pleydell-Bouverie, B., and Csicsvari, J. (2008). Inhibitory interneurons and network oscillations. Proceedings of the National Academy of Sciences *105*, 18079–18080. doi:10.1073/pnas.0810064105

Fenske, M.J., and Eastwood, J.D. (2003). Modulation of focused attention by faces expressing emotion: Evidence from flanker tasks. Emotion *3 (4)*, 327–343. doi:10.1037/1528-3542.3.4.327

Fredrickson, B.L. (1998). What Good are positive emotions? Review of General Psychology *2*, 300–319. doi:10.1037/1089-2680.2.3.300

Fries, P., Nikolić, D., and Singer, W. (2007). The gamma cycle. Trends in Neuroscience *30*, 309–316. doi:10.1016/j.tins.2007.05.005

Gage, P.J. (1995). New Approaches to Optimization in Aerospace Conceptual Design, Ph.D. (thesis, Stanford University).

Gray, C.M., König, P., Engel, A.K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature *338*, 334–337. doi:10.1038/338334a0

Heerebout, B.T., and Phaf, R.H. (in press). Emergent Oscillations in evolutionary simulations: Oscillating networks increase switching efficacy. Journal of Cognitive Neuroscience *22*. doi:10.1162/jocn.2009.21205

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

Hommel, B., Kessler, K., Schmitz, F., Gross, J., Akyürek, E., Shapiro, K., and Schnitzler, A. (2006). How the brain blinks: Towards a neurocognitive model of the attentional blink. Psychological Research *70*, 1–11. doi:10.1007/s00426-005-0009-3

Hutcheon, B., and Yarom, Y. (2000). Resonance, oscillation, and the intrinsic frequency preferences of neurons. Trends in Neurosciences *23*, 216–222. doi:10.1016/S0166-2236(00)01547-2

Isen, A.M. (1999). Positive affect. In: Handbook of Cognition and Emotion, T. Dalgleish and M.J. Power, eds. (New York: Wiley and Sons), pp. 521–539.

Johnston, V.S. (2003). The origin and function of pleasure. Cognition and Emotion *17*, 167–179. doi:10.1080/02699930302290

Kessler, K., Gross, J., Schmitz, F., and Schnitzler, A. (2006). Cortical Dynamics and synchronization related to multiple target consolidation under rapid-serial-visual-presentation conditions. Journal of Physiology *99*, 21–28.

LeDoux, J.E. (1996). The Emotional Brain (New York: Simon and Schuster).

Lengyel, M., Huhn, Z., and Érdi, P. (2005). Computational theories on the function of theta oscillations. Biological Cybernetics *92*, 393–408. doi:10.1007/s00422-005-0567-x

Luu, P., Tucker, D.M., and Derryberry, D. (1998). Anxiety and the motivational basis of working memory. Cognitive Therapy and Research *22*, 577–594. doi:10.1023/A:1018742120255

McClelland, J.L., and Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. Psychological Review *88*, 375–407. doi:10.1037/0033-295X.88.5.375

Most, S.B., Chun, M.M., Johnson, M.R. and Kiehl, K.A. (2006). Attentional modulation of the amygdala varies with personality. Neuroimage *31*, 934–944. doi:10.1016/j.neuroimage.2005.12.031

Olivers, C.N.L., and Nieuwenhuis, S. (2006). The beneficial effects of additional task load, positive affect, and instruction on the attentional blink.

Journal of Experimental Psychology: Human Perception and Performance *32*, 364–379. doi:10.1037/0096-1523.32.2.364

Öhman, A., Flykt, A., and Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. Journal of Experimental Psychology: General *130*, 466–478. doi:10.1037/0096-3445.130.3.466

Pareti G., and De Palma, A. (2004). Does the brain oscillate? The dispute on neuronal synchronization. Neurological Sciences *25*, 41–47. doi:10.1007/s10072-004-0235-2

Phaf, R.H., Dulk, P. den, Tijsseling, A., and Lebert, E. (2001). Novelty-dependent learning and topological mapping. Connection Science *13*, 293–321. doi:10.1080/09540090110085666

Phaf, R.H., Van der Heijden, A.H.C., and Hudson, P.T.W. (1990). SLAM: A connectionist model for attention in visual selection tasks. Cognitive Psychology *22*, 273–341. doi:10.1016/0010-0285(90)90006-P

Ritz, R., and Sejnowski, T.J. (1997). Synchronous oscillatory activity in sensory systems; new vistas on mechanisms. Current Opinion in Neurobiology *7*, 536–546. doi:10.1016/S0959-4388(97)80034-7

Rougeul-Buser, A., and Buser, P. (1997). Rhythms in the alpha band in cats and their behavioural correlates. International Journal of Psychophysiology *26*, 191–203. doi:10.1016/S0167-8760(97)00764-2

Rumelhart, D.E., and Zipser, D. (1985). Feature discovery by competitive learning. Cognitive Science *9*, 75–112

Schaefer, A.T., Angelo, K., Spors, H., and Margrie, T.W. (2006). Neuronal oscillations enhance stimulus discrimination by ensuring action potential precision. Public Library of Science Biology *4*, 1010–1024.

Sohal, V.S., Zhang, F., Yizhar, O., and Deisseroth, K. (2009). Parvalbumin neurons and gamma rhythms synergistically enhance cortical circuit performance. Nature *459*, 698–702. doi:10.1038/nature07991

Tan, H.K., Jones, G.V., and Watson, D.G. (2009). Encouraging the perceptual underdog: Positive affective priming of nonprefererred local-global processes. Emotion *9*, 238–247. doi:10.1037/a0014713

Todorović, A. (2009). Affective modulation of spatial and temporal attention in early vision (Unpublished Master's thesis. Psychology Department, University of Amsterdam).

Tsai, H., Zhang, F., Adamantidis, A., Stuber, G.D., Bonci, A., de Lecea, L., and Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science *324*, 1080–1084. doi:10.1126/science.1168878

Ulanovsky, N., and Moss, C.F. (2007). Hippocampal cellular and network activity in freely-moving echolocating bats. Nature Neuroscience *10*, 224–233. doi:10.1038/nn1829

Ward, R., and Ward, R. (2008). Selective attention and control of action: comparative psychology of an artificial, evolved agent and people. Journal of Experimental Psychology: Human Perception and Performance *34*, 1165–1182. doi:10.1037/0096-1523.34.5.1165

Womelsdorf, T., and Fries, P. (2007). The role of neuronal synchronization in selective attention. Current Opinion in Neurobiology *17*, 154–160, doi: 10.1016/j.conb.2007.02.002.

Appendix A

To create offspring, the first parent was chosen by 'tournament selection'. This was the agent with the highest fitness within a random subset of agents of one third of the population. To select the second parent agent, an adapted 'roulette-wheel selection' was applied. For each agent a portion of a roulette wheel (i.e., proportional to the probability of selection) was allocated. The probability of an agent (P_i) to be selected was calculated with the formula:

$$P_i = 0.5 \cdot \left(\frac{\mathrm{Re}\,versedDis\tan ce_i}{Sum\,\mathrm{Re}\,versedDis\tan ce} + \frac{\mathrm{Re}\,versed\,\mathrm{Re}\,ducedFitness}{Sum\,\mathrm{Re}\,versed\,\mathrm{Re}\,ducedFitness} \right) \quad (1)$$

This formula combined an agent's the resemblance to the first parent and the agent's *reduced* fitness. The resemblance to the first parent, its *ReversedDistance,* was measured as the inverse of the Euclidean distance between the networks, which in this context were represented as points in a multidimensional space, defined by the networks' connections. The agent's *ReducedFitness* was its fitness minus 90% of the lowest occurring fitness in the current generation.

The connection strengths of the parents were used to create two offspring. Half of the time each offspring was a copy of one of the parents and the other half they were a mixture of both parents. In case of a mixture, two random parts of the weights would be swapped. In addition, all the weights, ranging between –10 and 10, would mutate slightly by adding M, which was calculated with the formula:

$$M = -0.1 \cdot \log\left(\frac{1}{r} - 1 \right) \quad (2)$$

where r was a random value drawn from a uniform distribution between 0 and 1. The offspring were added to the population and the process was then repeated until the population size had tripled. All agents were considered as parents once again with the creation of each offspring pair.

Appendix B

The entities had a round shape with a radius of 10 length units. The environment was a torus shaped square which sides measured 400 length units. Smells were emitted by all entities in the environment and dissipated in all directions equally with geometrically decaying intensity (S) according to the formula:

$$\begin{aligned} \delta < \delta_{MAX} \quad &: \quad S = \frac{1}{1+\delta} \cdot S_{MAX} \cdot \left(1 - \frac{\delta}{\delta_{MAX}}\right) \\ \delta \geq \delta_{MAX} \quad &: \quad S = 0 \end{aligned} \tag{3}$$

with S_{MAX} = maximal smell intensity at the source (set to 25), δ = distance to the source (in length units) and δ_{MAX} = maximal distance at which the source was smelled (set to 100). The entities emitted multiple smells in different ratios. Plants emitted both smell 'a' and 'b' with the intensity of 'a' being half as strong as 'b'. With predators smell 'b' was half as strong as smell 'a'.

The agent's motor actuators exerted a force against the environments surface that propelled the agent straight ahead if both forces were equal or curved if the one was greater than the other. Each time step that force was exerted, the agent's energy was reduced by an amount $Energy_t$ according to the formula:

$$Energy_t = \frac{MaxEnergy}{1-\left(MotorAct_t - MaxAct\right)} \cdot \frac{MotorAct_t}{MaxAct} \tag{4}$$

where $MotorAct_t$ = activation of the actuator at time t, $MaxAct$ = the actuator's maximum activation (set to 1.0), $MaxEnergy$ = maximum consumption per actuator per time step (set to 0.001). In addition, the agent used a static energy cost of 0.001 per time step.

Appendix C

In the third stage agents are removed from the population. The probability of staying in the population for another generation would be determined in three steps. First a relative factor (*RF*) was calculated for each agent with the formula:

$$RF = \frac{b \cdot (F_i - F_{\min})}{F_{\max} - F_{\min}} + o \tag{5}$$

with F_i as the fitness of agent *i*, F_{min} as the lowest fitness in the current population, F_{max} as the highest fitness in the current population, *o* as the lower limit of *RF* (set to 0.15) and *b* as the range of *RF* values (set $1 - o = 0.85$). In the second step the relative factor was used to calculate the population factor which accounted for the influence of the current population size (*N*) with respect to the initial population size (N_0). If the current population was large, the agents' probability to be removed increased and vice versa. The population factor (*PF*) was calculated according to the formula:

$$\begin{aligned} N \geq N_0: &\quad PF = RF \cdot \frac{N_0}{N} \\ N < N_0: &\quad PF = 1 - \left((1 - RF) \cdot \left(1 - \left(1 - \frac{N}{N_0} \right)^2 \right) \right) \end{aligned} \tag{6}$$

In the third step the probability that the agent would remain in the population for another generation was calculated by multiplying the *PF* by 0.99. This ensured that even the fittest individual could sometimes be removed from the population.

Appendix D

The weights of the network's connection determine an agent's behavior. The network is symmetrical in the sense that all nodes have a mirror node and thus, for each connection running from node i to j there is a mirror connection from mirror node i' to j'. Activations are of course unique, otherwise the output nodes' activation would always be equal and the agent could only move forward. The symmetry greatly reduces the number of mutations required to evolve a functional behavior and is correct under the assumption that the agent should give an exact mirrored response in an exactly mirrored environment. The activation (y_i) of all nodes was calculated with the formula:

$$y_i = \sigma\left(\sum_{j=1}^{N} \omega_{ij} y_j + \theta_i + S_i\right) \tag{7}$$

The formula of the sigmoid was:

$$\begin{aligned} \xi \geq 0: &\qquad \sigma(\xi) = \frac{\xi}{1+\xi} \\ \xi < 0: &\qquad \sigma(\xi) = 0 \end{aligned} \tag{8}$$

with ω_{ij} = the weight of the connection from unit *j* to node *i*, *N* = the number of connections to node *i*, θ_i = the bias of node *i* (0.1 in the simulations) and S_i = the smell input to node *i* from the corresponding sensor (only for input layer). In order to create a temporal difference between processing over long neural pathways and short neural pathways, a time delay was introduced by storing a node's activation for one time step. Hence, the activation that was calculated for a node at time t was propagated over it's sending connections at time t+1. As a consequence, input signals were processed to the output nodes at t+1 via the direct projections and at t+2 via the indirect connections, through the hidden layer.

Part II

Nico H. Frijda
Amsterdam University

Impulsive Action and Impulse Control

Abstract

In this paper I will explore the phenomena of impulsive action, as a contribution to the psychological analysis of motivation.

I will be arguing the following points: 1). Emotions are motive states, including states of motive loss and motivational disorganization; 2). These motive states instigate impulsive actions; 3) Many emotional elicitors elicit multiple emotions; 4) Multiple emotions interact, producing a mixed emotion that differs from each of the multiple emotions; 5) Emotion regulation is one of the outcomes of such interaction; 6) Deliberate impulse control is due to failure of such multiple emotion interaction; 7) and so is emotional conflict.

1. Impulsive action and motivation

Only a short time ago I happened to come across a book by Joseph LeDoux that I was unfamiliar with: *Synaptic Self* (LeDoux, 2003). In that book I found a chapter on “The lost world”. It discussed the almost total absence of the basic study of motivation in contemporary psychology and neuroscience. After the disappearance of needs, drives, and instincts from psychology, *why* do humans do what they do?

I think LeDoux was right. The last monographs on motivation that I know of appeared about 30 years ago: Gallistel's (1980) *The organization of action* and Toates' (1986) *Motivational Systems.*

True, great advances have been made since. Much has come from brain science, as LeDoux's (2003) book attests. Much has also come from psychology (Davidson et al., 2003; Morsella et. al., 2009). But both worlds are still more separated than they should. Most work on action in psychology has focused the organization and execution of goal-directed action. Little work has been devoted, so far, on action that is best called "impulsive action": action that is not preceded by a goal about what one plans to achieve, but instigated by stimulus events and subsequent desire, and that is often difficult to control.

It may be profitable to examine what emotion psychology has to say about impulsive action. It is the more profitable, I think, because confusion has been created by the emergence of "dual processing models", and notably the distinction made between two different processing systems: an impulsive and a reflective system (Strack and Deutsch, 2004). These systems have been proposed to employ different processing principles and information representation modes. However, the presumed impulsive system lumps together two quite different kinds of behavior control whose only common factor is that both are unpremeditated[1].

2. Drive actions: Wundt

Wundt (1896), at the time, distinguished not two systems, but four different kinds of behavior: reflexes, habits, voluntary actions, and *Triebhandlungen, "*driven actions". Although both are unpremeditated, *Triebhandlungen* are not habits. The conditions for both sorts

---

[1] The distinction between processing systems that deal with action should not be confused with a distinction between modes of information representation and handling, as proposed by Kahneman (2012).

of behavior differ profoundly. Habits are acquired by learning (Strack and Deutsch, 2004). *Triebhandlungen* are not necessarily. McDougall (1923) translated the term as “instincts”. What he meant was that they are emotional actions. Though not deliberate, these are purposive and, furthermore, they do not generally result from overlearning. Impulsive actions in the sense of “Triebhandlungen” need result from learning. Killing someone in anger is not the outcome of long practice, except perhaps in a hired killer. They may even be made on the spot, as evidenced by the individual who, in a fit of rage, pauses to construct a stinging insult. I will use the term “impulsive actions” to designate them. They do not constitute a separate class of actions shaped by separate processing principles but rather form a separate mode of action initiation and selection, as I hope to show.

This paper has two aims. First, I will try to account for impulsive action as action elicited and driven by an acute motive state, often an “impulse” indeed. This is why emotional actions impress as “impulsive”. It is not their having been overlearned, although overlearning leads to using the same word “impulsive” for an ingrained habit. I will call the acute motive state a “state of action readiness” (Frijda, 1986; 2007).

Second, I will try to account for the notion that emotional actions are often actions that one seeks to inhibit or control. I will try to show that there are not two souls in one brain: one that has emotions, and one, using reason, that controls them.

I will argue the following main points.

First, emotions involve motive states, including states of motive loss. Motive states are mental states that instigate and maintain impulsive actions.

Second, complex events can elicit multiple motive states, and thereby lead to multiple impulsive actions, and to mixed emotions.

Third, multiple motive states can interact, and thereby entail impulse control or they lead to emotional conflict.

2.1. Analysis of impulsive action

What I mean by “impulsive actions” is best made clear by some examples. They are in part taken from Pacherie (2001)

Impulsive actions: examples

Fleeing in fear
Flaring up against an annoyer in a pub brawl
Following a person one is charmed by when passing in the street
Being unable to sleep when preoccupied by a problem one could not solve before going to bed
Eating more peanuts than one likes to eat
Drinking more than one is good for you
Buying things during a sale one does not need
Saying things that one will later regret
Interrupting someone speaking

This is how one can characterize in general what I call impulsive actions, in addition to the absence of prior goal setting and brief onset latency. Like the state of action readiness that motivates them, they are elicited by an event, the thought of an event, or some mental imagery of an event. Although not preceded by a goal, they do have direction. They have an *aim,* which specifies the aim of the motive state of action readiness in terms of the current situation: to achieve, modify, or maintain the relationship to the object, upon which achievement the action will terminate.

The actions include cognitive actions: thoughts and fantasies, modifications of beliefs and expectations. They usually manifest dynamic qualities, designated as “control precedence”. Control precedence includes persistence over time, overcoming obstacles and interruptions, and attentional selectivity. Illustrations include the frenzy of an alcoholic in search of a drink, the persistence of a trapped frog, dog, or cat that seeks to escape till exhaustion, continuing to scold angrily even after an apology, the average number of stab wounds in violent marital quarrels, interrupting someone who is talking, and the stubborn returning of one’s thoughts to a person one fell in love with. They manifest the consequences of control precedence, such as the limitation of intake of information from memory or the current situation to what is relevant to the dealing with the current aim. It is why impulsive actions often fail to reckon with harmful or disadvantageous consequences they might entail, but that could have been foreseen if one had taken the time and the cognitive trouble to do so.

In the present perspective, a motive state with an aim is thus hypothesized to intervene between occurrence of an event and an action. The reason to assume such an aim is the *equifunctionality* of different actions triggered by a given kind of event. An insult or an offense may elicit shouting, attacking, making harmful remarks, destroying a target’s cherished things, and turning one’s back, that all harm the target and may discourage the target from continuing what he or she does. An aim can be conceived of as the neural representation of the orientation towards actions to come (something like the “efferent copies” in the neurology of movement; see Jeannerod, 1995)

2.2. Core hypothesis of emotions

This embodies the present core hypothesis. “Emotion” involves a motive state or variant of state of action readiness. That is to say that motive states do not only occur as actual readiness for action, with muscles tensed. They can also exist as states of *latent* readiness, to be released when conditions are proper, and which sustains or "drives" the action during execution.

But the notion of action readiness is somewhat wider. States of action readiness include states of the loss of action readiness. All previous action attempts may have failed, or one may have been unable to identify any action program that appeared meaningful under the circumstances (like crying out loudly alone at night in the desert). Such states of motivation loss can persist. Apathy can persist for as long as anger can and, indeed, some cases of depression are exactly in this mould, as is acute anxiety.

States of action readiness can exist as mere orientations, as subthreshold neural states without any motor implementations (Jeannerod, 2006). They form the main substrate of feelings of anger or fear or any other emotional urge. Such virtual states of action readiness most probably form the mainstay of emotional reactions to fleeting impressions, value judgments, passing objects, imaginary objects, empathic participations in witnessing the emotions of other people, and witnessing the emotional meaning of objects of art (Frijda and Sundararajan, 2007).

Motive states have several features that describe or specify each of them, in the manner that Jackendoff (2007) describes the informational structure of actions. These features are derived from behavioral evidence, as well as from self-reported mention of aims, desires, and urges in recalled emotion incidents (Frijda, 1986, 2007; Davitz, 1969).

States of action readiness first of all have a particular content: their aim, the satisfaction condition for the actions which that state of readiness elicits, and that distinguishes the different modes of action readiness. They all are modes of *relational* action readiness: readiness to establish, maintain, or modify the subject's relation to some object. "*Approach*" specifies the aim of an increase of interaction; "*apathy*" specifies that no relational aim is held with any object, including information acquisition and so on. The aims involve readiness or unreadiness for action, that is: assessing some state of action readiness involves assessment that action with the given aim is occurring, or predicting that such action is likely. The notion is close to those of *attitude*, introduced for the same phenomena and functions by Claraparède (1928), Bull (1951), and Deonna and Teroni (2012), and of *positionality* by Frijda (1953). They are all focused on the relational nature of the construct, and its applicability to action.

Second, they permit a general flexibility of actions to form: they each allow a variety of specific actions, in accordance with the current situation, but that all instantiate the momentary action readiness's aim; it renders emotional states different from reflexes and habits. States of action readiness are each forms of being set to modify the given relationship, or to retain the current subject-object relationship, in states of readiness for acceptance, towards maintaining the current one as against disturbances and interferences; or they are oriented towards actions that produce a certain effect: for instance, a particular change in the current subject-object relationship.

And, third, they retroactively point to their source: how the eliciting event is or was appraised. But that source may not be accessible to conscious reflection. The source can consist of priming: the effect of stimuli that were not remarked, did not draw attention, and yet activated or facilitated some response. It has been observed and examined as a response to seeing someone else manifesting an emotional expression, or by an aim or urge to be mentioned during conversation (Bargh and Ferguson, 2000).

3. Appraisal

States of relational action readiness are instigated by how an event is appraised. The term *appraisal* is used for the information processes that enrich incoming event information, providing it with meaning that makes the event relevant to the individual. The information that can do that stems from the event as such, its current context, stored information associated to the event itself, and actual interactions with objects and other individuals (Ellsworth and Scherer, 2003; Scherer et al., 2001). Appraisal amounts to retrieving information on what can be expected from the event, what it may do or offer to the perceiver, or prevent him or her from doing, and to what he or she may do to cope or deal with that (Elster, 1999; Frijda, 1986, 2007; Lazarus, 1991; Oatley, 1992). Central for the emergence of emotions is that appraisal processes have assessed the event's consequences for the satisfaction or frustration of the individual's concerns and currently active goals. Events can be appraised as beneficial or harmful to some of those concerns and goals (Lazarus, 1991; Scherer, 2005).

These appraisals in particular lead to an evaluation of the event as pleasant or unpleasant, or of mixed affective value.

Appraisal processes mostly proceed automatically and unconsciously, as does so much other thinking (Dijksterhuis and Nordgren, 2006). They produce a profile of appraisal features that characterize the content of a given emotional reaction. Appraisals are examined by varying antecedent variables, for instance in studies with causal vignettes (Ellsworth and Scherer, 2003; Lazarus, 1991).

Different appraisals tend to activate different states of action readiness. Correlations between patterns of appraisal, assigned emotion labels, and motive states are observed (Frijda at al., 1989; Roseman, 2001). There is evidence for a fairly strict correspondence between patterns of appraisal and the resulting mode of action readiness.

4. Impulsive action selection: aims not goals

The major point about appraisal is whether or not the implications of events as appraised instigate a motive state: a state of action readiness for establishing, changing, or maintaining a particular relation between the subject and some object. The total impact of the event determines the nature of the motive, the strength of the impulse, and the motive state's control precedence. The appraisal can be taken to be the stimulus for these latter aspects of action and feeling. It activates the midbrain dopamine circuit that confers incentive value to the object or event at hand, energizes the processes of control precedence and subsequent degree of striving (Berridge, 2004; LeDoux, 2003).

Remains identifying an appropriate action. Generally actions are selected and guided by a goal representation, that is, the representation of a future state, forming and preceded by an intention. But emotions of some intensity tend to induce impulsive actions that are not preceded by deliberation, nor are they intentional in the strict sense.

But, action being impulsive by definition, there is no prior intention of what outcome to reach. There is no prior goal. There is no prior representation of which situation to achieve, which is a condition for goal directed action (Kruglanski, 1996). How then to find an appropriate action?

Actually, under emotional conditions there is usually no need for such a representation of a future state. What calls and guides action is the aim of the current disposition to act. The action readiness aims at the present situation, with its affordances and obstacles. That is: the appraisals that elicit the motive state dictate a change from *now*, or *no* change from now, and in which the aspect that demands change is located at the core of the affective appraisal. When afraid, one is not impelled to get to a place of safety, but to get anywhere where the discomfort is less. When vividly enticed, one is not impelled to obtain satisfaction, but to give in to the attraction of the object one faces or expects, and to permit consummatory actions to go that already were put in readiness, such as stretching arms and pointing lips. Both actions operate as if obeying attractors in the dynamic systems sense of the term. Images of what is waiting there upon close proximity, in being attracted, or where you are when further from the threat – formation of a goal representations – can come later, if at all. You were engaged in impulsive action!

What may be evident from the analysis is that impulsive actions do not form a special class of actions – overtrained ones, or innately prepared ones. It is their being fed by action readiness and their fitting a pre-existing aim.

I have no space here to enlarge on what happens when no action program can be found, no aim can be formed, by profound disorientation, by full despair or by exhaustion. It must suffice to point out that this is the domain of responses of emotional failure: anxiety, panic, deep despair, and apathy.

5. Complex events elicit complex appraisals

My analysis of impulsive action might suggest that emotionally meaningful events, for a given person and at a given moment, give rise to one given appraisal profile that fits one mode of action readiness. This is not the case.

Emotional events rarely occur in isolation. They consist of more or less extended events and interactions; and within some context. Each may represent a loss as well as a gain, a challenge as well as a threat. That is to say: one and the same event allows very different appraisal profiles to emerge, and often at the same time. By consequence, multiple motive states, and thus multiple emotions, can be aroused at the same time.

“Multiple emotions” means primarily that several different action tendencies are evoked at the same time. They make one inclined to approach as well as to avoid. They make one inclined to hostility as well as to implement tenderness, or to withdraw at the same time as desiring to come closer to someone. One can want to yield to temptation, at the same time as one wants to resist it. The constellation is commonplace. Being treated unkindly by one’s spouse may evoke anger, but one’s very anger will hurt one’s beloved spouse. At least, it threatens to spoil one’s evening together.

There is nothing mysterious in multiple simultaneous action tendencies. They often show a complexity of expressive behavior. For instance, a person may look at someone, but thereby not facing the other in full. Or one can be tense, show a defiant glance, and clench one’s fists, but at the same time lean backwards so as to disengage from more encompassing personal interaction. One can intently focus on the person one is dealing with, entertain close contact, and also extend care. And one can mingle affection, invitation, and abandoning oneself, in a unique blend that closely appears to mirror the momentary shades of feeling and readiness for further interaction.

The motive state interpretations of expressive behaviors are in the process of validation, by having subjects score pictures of facial expressions in terms of action tendencies, and then verify how well the subjects agree in their assignments. Agreements were of the same order of magnitude as those of emotion name assignments to those same pictures (Tcherkassof et al., 2007).

6. Multiple emotion interaction

Multiple states of action readiness interact and tend to fuse in experience, as orange results from red and yellow. Their action preparations may interfere, or claim the same resources.

They may result in mixed emotions (Shimmack, 2001), “mixed feelings”, in which each motive state is attenuated or modified by the other. In approach-avoidance conflict (Miller, 1959), the rat running to the goal box that contains food but also emits shocks slows down and comes to a stop at some distance from the box.

The interactions of simultaneous affects have been studied extensively by Cacioppo and colleagues (e.g. Cacioppo et al., 2004). Mixed feelings often lack the transparency of simple feelings. They may be sensed as instable (e.g. Bargh et al., 1992; Wilson et al., 2001).

An example of a mixed feeling is nostalgia. The feeling fuses pleasure and pain, with pleasure from an old memory moderated by that pleasure being past, and the pain that its pleasant content is over (Bellelli and Saldarelli, 1990).

Multiple emotions may lead to outcomes other than attenuation. A major one is that one of the motive states fully overrides the other one. Control precedence goes fully to one of them, and attention goes fully to its content. “Come hell and high water!” implies neglect or avoidance of what is relevant to the other one, at least with respect to conscious processing.

7. Emotion regulation results from the interaction of multiple motive states

This perspective of interacting states of action readiness finds support in the phenomena of emotion regulation or impulse control. These phenomena include action attenuation, action suppression, seeking alternative appraisals of the event, and efforts to get rid of unwanted motive states, like anger, temptation, or greed. In fact, regulation appears to be largely the result of emotion interactions. It notably comes from foreseeing the aversive emotional consequences of one’s unrestrained emotional impulse. Anger is often held back by fear of retaliation, by dislike of offending social propriety, or by fear of loss of love. Fear for the risks of facing some menace conflicts with fear of appearing a weakling, or with the sheer pain of fear itself. Desire for drink or drugs faces the foresight of a hangover, a blackout, or loss of self-regard.

The antecedents of emotion regulation suggest this generalization: emotion regulation itself is emotional. It is frequently as emotional as is the regulated emotion. One does not want to hurt one’s spouse, even if he or she just hurt you. One does not want to become a useless or despised drunk, but one also wants to escape from the pains of abstention and so on.

This can be said somewhat more precisely. Emotions are regulated and impulses are controlled to the extent that one cares about the consequences of giving in to unregulated emotions. Whether one does or does not care depends on the relative strength of the competing motive states: such as concern for the wellbeing of one’s spouse, for social harmony, for one’s self-esteem, besides those for standing up for oneself, for the pleasures of drink or drugs. And the perspective of course explains the absence of signs of regulation when one is convinced that one's anger is fully justified, or when one feels entitled to one's sexual inclinations, when approving of them, or draws pride from giving in to them. Emotion regulation thus is to a large extent a matter of preference. It always is uncertain whether absence of regulation is due to weakness of restraining impulses or abilities, or to strength of emotional urges, and the priority of concerns. Whether one regulates or not depends on the price one is willing and capable to pay (Lewis, 2011), identity loss, risk of death, and suicide including (Frijda, 2010). The latter point is evident from the analysis of self-sacrifice (what Kuhl and Koole, 2004, called self-maintenance rather than self-control, such as resisting betrayal under torture, remaining faithful to a sick partner). It is also evident from the risk-taking of heroes and of terrorists (Reykovski, 2001; Kruglanski, 2008).

True enough, holding on to the selected choice or preference may well be beyond one's resources. (Baumeister at al., 2000; Frijda, 2010).

8. The control of impulsive action

Control of impulsive action is thus often due to the interaction between the motive state that generated the impulsive action under concern and a simultaneous other state of action readiness. A hostile impulse may interact with simultaneous affinitive readiness, which both together may generate friendly rather than angry reproach. Inclination to obtain the pleasures of one drink for the road interacts with the discomfort of envisaging the unpleasantness of a hangover and self-reproach, or just the efforts of driving home with an unclear mind faces the impulse to retain self-respect. The action may consist in declining the drink.

The point is this: Many such interactions are themselves impulsive, that is, non-deliberate. They are emotional, affect-driven. The processes include the interferences and summations caused by the two simultaneous motive states, as observed by Cacioppo et al., (2003). They may include the restrictions of attentional range contingent upon control precedence, as discussed by Easterbrook (1959), avoidance of cognitive exploration (Derakshan et al., 2007), slowing down, speeding up, or halting of approach behaviors (Miller, 1959), and inhibitory action tendencies elicited by conflict between the motive states (Gray and McNaughton, 2000; Rothbart and Sheese, 2007).

All these are automatic processes. They are part of emotional processes generally, to be understood from the dynamics of interacting multiple motive states. There is nothing here that calls for specifically regulatory processes.

One motive state overriding the other, too, is an automatic process, following differences in control precedence of the simultaneous motive states.

Regulation – response attenuation – is also achieved automatically, and without effort, as an outcome of the individual having acquired action skills that serve multiple aims. Such skills form the gist of socialization, by which ready-made actions with multiple aims have become available when one of the motive states is evoked, and the foresight of what it may evoke in the opponent evokes the other (Campos et al., 2004; Mesquita and Albert, 2007). Children learn the skills of combining hostile interaction and friendly interaction during rough and tumble play. They also learn the skills of negotiating rather than fighting about possessions. Variations in social interactions yield immediate profits, such as increase in social pleasures and both interactants sharing some other gain. One would not be inclined to view these skills as falling under the heading of "emotion regulation"; yet, that is their effect.

Note that such modulations of impulsive actions also occur among animals other than humans (De Waal, 1969) and that applying the relevant skills may get lost when the motive states are too strong, as when an enraged male chimpanzee swings an infant around by its arm under the terrified shrieks of onlookers, whereas under other circumstances he lets its pelt be pulled without bothering (Goodall, 1986).

Much emotion regulation, in other words, appears to proceed without effort or regulatory intent. Part at least of control of impulsive action is due to elementary processes of motive states, and their mutual influence. It would appear to result from automatic information processes of multiple appraisals, and interaction of simultaneous states of action readiness. It uses mental and bodily actions that are largely a part of the individual's standard action repertoires, as well as of his or her repertoire of action modes that each represent some action compromise or synthesis, as in the friendly reproach example.

9. Conflict

The domain of emotional reactions is, however, replete with contingencies in which motive states are incompatible. There is no automatic solution of handling the multiplicity and it forms an impasse from which there is no obvious escape.

There are various outcomes. One is emotional awareness of conflict: panic, aimless excitement, profound distress, fruitless search for a way out. A second one is a transformation of conscious awareness: dissociation (Hilgard, 1977), the sense of detachment, numbing, decay of any motive state, as anybody can experience during skidding in one's car, or when one's partner tells you that he or she is leaving, or during torture (Basoglu et al., 1997; Bryant, 2007).

10. Effortful emotion control

Yet a good measure of detachment can be achieved in a different way.

Unsolvable conflicts are probably the main incitements for the emergence of consciousness. Conscious awareness of conflict, according to Morsella's Supramodular Interaction Theory (SIT; Morsella, 2005), is the mechanism for integrating information from different supramodular response systems, such as different high-level concerns.

Articulate consciousness may well serve to explicitate information that might give access to ways for integration.

Impulse driven action then can make place for cognitive articulation. It may make room for the motive state of observing and acquiring information, as can be engaged in on day 2 or so of a human baby, witness its motionless gazing with wide-open eyes.

It leads to intentionally stepping out of interaction with ongoing events. One adopts an observational or a reflective stance, in which, for the moment, events do not concern oneself, – the same attitude as during skidding in one's car, but voluntarily so. One momentarily abandons being set to act. One postpones acting, and considers what actions there are that can be taken.

In adopting an observational stance, one is not engaging a "reflective system". Rather, one engages reflective or "executive" processes," such as monitoring one's ongoing cognitive processes, goal constructions based on conscious appraisals, and constructing and maintaining action plans, and maintaining goal-settings in working memory (Miyake et al., 2000).

Such goal construction replaces the mere aims of motive states by goal representations, and these goals, in turn, allow the various regulation strategies described in the bulk of regulation literature. One flees towards safety, one silences an antagonist and one seeks deliberately to suppress overt angry actions and seeks to reappraise the event.

Deliberate and effortful regulation is the rule in the absence of multiple emotions, but when the adverse effects of an ongoing motive are still known about. One smokes; one knows about lung cancer; but that event is so unreal, so far away, that it lacks emotional force. Its impact suffers from "time discounting": the law that future emotional impacts decrease in strength hyperbolically with the time in the future that they may materialize (Ainslie, 2001). Actions to forestall or prevent those effects can still be devised by rational considerations and, perhaps, by deliberately evoking emotions with the help of imagery of miserable future situations.

Effortful emotion control can be considered to represent a failure of impulse-driven emotion control that is the everyday way of emotion regulation.

11. Do we understand impulsive action and impulse control?

I have left out discussing the major and basic ingredient of the analysis of impulsive action and motive states: the determinants of how much we care and what we care about.

I have left out the basic sources of motivation. I discussed short-term motivation: that which drives you to scold and to embrace and to consume. But I almost entirely left out discussing long-term or dispositional motivation: that which allows us to have as well as to miss self-worth, affections, sexual interactions, social propensities, that which I called *concerns* (Frijda, 1986, 2007). I mentioned in passing that emotions arise when events are appraised as relevant to one or more of the individual's sensitivities or concerns**.** One has an emotion when an event makes one care. In older psychology, one tried to account for the bases for emotion and caring by notions like drives, needs, the conceptions of libido and Lorenz-like instincts, all grasped at the time by the hydraulic metaphor.

The hydraulic metaphor has been replaced by the conception of incentive motivation (Bindra, 1978; Bolles, 1975, and Toates, 1986). The incentive motivation conception has brought us a fairly coherent model of the dynamics of motivated or impulsive action, in the workings of dopamine, and perhaps other neuropeptides. But it has not, as yet, offered a coherent replacement for LeDoux's (2003) lost world. It offers no systematic treatment of concerns, let alone human concerns.

Current treatments of dispositional motivation or concerns discuss the domain in two unsatisfactory ways. One is by using a blanket term that avoids all theoretical problems: anything motivates that is or was a reinforcement. But “reinforcement” is not a theoretically satisfactory concept. It refers to an effect (future increase of that response), but not to a process. The other is by stating that a particular sensitivity is evolutionarily ever so useful, thereby leaving no way to account for the concerns that offer no obvious advantage for reproduction, and that may not be direct parts of the human biological make-up (Mesquita and Leu, 2007).

Yet, the world shows a number of concerns that make people happy when they are satisfied, or for which people are willing to die if they are not, in some places of the world (Schwartz, 1992). Recent experimentation shows how experimentally induced relevance of values enhances striatum responses to value-relevant stimulus items (Brosch et al., 2011).

One of the obstacles in trying to obtain a coherent account of dispositional motivation consist of the hierarchical or even heterarchical structure of concerns (Frijda, 2007, 2010a; Gallistel, 1980; Kortlandt, 1955). One has one’s partners and other close affiliates: their well-being constitutes as many concerns. One also has a “need to belong” (Baumeister and Leary, 1995). This structure does not make recuperating the lost world any easier. It is about time to set to the task, at more detailed level than examining and specifying the neural and neurohumoral mechanisms.

12. Concluding words

I hope this may contribute to understanding impulsive behavior. Impulsive action in the sense that I have examined is a kind of behavior *sui generis*, not to be confused with habits. On the contrary, it sheds light on the mechanisms that underlie short-term motivation, and variations in desire. It also points to the gaps of insight in long-term or dispositional motivation that make up the human mind in contrast to the minds of most other animal species.

I think that my analyses also shed light on some major puzzles of emotion regulation. It opens up, I think, some more dynamic and emotional perspectives on such regulation.

References

Ainslie, G. (2001). Breakdown of Will (Cambridge: Cambridge University Press).

Bargh, J.A., and Ferguson, M.L. (2000). Beyond behaviorism: On the automaticity of higher mental processes. Psychological Bulletin *126*, 925–945.

Basoglu, M., Paker, M., Paker, O., Ozmen, E., Marks, I., Incesu, C., Sahin, D., and Sarimurat, N. (1994). Psychological effects of torture: a comparison of tortured with nontortured political activists in Turkey. American Journal of Psychiatry *151*, 76–81.

Baumeister, R.F., and Leary, R.M. (1995). The need to belong: Desire for interpersonal attachment as a fundamental human motivation. Psychological Bulletin *117*, 497–529.

Baumeister, R.F., Muraven, M., and Tice, D. (2000). Ego depletion: A resource model of volition, self-regulation, and controlled processing. Social Cognition *18*, 130–150.

Bellelli, G., and Saldarelli, R. (1990). The representation of affective states: First observations on nostalgia. Paper presented at Meeting European Association of Experimental Social Psychology, Budapest, June 1990.

Berridge, K.C. (1995). Food reward: Brain substrates of wanting and liking. Neuroscience and Behavioral Reviews *20*, 1–25.

Berridge, K.C. (2004). Motivation concepts in behavioral neuroscience. Physiology and Behavior *81*, 179–209. (dozen)

Berridge, K.C., and Aldridge, W. (2009). Decision utility, incentive salience, and cue-triggered “wanting”. In: Oxford Handbook of Human Action, E. Morsella, J.A. Bargh, and P.M. Gollwitzer, eds. (Oxford: Oxford University Press), pp. 509–539.

Berridge, K.C., and Winkielman, P. (2003). What is an unconscious emotion? (The case for unconscious “liking”). Cognition and Emotion *17*, 181–211.

Bindra, D. (1978). How adaptive behavior is produced: A perceptual-motivational alternative to response-reinforcement. Behavioral and Brain Sciences *1*, 41–91.

Botvinick, M.K., Braver, T.S., Barch, D.M., Carter, C.S., and Cohen, J.D. (2001). Conflict monitoring and cognitive control. Psychological Review *108*, 624–652.

Brooks, V. and Hochberg, J. (1960). A psychophysical study of "cuteness". Perceptual and Motor Skills *11*, 205.

Brosch; T., Coppin, G., Scherer, K.R., Schwartz, S., and Sander, D. (2011) . Generating value(s): Psychological value hierarchies reflect context-dependent sensitivity of the reward system. Social Neuroscience *6*, 198–208.

Bryant, R.A. (2007). Does dissociation further our understanding of PTSD? Journal of Anxiety Disorders *21*, 183−191.

Cacioppo, J.T., Larsen, J.T., Smith, N.K., and Berntson, G.G. (2004). The affect system: What lurks below the surface of feelings? In: Feelings and Emotions: The Amsterdam Symposium, A.R.S. Manstead, N.H. Frijda, and A.H. Fischer (Cambridge: Cambridge University Press), pp. 223–242.

Claparède, E. (1928). Emotions and feelings. In: Feelings and Emotions: The Wittenberg Symposium, M.L. Reymert, ed. (Worcester, Mass.: Clark University Press), pp. 124–139.

Damasio, A. (1994). Descartes Error. Emotion, Reason, and the Human Brain (New York: Putnam).

Davidson, R.J., Scherer, K.R., and Goldschmidt, H.H. (eds.) (2003). Handbook of the Affective Sciences (Oxford: Oxford University Press).

Davitz, J.R. (1969). The Language of Emotion (New York: Academic Press).

de Gelder, B., Vroomen, J.H.M., and Pourtois, G.R.C. (2001). Covert affective cognition and affective blindsight. In: Out of Mind, B. De Gelder, E. de Haan, and C. Heywood, eds. (Oxford: Oxford University Press), pp. 205–221.

Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness. Basic evidence and a workspace framework. Cognition *19*, 1–37.

Deonna, J., and Teroni, F. (2012). The Emotions: A Philosophical Introduction (New York: Routledge).

Derakshan, N., Eysenck, M.B., and Myers, L.B. (2007). Emotional information processing in repressors. Cognition and Emotion *21*, 1585–1614.

De Waal, F.B.M. (1996) Good Natured: The Origins of Right and Wrong in Humans and Other Animals (Cambridge, Mass.: Harvard University Press).

Dijksterhuis, A., and Nordgren, L.F. (2006). A theory of unconscious thought. Perspectives on Psychological Science *1*, 95–109.

Ellsworth, P.C., and Scherer, K.R. (2003). Appraisal processes in emotion. In: Handbook of the Affective Sciences, R. Davidson, K.R. Scherer, K.R., and Goldsmith, H.H., eds., (Mahwah: Erlbaum), pp. 572–596.

Frijda, N.H. (1953). The understanding of facial expression of emotion. Acta Psychologica *9*, 294–362.

Frijda, N.H. (1986). The Emotions (Cambridge: Cambridge University Press).

Frijda, N.H. (2007). The Laws of Emotion (Mahwah: Erlbaum).

Frijda, N.H. (2009). On the nature and function of pleasure. In: Pleasures of the Brain: The Neural Basis of Sensory Rewards, K. Berridge, and M. Kringelbach, eds. (Oxford: Blackwell), pp. 99–112.

Frijda, N.H. (2010). Not passion's slave. Emotion Review *2*, 68–75.

Frijda, N.H. (2010a). Impulsive action, and motivation. Biological Psychology *84*, 570–579.

Frijda, N.H., and Sundararajan, L. (2007). Emotion refinement. Perspectives on Psychological Science *2*, 227–241.

Goodall, J. (1986). The Chimpanzees of Gombe: Patterns of Behavior (Cambridge: Belknap).

Gray, J.A., and McNaughton, N. (2000). The Neuropsychology of Anxiety (Oxford: Oxford University Press).

Jackendoff, R. (2007). Language, Consciousness, Culture: Essays on Mental Structure (Cambridge, Mass.: MIT Press).

Jeannerod, M. (1995). The Cognitive Neuroscience of Action (Oxford: Blackwell).

Jeannerod, M. (2006). Motor Cognition: What Actions Tell the Self (Oxford: Oxford University Press).

Kahneman, D. (1999). Objective happiness. In: Foundations of Hedonic Psychology: Scientific Perspectives on Enjoyment and Suffering, D. Kahneman, E. Diener, and N. Schwarz, eds. (New York: Sage), pp. 3–25.

Kahneman, D. (2012). Thinking, Fast and Slow (New York: Farrar, Strauss and Giroux).

Kruglanski, A.W. (1996). Goals as knowledge structures. In: The Psychology of Action: Linking Cognition and Motivation to Behavior, P.M. Gollwitzer, and J.A. Bargh, eds. (New York: Guilford Press), pp. 599–618.

Kruglanski, A.W., Chen, X., Dechesne, M., Fishman, S., and Orehek, E. (2008). Fully committed: Suicide bombers' motivation and the quest for personal significance. Political Psychology *30*, 331–357.

Kruglanski, A.W. (2004). The Psychology of Closed Mindedness (Hove: Psychology Press).

Kuhl, J., and Kazén, M. (1999). Volitional facilitation of difficult intentions: Joint activation of intention memory and positive affect removes stroop interference. Journal of Experimental Psychology: General *128*, 382–399.

Kuhl, J., and Koole, S,L, (2004). Workings of the will: A functional approach. In: Handbook of Experimental Existential Psychology, J. Greenberg, S.L. Koole, and T. Pyszczynski, eds. (New York: Guilford), pp. 411–430.

Lazarus, R.S. (1991). Emotion and Adaptation (New York: Oxford University Press).

LeDoux, J. (2003). Synaptic Self: How Our Brains Become Who We Are (London: Penguin Books).

Lewis, M.D. (2011). Dopamine and the neural "now": Essay and review of addiction: a disorder of choice. Perspectives on Psychological Science *6*, 150–155.

Lorenz, K. (1971). Part and parcel in animal and human societies. In: Studies in Animal and Human Behavior, vol. 2 (Cambridge, Mass.: Harvard UP, originally pub. 1950).

McDougall, W. (1923). Outline of Psychology (New York: Scribner).

Mesquita, B., and Leu, J. (2007), The cultural psychology of emotion. In: Handbook of Cultural Psychology, S. Kitayama, and D. Cohen, eds. (New York: Guilford Press), pp. 734–759.

Miller, N.E. (1959). Liberalization of basic S-R concepts: Extension to conflict behavior, motivation and social learning. In: Psychology: A Study of a Science, vol. 2., S. Koch, ed. (NewYork: McGraw-Hill), pp. 196–292.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., and Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cognitive Psychology *41*, 49–100.

Morsella, E. (2005). The function of phenomenal states: Supramodal interaction theory. Psychological Review *112*, 1000–1021.

Morsella, E., Bargh, J.A., and Gollwitzer, P.M., eds. Oxford Handbook of Human Action (Oxford: Oxford University Press).

Nafe, J.P. (1924). An experimental study of the affective qualities. American Journal of Psychology *35*, 507–544.

Pacherie, E. (2001). The role of emotions in the explanation of action. European Review of Philosophy, 53–91.

Posner, M.I., and Rothbart, M.K. (1998). Attention, self-regulation and consciousness. Philosophical Transactions of the Royal Society of London *353*, 1915–1927.

Posner, M.I., and Rothbart, M.K. (2007). Research on attentional networks as a model for the integration of psychological science. Annual Review of Psychology *28*, 1–23.

Quirk, G.J. (2005). Prefrontal-amygdala interactions in the regulation of fear. In: Handbook of Emotion Regulation, J.J. Gross, ed. (2007) (New York: Guilford Press), pp. 27–46.

Reykovski, J. (2001). Justice motive and altruistic helping. In: The Justice Motive in Everyday Life, M. Ross, and D.T. Miller, eds. (New York: Cambridge University Press).

Roseman, I.J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In: Appraisal Processes in Emotion: Theory, Methods, Research, K.R. Scherer, A. Schorr, and T. Johnstone, eds. (New York: Oxford University Press), pp. 68–91.

Rothbart, M.K., and Sheese, B.E. (2007). Temperament and emotion regulation. In: Handbook of Emotion Regulation, J.J. Gross, ed. (New York: Guilford Press), pp. 331–350.

Schimmak, U. (2001). Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? Cognition and Emotion *15*, 81–98.

Scherer, K.R. (2005). What are emotions? And how can they be measured? Social Science Information *44*, 695–729.

Scherer, K.R., Schorr, A., and Johnstone, T., eds. (2001). Appraisal Processes in Emotion: Theory, Methods, Research (New York: Oxford University Press).

Searle, J.R. (1983). Intentionality: An Essay in the Philosophy of Mind (Cambridge: Cambridge University Press).

Shizgal, P. (1999). On the neural computation of utility: Implications from studies of brain stimulus reward. In: Foundations of Hedonic Psychology: Scientific Perspectives on Enjoyment and Suffering, D. Kahneman, E. Diener, and N. Schwarz, eds. (New York: Russell Sage), pp. 500–524.

Strack, F., and Deutsch. R. (2004). Reflective and impulsive determinants of social behavior. Personality and Social Psychology Review *8*, 220–247.

Toates, F.M. (1986). Motivational Systems (Cambridge: Cambridge University Press).

Tcherkassof, A., Bollon, T., Dubois, M., Pansu, P., and Adam, J.-M. (2007). Facial expression of emotions: A methodological contribution to the study of spontaneous and dynamic emotional faces. European Journal of Social Psychology *37*, 1325–1345.

Weiskrantz, L. (1997). Consciousness Lost and Found: A Neuropsychological Exploration (Oxford: Oxford University Press).

Wilson, T.D., Lindsey, S., and Schooler, T.Y., (2001). A model of dual attitudes. Psychological Review *107*, 101–126.

Winkielman, P., Berridge, K.C., and Wilbarger, J. (2000). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. Personality and Social Psychology Bulletin *31*, 121–135.

Wundt, W. (1896). Grundriss der Psychologie (Stuttgart: Engelmann).

Zelazo, P.D., and Cunningham, W.A. (2007). Executive function: Mechanisms underlying emotion regulation. In: Handbook of Emotion Regulation, J.J. Gross, ed. (New York: Guilford Press), pp. 135–158.

James A. Russell
Boston College, Chestnut Hill, MA, USA

Emotion, Core Affect, and Psychological Construction

As an alternative to using the concepts of emotion, fear, anger, and the like as scientific tools, this article advocates an approach based on the concepts of core affect and psychological construction, expanding the domain of inquiry beyond "emotion". Core affect is a neurophysiological state that underlies simply feeling good or bad, drowsy or energised. Psychological construction is not one process but an umbrella term for the various processes that produce: (a) a particular emotional episode's "components" (such as facial movement, vocal tone, peripheral nervous system change, appraisal, attribution, behaviour, subjective experience, and emotion regulation); (b) associations among the components; and (c) the categorisation of the pattern of components as a specific emotion.

In fear, your heart races, your palms sweat, your face broadcasts fear, you scream, and you flee. But, does this happen in all cases of fear? Most? In the realm of the emotions, reflex-like consistency is the exception rather than the rule. The rule is differences–both between and within individuals and situations. As Barrett (2009) describes, there is more variety within fear and other categories of emotion than our standard theories suggest–indeed, more than commonsense supposes. To understand these differences in emotion requires a conceptual framework that anticipates differences and that supports a more idiographic approach to research on them. It requires a new way of thinking about emotion. This article outlines one such approach.

The major barrier to progress is not our ignorance but our illusion of knowledge. Increasingly, our traditional "knowledge" about emotion is being questioned at the most fundamental level. The traditional assumption that emotion is a unitary event is challenged by the idea that emotion is a multi-component process, with no one component identified with the emotion (Scherer, 2001). A similar challenge is Clore and Ortony's (2008) argument that an emotion is an emergent construction rather than a latent entity and their call for a shift from discussing "emotion" to discussing "affective processes". Minsky (2008) called for wholesale remodelling of this field. Kagan (2007) challenged the language used in this field and its methods for gathering knowledge.

There are many exciting avenues to explore and new ways to think about emotion. My version has been presented at length with supporting evidence elsewhere (Russell, 2003, 2006). Here I clarify, develop and defend that approach. First, however, I briefly summarise some problems with the traditional way of thinking about this topic.

1. The stranglehold of common sense

Citing Tomkins (1962, 1963) and his followers (Ekman, 1984; Izard, 1977), much of the psychology of emotion has been guided by a highly productive research program that I call basic emotion theory. The central idea of basic emotion theory is that human nature includes a small number of qualitatively distinct kinds of emotion, each of which produces (or is) a distinct, tightly organised and recurring pattern of manifest components. Basic emotion theory is a plausible approach that has stimulated the gathering of much evidence. As often happens in science, however, the evidence has revealed surprising problems, which I summarise in Table 1. No one or two problems are fatal, but, altogether, the problems point to a poor prognosis for this type of theory.

Table 1. Problems uncovered in studies of basic emotion theory

Area	Problem
Culture	There are cultural differences in all known aspects of emotion.
Language	Different languages lack a one-to-one correspondence between emotion terms.
Definitions	Theories based on traditional assumptions have not led to increased precision of terms. Each term lacks inclusion and exclusion rules.
Blends	Basic emotions rarely occur alone, and yet no accepted theory of how they co-occur or blend has been developed.
Facial expressions	Failure to find convincing evidence that emotions produce “facial expressions of emotion”.
Autonomic nervous system	Failure to find convincing evidence of a unique pattern for each emotion in the autonomic nervous system.
Subjective experience	Failure to find separate factors corresponding to basic emotions in studies of self-reported emotional experience.
Emotional behaviour	Failure to find a class of behaviour common to instances of a given emotion.
Coherence	Dissociation rather than predicted associations among manifest components.

Rather than elaborate on the problems listed in Table 1, let me instead suggest a diagnosis of their source. The problems encountered by basic emotion theory are not philosophical but empirical. They are problems encountered by most analyses of emotion, modern and ancient, and in diverse disciplines. They stem from the preconceptions that underlie not only much of scientific thinking but much of our everyday thinking about emotion. The source is deeper than simply the answers provided by basic emotion theory, which is, after all, just a more explicit and systematised version of common knowledge. The problem already exists in the questions that this family of emotion theories attempts to answer. Questions such as: What is an emotion? How many emotions are there? What are they? How are they generated? How does one emotion differ from another? What are their manifestations? What are their effects/functions?

Long before any psychologist asked these questions and theorised about the answers, everybody who speaks English knew about anger, fear, jealousy and other emotions. Indeed, we simply opened our eyes and saw a friend become frightened. We saw ourselves erupt in anger. We felt sad. Everyone knows that smiles mean happiness, tears mean grief, scared people run away, and angry people fight. Everyone knows that emotion is the antithesis of reason, emotional behaviour the antithesis of deliberate action. The science of emotion began with these compelling perceptions based on and reinforcing a set of preconceptions.

But, history and science combine to show us that what we see does not always reveal the way things are. We open our eyes and see one line longer than another in the Müller–Lyer figure, even though we know they are equal in length. We open our eyes and see the sun rise, move through the sky, and then set. Except, of course, that it does no such thing. “Seeing” is a powerful force on our thinking, and centuries of scientific work were needed to replace a geocentric theory of the solar system with a heliocentric one.

Science begins in the everyday human attempt to understand the world. Infants are scientists (Gopnik and Meltzoff, 1997) who develop their theories of the world filtering data through human perceptual and cognitive processes. As science progresses, however, everyday folk theories are typically left behind. Some everyday concepts are forged into scientific concepts; some not; all must be scrutinised. Contrast traditional beliefs with what is now known about the size of the universe, the age of planet earth, the origin of species, or quantum mechanics. None of these breakthroughs in science has been easy or without powerful resistance.

In the realm of emotion, scientific progress has been hindered by our everyday vocabulary and the assumptions it carries. One source of problems is the very word *emotion*. Frijda (2008) wrote, "The psychological point of view is that 'emotion' represents a meaningful and necessary concept" (p. 68). Despite decades of attempts to forge the concept of emotion into something scientific, we are still left with too many problems. The everyday folk concept of emotion is culture specific and vague and comes with misleading presuppositions. Its extension is heterogeneous, so that writers have tried to subdivide it into basic and non-basic emotions, affects, sentiments, moods, and so on, but with little consensus. Emotion is divided into joy, anger, fear, jealousy, and so on–but the problem lies in the "and so on" because no one knows where the border is between emotions and non-emotions. The everyday concept of emotion presupposes: that an emotion is separate from its causes, its manifestations, and its consequences, although different writers have proposed equating emotion with one or more of these; that an emotion (or its equivalent, an affect programme or neural module) is an entity that causes these manifestations (expression, instrumental action, peripheral nervous system changes, and so on); that emotion is an entity that is qualitatively different from other psychological entities, especially the mechanisms of rational thought, but also behaviour, conation, and so on. From these assumptions it follows that the word *emotion* (perhaps after some tidying up) demarks a special scientific domain of inquiry. Writers continue to seek the essence of emotion and to ask unanswerable questions.

Anger, fear, jealousy and other concepts for kinds of emotion are similarly embedded in a way of thinking that must similarly be questioned. They too are vague, culture specific, and come with misleading presuppositions. Commonsense presupposes separate discrete emotions, each a thing-like entity with causal powers. Thus anger is "expressed" in the face and voice, is "felt" in consciousness, "causes" the heart to speed up, and "urges" us to intemperate action. It is further assumed that because the various components stem from a single entity, they cohere in tight packages, each named in natural language. As a result, it is assumed that different emotions are separable into scientifically useful discrete kinds, each captured by a familiar word. Each has an essence.

Abandoning the concepts presupposed in the language we speak–indeed, in our perceptions of the world–as scientific tools is not easy. Without these concepts, it is difficult to speak or write about this domain. Rejecting these concepts as scientific terms appears to deny commonsense and obvious facts of the case. Doing so appears to abandon not only the answers our field has developed, but the very questions that motivated our work and the optimism that we will soon succeed. Loss of vocabulary and questions leaves theorists with a sense of vertigo. Critiques of current practices have then been dismissed as anti-emotion theories. Zachar (2006) characterised my alternative as Eliminativist. Colombetti (in press) suggested that my alternative implies that "emotions have no existence independent of our categorization" (p. 2). For many readers it might seem as if I am proposing that we toss out everything of importance. So, what remains?

In abandoning the geocentric theory of the solar system, some facts survived the move to a heliocentric system: the earth and sun remain, their relative positions remain, motion remains, observed time between events remain. Similarly, in abandoning an emotion-centric view, many of our observations about events called emotion remain. Indeed, all the observable phenomena remain in my account: faces move, prosody changes, hearts speed up and slow down, people aggress and they flee–all such events really occur, independent of our categorisation. Some people see anger in faces of others and they feel anger in themselves[1] – these events really occur, but do imply categorisation via the folk concept of anger. Emotion's manifestations (although on my account the phrase is a misnomer) are real and im-

---

[1] To be clear: when I write of "seeing" anger or another emotion, of "seeing" two lines in the Müller-Lyer figure as of unequal length, or of "seeing" the sun set, I mean to describe what the perception is like for the perceiver. In the realm of the emotions, we see faces move, which they really do, hear the voice change, which it really does, see palms sweat, which they really do, and so on, all in some context. We then name the combination of these events as anger, thereby inferring anger as their cause

portant and cannot be eliminated. A scientific account of these manifestations can be had without the concepts of emotion, anger, etc., used in a scientific role and without assuming them to be manifestations of something. Even the concepts of emotion, fear, anger, and the rest, remain–although as everyday folk concepts rather than as scientific ones. In that role, they are important topics, much as are all other folk concepts from *angel* to the *zodiac*. *Anger*, *disease*, *emotion*, *marriage*, *meal*, *money*, *poker*, *tax*, *tornado*, *soap* and many other folk concepts play a role in human affairs that must be understood.

On my account, events now called *emotion* (in human adults and infants and non-human animals) bear a family resemblance to one another. Although it does not follow that emotion/not-emotion divides nature at its joints, the word emotion does point in an interesting direction, and the events it refers to are real and important. The same can be said for fear, anger, and the like.

In short, on my account, much remains but is understood in a different framework. My account does not allow the concept of emotion to determine boundaries of the domain to be explained. My account postulates no single causal entity–the emotion of anger, fear, or their neural equivalent, an affect programme–to explain the occurrence of the manifest components of that emotion or to explain the pattern or correlations among these components. In the remainder of the article, I summarise my account by clarifying two key concepts: Core affect and psychological construction. Core affect is part of, but not the whole of, what are called moods and emotions. Psychological construction is an umbrella term for the set of processes that produce a token emotional episode's manifestations and its categorisation as an emotion[2].

---

[2] The influence of the concept of emotion is pervasive and subtle. For example, critics of my account (Russell, 2003) have mistakenly assumed that mine is a "dimensional" rather than "categorical" account of emotion, i.e., that the domain of emotion is to be subdivided by dimensions of core affect rather than by discrete categories such as fear and anger. Another false interpretation is that core affect carries the same assumptions that emotion does, i.e., that core affect must account for facial and vocal expressions, autonomic changes, emotional behaviour and the like.

2. Core affect

Core affect is a pre-conceptual primitive process, a neurophysiological state, accessible to consciousness as a simple non-reflective feeling: feeling good or bad, feeling lethargic or energised. There is something it is like to feel core affect. Its presence in consciousness varies from being focal to peripheral to out of sight. A structural description of core affect is an empirical matter, and I have proposed a circumplex with the two underlying dimensions of pleasure–displeasure and activation–deactivation (Russell, 1980, 2005; Yik, Russell, and Steiger, 2011). Although two-dimensional, core affect is, subjectively, a single feeling. That is, the two dimensions combine in an integral fashion to form one unified feeling. Pleasure and arousal combine to form the single feeling of ecstasy, for example. An analogy is the sensation of a specific colour, such as the red of the autumn leaf outside my window. The dimensions of hue, saturation, and brightness combine in an integral fashion to form one unified sensation of any particular colour.

It may help to say what core affect is not. So, here is a series of conceptual distinctions. The idea is to point to ostensibly different phenomena, without precluding the possibility that core affect is contingently related to them.

Core affect, although it may be empirically involved, is not a judgement that something is morally good or bad or tactically advantageous or harmful. Core affect has been shown empirically to be related to changes in the autonomic nervous system, facial and vocal behaviour, instrumental behaviour, cognitive processes, reflexes, and a host of other things (Russell, 2003). Still, core affect is conceptually distinct from them.

Core affect is not a substitute term for emotion, nor is it the essence of emotion; it is not a necessary feature of emotion[3]. Prototypi-

---

[3] Core affect does provide the pleasant or unpleasant hedonic tone to those token emotional events that have that tone. Core affect is also a central feature of the mental prototypes of some emotions but not a necessary feature of all cases of those emotions. For example, cases of fear can be found without a core affect of unpleasant arousal,

cal emotional episodes are said to begin and then, after a short time, end. In contrast, one is always in some state of core affect, which simply varies over time, sometimes slowly, sometimes rapidly, without beginning or end. Prototypical emotional episodes are directed at something (one is angry with, afraid of, or sad about something). In contrast, core affect is not necessarily directed at anything. (In this regard, core affect is closer to the English word *mood* than to *emotion*, although mood prototypically is long lasting and mild.) Core affect per se can be free floating (as in feeling down but not knowing why), but it can come to be directed at something. The full experience of core affect can thus become intentional in the philosophical sense, in much the same way that the full experience of a pain can become intentional (Searle, 1992, p. 251).

Core affect is also distinguishable from the affective properties we perceive in objects, events, and features. We perceive objects to be beautiful, ugly, awful, soothing, and so on. We perceive how pleasant or unpleasant or how energising or soporific something promises to be. Typically, perceiving something to be beautiful is a pleasant experience, perceiving something ugly unpleasant. And, such perceptions are logically related to core affect: to perceive something as pleasant is to judge it capable of producing pleasure. Nevertheless, I want to allow the possibility that the perception of an affective quality can occur even when we're not personally moved by the object or event. That is to say, such perceptions are not, by definition, accompanied by changes in core affect. Distinguishing the two thus leaves as an empirical hypothesis the idea that core affect is involved in the perception of affective quality.

---

but they are mediocre or borderline rather than prototypical cases. (Fear without arousal: contemplating a distant danger; fear without displeasure: thrill seeking.) Charland (2005) underscored the distinction between core affect and emotion by observing that, on Panksepp's theory, some emotions (lust, seeking) can, depending on context, involve very different values of core affect prototypically. Of course, the concepts of emotion, fear, and so on, could be redefined for scientific purposes such that core affect is a necessary feature, but I'm not advocating that route.

Even when non-intentional, even when we have no clue why we feel as we do, core affect is caused. Individuals vary greatly in how their core affect is experienced (Barrett, 2009) and changes over time (Kuppens, Van Mechelen, Nezlek, Dossche, and Timmermans, 2007). Core affect changes in response to many simultaneous influences. Sometimes the influence is a single powerful and obvious external event, as in James's prototypical emotional episode of the bear in the woods. More typically, however, there are many simultaneous influences, including some beyond human ability to detect. Some influences are internal, such as diurnal cycles, hormonal changes, and immune responses. Core affect is changed directly by chemicals, both uppers and downers and euphoric and dysphoric drugs. Indeed, it is precisely the core-affect-altering properties of drugs that make them objects of abuse or avoidance. Addicts seek out drugs that make them feel good, and some people fail to take needed medicines that, as a side effect, make them feel bad. Core affect is altered by real events, but also by imaginary, remembered, and foretold events; core affect responds to virtual reality in art, imagination, fantasy, and entertainment.

There are too many such influences for a person to track them all and hence to know what caused his or her current core affect. So, although we often have a good idea of why we feel the core affect we do, we sometimes don't, as in free-floating emotions and moods and everyday feelings. For this reason, attributions and misattributions often play a role in emotional episodes (Neumann, 2000; Weiner, 1985).

Core affect appears to have many of the features of modularity (Faucher and Tappolet, 2006): fast, mandatory, unique output, an evolutionary explanation, dedicated brain circuitry, and encapsulation (Russell, 2006). At first blush, encapsulation (an example of which occurs in the Müller-Lyer illusion in which one simultaneously knows that two lines are of equal length and yet sees them as unequal) seems an unlikely property of core affect. After all, cognitively processed information can be a powerful influence on core affect, obviously: core affect responds to news of winning a lottery, learning of the death of a loved one, reading about injustice, and pondering a distant danger such as global warming–all of which require cognition. Such examples, however, also have another feature in common: conscious attention to the information. When the information fades from consciousness, its influence on core affect similarly fades. Consider the excitement of winning the lottery. If the news continues to invade consciousness (shopping, congratulatory remarks, memories), then it continues to shape core affect. But if one is distracted such that thoughts about the win are pushed out of consciousness, then the excitement fades. Thus, I offer the hypothesis that core affect is largely encapsulated from cognitively processed information, with the major exception of knowledge seen through the window of consciousness. Thus encapsulation here means that knowledge per se does not alter core affect unless we are conscious of it. Encapsulation thus resonates with the response of core affect to virtual reality: in both cases, core affect responds to what is in consciousness, rather than to what is known to be real.

3. Psychological construction

The science of emotion has historically sought a single process that explains (or defines) emotion and that distinguishes emotions from non-emotions. Some of the proposals are that emotion requires (or is): somatosensory feedback from the bodily reaction to an emotional stimulus; the output of a small set of pre-wired reflex-like neural modules (affect programmes); the enactment of a core relational theme; cognition (labelling, attribution) triggered by bodily arousal; the valenced response to a set of appraisal processes; an act of categorisation; and the enactment of a socially constructed role. The issue I want to raise is not the precise nature of these theories–I have undoubtedly oversimplified them–or the proper verb (*requires* vs. *is*). The issue is the assumption that the number of processes needed to explain or define emotion is one.

The concept of emotion lacks necessary and sufficient features. Given the heterogeneous nature of the extension of emotion, I believe that each of these theories is likely true for some cases, but that no one theory will be true for all cases called emotion without being true for pretty much all nonemotional psychological states as well. My proposal, psychological construction, is thus not the claim that emotion requires (or is) core affect, an act of categorisation, a socially constructed role, or an attribution of core affect to an event. Indeed, what gets psychologically constructed is not emotion as a generic process or anger, fear, etc., as generic kinds; rather, what gets psychologically constructed are individual token events, which may (or may not) then be classified as emotion, fear, anger and the like by means of a folk concept.

Each actual instance called an emotion is real and requires an explanation. Psychological construction is not just another such single mechanism offered as an explanation for all and only cases of emotion. Psychological construction is based on assumption that no one specific mechanism explains all, or even a good number, of cases called emotion. Each token's components are cobbled together on the fly. Actual instances called emotion are constructed in a variety of ways, depending on the particular circumstances of the particular individual.

The phrase *psychological construction* is meant to contrast both with biological construction, assumed in basic emotion theory, and with social construction, often assumed by anthropologists and sociologists. Both biological and social construction seek the origin of emotion as one generic process and seek the origin of fear, anger, and so on, each as one general mechanism. Both thus assume that all emotional episodes are largely predetermined, by nature in the former and by culture in the latter. On my account, both nature and nurture are essential, but the important point is that each particular instance of emotion is not predetermined but comes together at a psychological level at the time of its occurrence. There is no one general mechanism. Put differently, the traditional assumption is that a "theory of emotion" will differ from a theory of cognition, or behaviour, or conation. My claim, in contrast, is that any theory that explains all cases called emotion will be close to the whole of psychology, a theory that of course will not be limited to emotion but will extend to all psychological processes.

My proposal raises three questions: (1) how to explain any given emotional episode, each token event; (2) how to explain the patterning of emotion's components; and (3) how to explain our compelling perceptions that particular episodes are instances of a general category, such as fear or anger.

3.1. Explaining emotion's manifest components

So, how specifically does psychological construction explain any given event called emotion? The first step is to recognise that each token consists of manifest components (Scherer, 2001, 2009), such as facial and vocal expression, changes in the autonomic nervous system, subjective experience, and so on. Let us begin by examining each component, with an eye toward how we might begin to explain it without positing an emotion as its cause. Doing so also explains why there are individual and situational differences that occur even within a given category of emotion.

Facial and vocal expressions. For the face or voice to "express" emotion implies, among other things, that the emotion causes the facial or vocal changes. Explanations of facial and vocal changes in terms of emotion have dominated the research agenda, but alternative explanations are starting to appear. Faces move as part of perception: to look away, to stare, to smell, etc.: We wrinkle our noses at smells, gag at noxious tastes. Faces move as part of cognitive processes such as attention and appraisal (Ortony and Turner, 1990; Scherer and Ellgring, 2007; Smith and Scott, 1997). Faces move as part of action, including speech and other goal-directed actions:We turn our faces in the direction we are heading, and we make faces as we tell stories. Some facial signals evolved (either genetically or epigenetically) to guide social interactions, such as threat, greeting, or submission (Fridlund, 1994). The face and voice change with general arousal and positive versus negative core affect (Lang, Greenwald, Bradley, and Hamm, 1993). My hypothesis is that facial and vocal behaviour during an emotional episode can be explained through such non-emotional processes (Russell, Bachorowski, and Fernandez Dols, 2003).

Autonomic nervous system (ANS). The search for an ANS signature for each emotion is based on the commonsense assumption that our heart races because of fear, and so on. Such signatures have not been found (Larsen, Bernstson, Poehlmann, Ito, and Cacioppo, 2008), and, more generally, ANS changes can be accounted for without attributing them to an emotion. ANS activity is on going, with both general features (degrees of arousal) and specific chores. Perhaps one can also detect broad response patterns, mobilisation for action versus relaxation for homeostasis or perhaps preparation for approach versus avoidance (Larsen et al., 2008). Long ago, Lacey (1950, 1967) found individual- and context-specific patterns of autonomic response.

Emotional behaviour. In the traditional way of thinking, emotions cause behaviour: we strike because are angry, we flee because we are frightened, and so on. James (1884) turned the causal arrow around: we are angry (or, more precisely, perceive ourselves to be angry) because we strike and are frightened because we flee. Baumeister, Vohs, DeWall, and Zhang (2007) reviewed evidence that questioned whether there is any causal arrow between anger and aggression, or between fear and flight, or, more generally, between emotion and behaviour. Scherer (2009) cites as a feature of emotion the looseness of its link to behaviour.

Fight, flight, and other emotional behaviour can be accounted for without emotion. Core affect, even when outside consciousness, influences behaviour (Winkielman, Berridge, and Wilbarger, 2005). Appraisal of the current situation can lead to formation of goals and plans and their execution (a process that can be implicit as well as explicit, and quick and ill advised as well as slow and wise). Strack and Deutsch (2004) brought together various lines of evidence and theorising to propose a two-system account of behaviour. An "impulsive system" influences behaviour through associative and motivational mechanisms, whereas a "reflective system" influences behaviour through a knowledge-based mechanism. On their account, emotional behaviour is not thought of as resulting from an emotion, but as the joint product of these two interacting systems.

Experience of having an emotion. Persons sometimes feel afraid, angry, sad, and so on. Such conscious feelings are real, but we need not then go on to assume that such feelings are veridical detections of (caused by) an emotion. On my alternative account, emotion is not their cause. Rather, as James (1884) proposed, such experiences are perceptions: to feel afraid is to perceive oneself as afraid (see Prinz, 2004). More technically, I characterised this feeling as a meta-experience because the raw data on which it relies include other experiences: core affect, somatosensory feedback, appraisal of the eliciting event, attribution, beliefs, desires, plans, and behaviour. Similarly, Lambie and Marcel (2002) characterised the experience of having an emotion as a second-order experience–that is, an experience that emerges out of first-order experiences.

Perceiving oneself as having an emotion is no different in kind from other perceptions. Percepts are often compelling, but they are not simple. Nor are they infallible. Like other acts of perception, an emotional experience is not entirely "bottom-up", not entirely data driven. A percept is the end product of a complex process involving raw data, concepts, learning, and context. Smith and Neumann (2005) developed a sophisticated account along these lines by drawing on dual-process models of perception.

To perceive oneself as afraid is to categorise oneself by means of the concept of fear. It is to establish the meaning of one's state via the concept of fear. In turn, the concept of fear can be unpacked as a script laying out a series of subevents (the components) in a temporal and causal order (Fehr and Russell, 1984). To perceive oneself as afraid is to see a resemblance between one's current state and a mental script (the concept) for that emotion.

A separate hypothesis is that there are both similarities and differences in emotion concepts across cultures and languages (Russell, 1991; Wierzbicka, 1999). Although some of the raw data on which emotional experience depends are universal (core affect, somatosensory feedback, attributions, appraisals, etc.), where there are differences in emotion concepts, the corresponding emotional experiences would then vary as well. Evidence from a surprising source supports this prediction. In their effort to find a universal ANS signature for each basic emotion, Levenson, Ekman, Heider, and Friesen (1992) studied the Minankabau of West Sumatra. Participants were instructed to contract facial muscles into the prototypical configurations hypothesised for basic emotions. Doing so, in turn, alters ANS activity. For North Americans, this alteration of facial muscles and ANS activity resulted in reports of the experience of specific emotions. For the Minankabau, however, the same procedure failed to produce the emotional experience, presumably because their emotion concepts differed from those of North Americans.

Appraisal. On more traditional accounts, emotion is separated from thought and reason, as in theories in which emotion is encapsulated from cognition (Griffiths, 1997; Prinz, 2004). In other accounts, an appraisal is thought of as a part of the emotion or as an event that mediates between an antecedent event and the emotion. The dichotomy between rational thought and irrational emotion was long ago undermined by research showing, on the one hand, that cognitive processes emphasise economy and speed as much as rationality and, on the other, that cognitive processes are relevant to emotion.

Interestingly, appraisal was initially thought of as a simple evaluation (Arnold, 1960), but then more and more dimensions were added (Ellsworth and Scherer, 2003). Scherer's (2009) latest theory draws on dual-process accounts to characterise appraisal as a complex, dynamic, multi-component process. To me, this suggests appraisal is involved not only in emotion but in the continuous monitoring of the environment by any sentient creature. We are constantly appraising the world around us, including perceiving the affective qualities of events, objects, and features. Doing so entails our full sensory-perceptual-cognitive apparatus, including rational thought. I therefore anticipate the addition of ever more dimensions as relevant. Rather than listing those cognitive processes involved in emotion, we might ask which sensory-perceptual-cognitive process is not involved. It might be the shorter list. For example, memoric processes are rarely mentioned, but emotional episodes can result from a situation via its association with a remembered event (Clore and Ortony, 2008; LeDoux, 1996). In the end, to account for the appraisal component of emotion may require nothing less than the full psychology of sensation, perception and cognition.

Appraisal has also been thought of as the principal source of individual differences in emotion: the reason the same antecedent can produce different emotions in different individuals. This enterprise will help underscore the great variety of individual differences within each category of emotion.

Attribution. Attribution, too, has typically been thought of as mediating between an antecedent and the emotion. A very common experience is feeling happy or unhappy about something. For example, one feels happy to hear from a friend, unhappy about the unexpected rain, and so on. My hypothesis is that these experiences require an attribution of core affect to the something, hearing from the friend, the unexpected rain, and so on. Attribution plays a greater role in emotional episodes than is sometimes appreciated. For example, Neumann (2000) placed participants in an ambiguous but upsetting situation. He used cognitive priming to influence attribution. Attribution, in turn, influenced subjects' self-reported emotion: attributing their upset to another person resulted in their reporting anger, whereas attributing their upset to themselves resulted in their reporting guilt. This study highlights subtle situational variation and individual attributional styles as a source of the variety of emotional episodes.

Emotion regulation. Emotion regulation is implicitly thought of as a consequence of the emotion, but I suggest that it be thought of as another, on-going process, much like I'm thinking of the other components of the emotion. I also suggest distinguishing two different processes. One is a fundamental motive in life and concerns core affect. We often (although not always; Tamir, 2009) behave so as to maximise pleasure and minimise displeasure–or at least take pleasure and displeasure into account. We also often behave so as to adjust our arousal level, sometimes raising it and sometimes lowering it. I refer to these two ubiquitous processes as *core affect regulation*. An account of this process will be had from a general account of behaviour, motives, goals, plans, and action. In a separate process, we seek to cultivate, avoid, or leave certain specific emotional states as defined within our given culture. Perceiving that I am afraid–or at least that others see me that way–I may put on a brave face. An account of this process requires an account of self-perception of emotion and of the role of folk concepts and cultural roles.

3.2. Patterns among components

In the last section, I divided each token emotional episode into its manifest components, claiming that each component can be explained without positing an emotion as its cause. Indeed, all components (with the arguable exception of the experience of having the emotion) occur outside as well as inside emotional episodes and in such cases need to be explained without recourse to emotion. My list of components is not likely to be exhaustive, nor each explanation the final word, but my research agenda should be clear enough. One could, for the sake of argument, concede my point about the explanation of individual components, but still suppose that theorists need emotion concepts to explain the co-occurrence of the components within the emotional episode. That is, the concept of emotion not only specifies individual components but carries the assumption that various components recur in a highly coherent organised pattern. And of course, in science, a recurring pattern would require an explanation. So, how about patterns?

Patterns are loose. First, our traditional assumption about patterning must confront the accumulating empirical evidence. Correlations among the components appear on the evidence to be weak. Some of the entries in Table 1 point to empirical evidence that an anticipated association between two components is unexpectedly weak or nonexistent. Tradition anticipates that in cases of fear, one finds all or most of fear's components such as flight, the conscious experience of fear, a frightened face, an ANS signature unique to fear, and so on. This expectation has not stood up to empirical scrutiny (Lang, 1968, 1979, 1988; Lazarus, 1991; Mandler, Mandler, Kremen, and Sholiton, 1961; Rachman, 1984; Rachman and Hodgson, 1974). People who experience fear show a variety of ANS changes, engage in various behaviours, and so on. More recently, studies of the correlations among components of surprise and among those for disgust have been reported, with similar results (Reisenzein, 2000, 2007). There is much less patterning to be accounted for than is traditionally assumed.

Many of the components are continuous psychological processes: core affect is always present; the ANS is continuously active; people are always behaving, always perceiving, always appraising, and so on. Because emotion's components are on-going processes, they necessarily occur in combination. The weakness of the correlations among components implies a large number of possible combinations. The question has not to my knowledge been raised which of these combinations constitute organised patterns and which mere combinations. For now, I won't pursue this distinction and will call all combinations patterns.

Explaining the patterns that do exist. Correlations found so far among components are weak, but not zero. Besides, patterns could exist that have not yet been established. Correlations that are found among components thus require explanation. Traditionally, theorists have sought a single central mechanism to account for what correlations exist–a fear-producing mechanism such as an affect programme, for example, to account for what associations occur among fear's components. A single central mechanism, however, is not warranted until simpler explanations are exhausted.

Some writers have recently acknowledged the growing evidence on the weakness of the correlations among emotion's manifest components and have begun exploring dynamic systems theory as a conceptual framework that accommodates the tremendous variation across situations and individuals within a given category of emotion (Camras, 2000; Colombetti, 2009; Fogel et al., 1992; Lewis, 2000; Scherer, 2000). I believe a much simpler approach is possible, one that relies on already known sources of correlation. There are at least three, not mutually exclusive, alternatives.

First, features in the environment have a correlational structure. When two correlated environmental features each elicit a separate response, then those two responses will be correlated. For example, suppose that novel events are more likely than familiar events to block a goal. (Perhaps problems presented by familiar events have been previously encountered and solutions found.) Suppose further that goal blockage elicits an ANS pattern of cardiac acceleration and that novel events elicit frowns. If so, cardiac acceleration and frowning will be correlated even if no internal process links the two.

Second, one component process can influence another. For example, suppose that forming the face into a threat expression (the "anger face") alters breathing and muscle tension, which in turn alters ANS activity, perhaps cardiac acceleration. The consequence would be that the threat face is correlated with cardiac acceleration. The component of emotional experience uses the other components as raw data in the formation of the percept. This hypothesis predicts that emotional experience is correlated with other components–more highly than they are with each other. Appraisal of current situation and attribution lead to emotional behaviour.

And, third, two components will be correlated when they are both influenced by a central mechanism other than emotion. For example, suppose that focused attention produces both muscle tension in the face and cardiac acceleration. If so, muscle tension in the face will be correlated with cardiac acceleration. The central mechanisms responsible for emotional behaviour (on Strack and Deutch's account, impulsive and reflective systems) influence other components such as facial and vocal behaviour and ANS activity.

In short, if my thesis pans out, the explanation for the pattern of manifest components that occurs in each instance of emotion is the explanation for the individual components plus well-known and simple explanations for the correlation among those components. No additional emotion-producing mechanism is needed. As an analogy, consider a fair game of poker. When cards are dealt fairly, each hand consists of a pattern of cards (a pair or straight, or whatever, although most patterns have no name within the rules of poker). Shuffling and dealing are the mechanisms that produce patterns; there is no additional pair-producing or straight-producing mechanism (other than cheating). Of course, in poker and in emotion, we see certain patterns, and doing so can be very important. Which brings us to the next topic.

3.3. Perception of emotions

As outlined so far, psychological construction is a research agenda that hopes to explain each particular emotional episode by explaining its manifest components and the correlations among the components. To many readers, this programme, even if it were successful, would still leave something out. The available empirical evidence on correlations among components would not lead scientists to hypothesise fear or other types of emotion as their explanation. But then most of us do not hypothesise fear–we see it. We see discrete emotions in others and experience them in ourselves. We remember prototypical blue-ribbon cases of anger, fear, etc.–each with all the components in the right pattern. These compelling observations, experiences and memories of emotions do not entail the traditional view of emotions, but they do raise a question: How to account for these perceptions and memories? My account has two parts: the event perceived and the perceiver.

Consider the event perceived. Someone is going about his or her daily life, encountering various situations and pursuing various goals. He or she will necessarily be undergoing some of the components listed above, since they are continuous on-going processes. Some component processes are therefore always present and always in some pattern. As plans unfold and situations occur, the components change and hence form new patterns. Therefore, even if the component processes were completely independent of one another, certain nameable patterns would form from time to time. That the components are somewhat correlated with each other means that certain patterns are more likely to form than others.

Now consider the perceiver observing this person. The perceiver does not simply register an external reality. Rather, the perceiver brings to the task a set of concepts embedded in a set of implicit assumptions and inherited from our linguistic ancestors–concepts for English speakers such as *emotion*, *fear*, *anger*, and so on. A concept such as *fear* is a mental script that specifies a temporal and causal pattern among various components. The perceiver observes the other's face, voice, behaviour, signs of physiological state, current situation, and so on. Perhaps the perceived person reports a subjective experience. The perceiver notes the similarity between the pattern of observed components and the script defining a category such as fear or anger. Occasionally, there is sufficient resemblance for that pattern to count as a member of the emotion category.

The key here is that membership in the emotion category does not require a set of necessary and sufficient features. Resemblance is a matter of degree. Occasionally, perceiving a single feature of the script will be enough for the instance to count as a member. Different observers may arrive at different categorisations of the same event. Although the script specifies a temporal and causal pattern, these aspects of the concept are not necessary either. Thus, uncoordinated combinations of components may still count as an instance of the category. The border between fear and not-fear is fuzzy.

The upshot is that members of the fear category resemble one another along different dimensions, and the set of all cases of fear is much more heterogeneous than most of our current scientific theories assume. This hypothesis is consistent with the general trend in the study of semantic categories. Traditionally, it was assumed that all the instances of a category must possess the same features, in this case the recurring pattern of defining components. Classically, the features were assumed individually necessary and collectively sufficient to determine membership in the category. Wittgenstein pointed out in his analysis of the concept of game that these assumptions are not necessarily so. We neither have nor need classical definitions. Analyses of the concepts of emotion, anger, and the like have similarly challenged the classical assumption (Fehr and Russell, 1984; Russell and Fehr, 1994). Further, memory is biased toward the more prototypical exemplars, which are first to spring to mind at mention of the category name.

Combining the nature of our mental categories for emotion with the ongoing nature of the component processes explains how we see emotions from time to time. In a person's life, at each instant, the processes called components are occurring. As the processes change, they necessarily form patterns. Many such patterns will seem random and go nameless, but sometimes the pattern will resemble the mental script for a specific emotion. Resemblance is a matter of degree, and so while some cases will be excellent examples of the concept, many will be mediocre examples, and some will be borderline such that one is not sure if it belongs inside or outside the category. The same event can resemble more than one script, albeit typically to different degrees. When resemblance is sufficient, we see the emotion in another or experience it in ourselves. We better remember the excellent examples, and excellent examples are more available when thinking about the category.

Even so, one might argue, events we categorise as, say, fear do in fact have some of the predicted components, and the blue-ribbon cases do in fact have all or most of the components and in the right order. All the instances named *fear* by an observer will indeed have some subset of the components specified in the mental script for fear. This fact follows trivially from the fact that all the cases were *selected* on the basis of their resemblance to the script. All the instances named as blue-ribbon examples of fear will have all or most of the components in the right order. This fact follows trivially from the fact that all the blue-ribbon cases were *selected* on the basis of their very close resemblance to the script.

What humans perceive, categorise, conceptualise and remember tells us less about the nature of reality than was traditionally supposed. As an analogy, consider the constellations we see in the sky. When we perceive the Big Dipper, we perceive real features: the stars comprising the Big Dipper are real and the pattern among the stars is real in the sense that the stars are really positioned in the universe such that they form a certain geometric pattern when viewed from earth. But stellar constellations, contrary to the beliefs held in many traditional cultures, are not interesting scientific entities. The Big Dipper does not explain the presence of the stars or the pattern among those stars. Constellations are not part of the causal story of astronomy. There is no use in asking for an astronomical account of why the Big Dipper exists or why it is structured the way it is. Astronomy long ago abandoned questions such as: What is a constellation? How many constellations are there? What are they? How were they generated? How does one constellation differ from another? What are their effects/functions?

4. Defence of the current proposal

Colombetti (in press) defended the existence of discrete emotion kinds as coherent and recurring packages (biological entities) by critiquing the account I have offered. She focused on the fact that humans perceive emotions in others. She argued that perception of an emotion entails the existence of components in the event so perceived. She wrote, "recognizing emotions is a matter of matching an acquired script with the features of a perceived event... but we are able to [do so] because *there are* features in the episode that we identify as matching the script" (p. 20). She further argued that recognising an emotion also implies the existence of a pattern: my account does "not explain how categories (and/or mental scripts) can ever come to be applied to the various components that eventually make up an emotion episode. The problem, as I see it, is that it is impossible to provide such an explanation if one does not also posit that there are already coherent packages, or patterns, in the organism. It must be because there are already such packages that we are able to recognize and identify them by matching them with available categories and/or mental scripts" (p. 8).

Colombetti is surely right on two points. For any given token event, when an observer perceives another as, say, afraid, then normally (1) the other is showing one or more individual components of fear and (2) the components occur in a pattern. Nevertheless, it does not follow that the same component or the same pattern recurs in other cases called fear. What would be required to establish Colombetti's claim is not just the existence of components in a pattern *in a given instance* but that the same components in the same pattern *recur in all or most instances of that emotion*. The assumption of a recurring pattern is widespread. It has been the assumption motivating a century of research seeking the features common to the instances of each kind of emotion.

So, it has been a great surprise when evidence has failed to support this assumption. For example, psychophysiologists assumed that all or most instances of fear would show a similar ANS pattern, but no such recurring pattern has been found (Larsen et al., 2008). Many assumed that people reporting fear would show the same facial expression, but little evidence has been presented in support of this assumption, and some evidence suggests the opposite (Carroll and Russell, 1997; Fernandez-Dols and Ruiz-Belda, 1997; Fernandez-Dols, Sanchez, Carrera, and Ruiz-Belda, 1997). Similarly, many assumed that all or most instances of fear would have a behaviour in common. Yet, no such common behaviour has been found. Consider the spider phobic who jumps back from a spider, a mother afraid that her child is sick and rushes him to a hospital, the person sitting in a theatre watching a thriller, someone riding a roller coaster, and the person afraid of global warming (Russell, 2003).

One could of course argue that all the behaviours that occur in dangerous situations are "avoidance behaviours" or suppose the existence of an action tendency to avoid. Doing so, however, encounters three problems: First, there are counterexamples (thrill-seeking and bravery). Second, labelling a variety of different behaviours as "avoidance behaviours" adds nothing to their explanation, for the particular behaviour that actually occurs (jumping back from the spider, rushing to the hospital, etc.) remains to be explained. And, third, one cannot classify an isolated behavioural act as avoidance except in the context of an interpretation of the situation as dangerous. For example, to interpret running as fleeing rather than as approaching something else or as running for exercise requires knowledge of the context. Thus, the association between danger and a behaviour being avoidance borders on circularity.

Colombetti (personal communication, 28 August 2008) replied that, "These examples all support the view that there are common behaviors! Not across these examples, but within . . . there seem to be recurrent behaviors in specific contexts". Commonsense supports Colombetti's assumption, and many would assume she's correct. But, her claim is an empirical one, and I predict that it will not be supported. Although not much evidence is available, what is available is not encouraging for her claim. The first context I mentioned, fear in phobics, has been studied. Indeed, this was the context that initially exposed the lack of consistent behaviour in fear (Lang, 1968, 1979, 1988; Lang et al., 1993; Rachman and Hodgson, 1974) and led Rachman (1984) to declare that "fear is not a lump". In experimental tests, individuals are all subjected to an identical situation, and yet individual differences prevail. I predict that within each of the contexts that I listed–being afraid while having a sick child, watching a film, riding a roller coaster, and contemplating global warming–a variety of behaviours will be found.

The defensive behaviour of rodents in dangerous situations illustrates my view. Rats show a range of behaviours including active investigation of the threat stimulus, alarm vocalisations, escape, freezing, attack, avoidance, information seeking and risk assessment. Mother rats move pups to safer ground. As Blanchard, Hebert, and Blanchard (2007) summarised, "Defensive behaviours are modulated by features of both the threat stimulus and the situation in which it is presented" (p. 653). Perhaps a cage environment can be so constrained that all rats show the same behaviour, but maybe not. For example, two rats in a small empty cage typically leap into the air when the floor delivers mild electric shock, but occasionally they attack one another.

Finally, even if Colombetti's claim were correct (behaviour is consistent within although not between situations), it would not justify postulating fear as an entity. In that case, the behaviour would be explained by the situation and not by fear. For the hypothesis of fear to have a scientific use requires evidence of a similar pattern across the various contexts in which fear is said to occur.

5. Future directions

I have laid out an ambitious research agenda, which, needless to say, is subject to revision as data are gathered. I end with an even broader agenda. Core affect has been criticised because it does not capture all there is in emotion. My phrase “psychological construction” seems naturally to imply the psychological construction *of emotion*. I have been asked what evidence I can offer that core affect plays a central role *in emotion*, more central, for example, than appraisal or somatosensory feedback. Implicitly, the domain for our field is limited to emotion and the questions asked implicitly assume that emotions are what are to be explained. It is just such preconceptions that must be challenged. Admittedly, I presented psychological construction as a way of accounting for emotion. But doing so is meant to be a bridge between current concerns and a new approach. Psychological construction is an umbrella term for a host of mechanisms that explain not only those events called emotion, but a much larger set of events. Core affect is important in psychology; its role in those events called emotion is an empirical question and range from central to nonexistent in individual cases. Core affect extends beyond the domain of emotion.

When the general concept of emotion is treated as the folk concept that it is, without authority to determine scientific boundaries, then more real and important events become more visible and present themselves to be explained. Once the concept of emotion is relegated to a chapter title, with no real scientific work to do, then the domain of our inquiry is revealed to be much wider. Some of the events are what are now called emotions, some moods, some feelings, and many have no name. Blue-ribbon emotional episodes (those that fit scripts for specific emotions closely) are included, but have no special status. Borders between blue-ribbon cases, mediocre cases, borderline cases, and non-cases are fuzzy and serve no scientific purpose. Instead, let us be open to a wider range of events. Feeling good that the sun is shining or feeling bad that the weather is turning too warm may not qualify as emotions, but they are frequent events, influence other behaviours, and require explanation. More generally, even without emotion, anger, and the rest as scientific concepts, we are left with each of the components discussed here. The task is to explore each component and the relations among them.

The reader may find the analysis outlined here pessimistic. I find it optimistic. I believe that progress in the science of emotion has been blocked by unwarranted but hidden assumptions. On my suggested strategy, we will have no grand theory of emotion, but we will have mini-theories of individual components. Research in practice already does just this. Freed from these roadblocks, progress may be rapid. Indeed, much has already been achieved, but not recognised because of these hidden assumptions. For example, consider the various theories of emotion proposed over the last century or so. Each theory has been found wanting, because it has been asked to perform an impossible task: account for all cases of emotion. Freed from this goal, these theories complement one another and point to important processes. Each may solve part of the puzzle. Cumulatively, they may together provide a viable account. The analysis of this article is aimed encouraging the development of separate accounts of each "component" of emotion and the empirical search for links and patterns among those components.

References

Arnold, M.B. (1960). Emotion and Personality (Vol. 1). (New York: Columbia University Press).

Barrett, L.F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. Cognition and Emotion *23*, 1284–1306.

Baumeister, R.F., Vohs, K.D., DeWall, C.N., and Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. Personality and Social Psychology Review *11*, 167–203.

Blanchard, D.C., Hebert, M., and Blanchard, R.J. (2007). Defensive behaviors. In: Encyclopedia of Stress (Vol. 1, 2nd ed.), G. Fink, ed. (Oxford, UK: Academic Press), pp. 722–726.

Camras, L.A. (2000). Surprise! Facial expressions can be coordinative motor structures. In: Emotion, Development and Self-Organisation: Dynamic Systems Approaches to Emotional Development, M.D. Lewis and I. Granic, eds. (Cambridge, UK: Cambridge University Press), pp. 100–124 .

Carroll, J.M., and Russell, J.A. (1997). Facial expressions in Hollywood's portrayal of emotion. Journal of Personality and Social Psychology, *72*, 164–176.

Charland, L.C. (2005). Emotion experience and the indeterminacy of valence. In: Emotion and Human Consciousness, L.F. Barrett, P.M. Niedenthal, and P. Winkielman, eds. (New York: Guilford), pp. 231–254.

Clore, G.L., and Ortony, A. (2008). Appraisal theories: How cognition shapes affect into emotion. In: Handbook of Emotions (3rd ed.,), M. Lewis, J.M. Haviland-Jones, and L.F. Barrett, eds. (New York: Guilford Press), pp. 628–642.

Colombetti, G. (2009). From affect programs to dynamical discrete emotions. Philosophical Psychology *22*, 407–425.

Ekman, P. (1984). Expression and the nature of emotion. In: Approaches to Emotion, K. Scherer and P. Ekman, eds. (New York: Erlbaum), pp. 319–343.

Ellsworth, P.C., and Scherer, K.R. (2003). Appraisal processes in emotion. In: Handbook of Affective Sciences, R.J. Davidson, K.R. Scherer, and H.H. Goldsmith, eds. (New York: Oxford University Press), pp. 572–595.

Faucher, L., and Tappolet, C. (2006). The modularity of emotions. Canadian Journal of Philosophy, Supplement Volume *32*.

Fehr, B., and Russell, J.A. (1984). Concept of emotion viewed from a prototype perspective. Journal of Experimental Psychology: General *113*, 464–486.

Fernandez-Dols, J.M., and Ruiz-Belda, M.A. (1997). Spontaneous facial behavior during intense emotional episodes: Artistic truth and optical truth. In: The Psychology of Facial Expressions, J.A. Russell and J.M. Fernandez-Dols, eds. (New York: Cambridge University Press), pp. 255–274.

Fernandez-Dols, J.M., Sanchez, F., Carrera, P., and Ruiz-Belda, M.A. (1997). Are spontaneous expressions and emotions linked? An experimental test of coherence. Journal of Nonverbal Behavior *21*, 163–177.

Fogel, A., Nwokah, E., Dedo, J.Y., Messinger, D., Dickson, K.L., Matusov, E., et al. (1992). Social process theory of emotion: A dynamic systems approach. Social Development *1*, 122–142.

Fridlund, A.J. (1994). Human Facial Expression (San Diego, CA: Academic Press).

Frijda, N. (2008). The psychologists' point of view. In: Handbook of Emotions (3rd ed.), M. Lewis, J.M. Haviland-Jones, and L.F. Barrett, eds. (New York: Guilford Press), pp. 68–87.

Gopnik, A., and Meltzoff, A.N. (1997). Words, Thoughts, and Theories (Cambridge, MA: MIT Press).

Griffiths, P.E. (1997). What Emotions Really Are (Chicago: University of Chicago Press).

Izard, C. (1977). Human Emotions (New York: Plenum Press).

James, W. (1884). What Is an Emotion? Mind *19*, 188–205.

Kagan, J. (2007). What Is Emotion? History, Measures, and Meanings (New Haven, CT: Yale).

Kuppens, P., Van Mechelen, I., Nezlek, J.B., Dossche, D., and Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. Emotion *7*, 262–274.

Lacey, J.I. (1950). Individual differences in somatic response patterns. Journal of Comparative and Physiological Psychology *43*, 338–350.

Lacey, J.I. (1967). Somatic response patterning and stress: Some revisions of activation theory. In: Psychological Stress: Issues in Research, M.H. Appley and R. Trumbull, eds. (New York: Appleton-Century-Crofts), pp. 14–37.

Lambie, J.A., and Marcel, A.J. (2002). Consciousness and the varieties of emotion experience: A theoretical framework. Psychological Review *109*, 219–259.

Lang, P.J. (1968). Fear reduction and fear behavior: Problems in treating a construct. In: Research in Psychotherapy (Vol. 3), J.M. Schlien, ed. (Washington, DC: American Psychological Association), pp. 90–103.

Lang, P.J. (1979). A bio-informational theory of emotion imagery. Psychophysiology *16*, 495–512.

Lang, P.J. (1988). What are the data of emotion? In: Cognitive Perspectives on Emotion and Motivation, V. Hamilton, G.H. Bower, and N. Frijda, eds. (Norwell, MA: Kluwer Academic), pp. 173–191.

Lang, P.J., Greenwald, M.K., Bradley, M.M., and Hamm, A.O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. Psychophysiology *30*, 261–273.

Larsen, J.T., Bernstson, G.G., Poehlmann, K.M., Ito, T.A., and Cacioppo, J.T. (2008). The psychophysiology of emotion. In: Handbook of Emotions (3rd ed.), M. Lewis, J.M. Haviland-Jones, and L.F. Barrett, eds. (New York: Guilford Press), pp. 180–195.

Lazarus, R.S. (1991). Emotion and Adaptation (New York: Oxford University Press).

LeDoux., J.E. (1996). The Emotional Brain (New York: Simon and Schuster).

Levenson, R.W., Ekman, P., Heider, K., and Friesen, W.V. (1992). Emotion and autonomic nervous system activity in the Minankabau of West Sumatra. Journal of Personality and Social Psychology *62*, 972–988.

Lewis, M.D. (2000). Emotional organization at three time scales. In: Emotion, Development and Self-Organization: Dynamic Systems Approaches to Emotional Development, M.D. Lewis and I. Granic, eds. (Cambridge, UK: Cambridge University Press), pp. 37–69.

Mandler, G., Mandler, J.M., Kremen, I., and Sholiton, R. (1961). The response to threat: Relations among verbal and physiological indices. Psychological Monographs *75* (No. 513).

Miener, S., and Reisenzein, R. (2009). More on emotional expression and experience: The case of disgust. Unpublished manuscript, University of Greifswald, Germany.

Minsky, M. (2008). A framework for representing emotional states. In: Handbook of Emotions (3rd ed.), M. Lewis, J.M. Haviland-Jones, and L.F. Barrett, eds. (New York: Guilford Press), pp. 618–627.

Neumann, R. (2000). The causal influences of attributions on emotions: A procedural priming approach. Psychological Science *11*, 179–182.

Ortony, A., and Turner, T.J. (1990). What's basic about basic emotions? Psychological Review *97*, 315–331.

Prinz, J. (2004). Gut Reactions: A Perceptual Theory of Emotion (New York: Oxford University Press).

Rachman, S. (1984). Fear and Courage (2nd edn) (New York: Freeman).

Rachman, S., and Hodgson, R.J. (1974). Synchrony and desynchrony in fear and avoidance. Behavior Research and Therapy *12*, 311–318.

Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. Cognition and Emotion *14*, 138.

Russell, J.A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology *39*, 1161–1178.

Russell, J.A. (1991). Culture and the categorization of emotion. Psychological Bulletin *110*, 426–450.

Russell, J.A. (2003). Core affect and the psychological construction of emotion. Psychological Review *110*, 145–172.

Russell, J.A. (2005). Emotion in human consciousness is built on core affect. Journal of Consciousness Studies *12*, 26–42.

Russell, J.A. (2006). Emotions are not modules. Canadian Journal of Philosophy. Supplement Volume *32*, 53–71.

Russell, J.A., and Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of anger. Journal of Personality and Social Psychology *67*, 186–205.

Russell, J.A., Bachorowski, J., and Fernandez-Dols, J.M. (2003). Facial and vocal expression of emotion. Annual Review of Psychology *54*, 329–349.

Scherer, K.R. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In: Emotion, Development and Self-Organization: Dynamic Systems Approaches to Emotional Development, M.D. Lewis and I. Granic, eds. (Cambridge, UK: Cambridge University Press), pp. 70–99.

Scherer, K.R. (2001). Appraisal considered as a process of multilevel sequential processing. In: Appraisal Processes in Emotion: Theory, Methods, Research, K.R. Scherer, A. Schorr, and T. Johnstone, eds. (New York: Oxford University Press), pp. 92–120.

Scherer, K.R., and Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? Emotion *7*, 113–130.

Scherer, K.R. (2009). The dynamic architecture of emotion: Evidence for the component process model. Cognition and Emotion *23*, 1307–1351.

Searle, J.R. (1992). The Rediscovery of the Mind (Cambridge, MA: MIT Press).

Smith, C.A., and Scott, H.S. (1997). A componential approach to the meaning of facial expressions. In: The Psychology of Facial Expressions, J.A. Russell and J.M. Fernandez-Dols, eds. (New York: Cambridge University Press), pp. 229–254.

Smith, E.R., and Neumann, R. (2005). Emotion processes considered from the perspective of dual-process models. In: Emotion and Consciousness, L.F. Barrett, P.M. Niedenthal, and P. Winkielman, eds. (New York: Guilford Press), pp. 287–311.

Strack, F., and Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. Personality and Social Psychology Review *8*, 220–247.

Tamir, M. (2009). What do people want to feel and why? Pleasure and utility in emotion regulation. Current Directions in Psychological Science *18*, 101–105.

Tomkins, S.S. (1962). Affect, Imagery, Consciousness: Vol. 1. The positive affects (New York: Springer).

Tomkins, S.S. (1963). Affect, Imagery, Consciousness: Vol. 2. The negative affects (New York: Springer).

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. Psychological Review *92*, 548–573.

Wierzbicka, A. (1999). Emotions aross Languages and Cultures (New York: Cambridge University Press).

Winkielman, P., Berridge, K.C., and Wilbarger, J.L. (2005). Emotion, behavior, and conscious experience: Once more without feeling. In: Emotion and Consciousness, L.F. Barrett, P.M. Niedenthal, and P. Winkielman, eds. (New York: Guilford Press), pp. 335–362.

Yik, M., Russell, J.A., and Steiger, J.H. (2011). A 12-point circumplex structure of core affect. Emotion *11*, 705–731.

Zachar, P. (2006). The classification of emotion and scientific realism. Journal of Theoretical and Philosophical Psychology *26*, 120138.

Part III

Łukasz Kurek
Copernicus Center for Interdisciplinary Studies

Some Remarks on the Relevance of the Concept of Emotion

1. Introduction

In recent years, the relevance of the concept of emotion for scientifically oriented inquiry has come under considerable scrutiny both in philosophy and science. The doubts concerning the relevance of this concept stem from difficulties in finding a significant number of interesting generalizations concerning the phenomena to which it is supposed to correspond. To many authors this suggests that the concept of emotion corresponds only to folk-psychological artifacts and it fails to 'carve nature at its joints'. In other words, the concept of emotion does not correspond to a natural kind. On the assumption that only natural kinds are the classes which matter for empirical sciences because they are the classes about which one can formulate scientifically relevant generalizations, one can argue that the concept of emotion is not only useless in the inquiry concerning phenomena traditionally described as 'emotions'. Furthermore, it can stand in the way of progress in the inquiry concerning these phenomena by preventing the search for categorizations which correspond to the relevant natural kinds. Therefore, the concept of emotion should be eliminated from scientific and, one might probably add, scientifically oriented philosophical discourse.

However, in this chapter it will be argued that this conclusion is premature. The reason for this is that the argument for eliminating the concept of emotion appeals to a restricted understanding of the relevance of this concept. Even if the concept of emotion does not refer to a natural kind, it still plays an important cognitive role in the inquiry concerning phenomena identified with emotions. Namely, this concept allows for a unification of the inquiry concerning emotions carried out in different disciplines, such as neuroscience, psychology, or philosophy, to name but a few.

In the first part of the chapter, the discussion concerning the elimination of the concept of emotion appealing to the notion of a natural kind will be assessed. It will be argued that this type of eliminative strategy presents an altogether different challenge than the more traditional versions of concept eliminativism. Subsequently, the connection between the notion of a natural kind and the concept of emotion as it is present in the science of emotions will be assessed. The discussion concerning natural kinds in this context will reveal that the assumption that emotion is a natural kind is a working hypothesis among scientists studying emotion. In the third part of the chapter, Joseph LeDoux's proposal of a new conceptual framework for neuroscientific research concerning some features of the phenomena usually understood to correspond to the concept of emotion will be assessed. The last part of the chapter pertains to the argument in defense of the relevance of the concept of emotion as a theory-binding concept which unifies the inquiries concerning emotions undertaken in different disciplines.

2. Eliminating the concept of emotion

There have been various attempts to eliminate the different types of concepts which were understood to be scientifically irrelevant. Perhaps the most well-known proposal of such an elimination was made by Paul Churchland. He argued that folk-psychological concepts such as the concept of belief or the concept of desire, which are supposed to refer to certain kind of objects called propositional attitudes, should be eliminated from all types of discourse (Churchland, 2002). The argument supporting this claim is quite straightforward. Namely, beliefs, desires and other concepts referring to propositional attitudes are defined by the roles of the those concepts in the psychological generalizations in which they feature. However, as Churchland argues, scientific evidence reveals that there are no phenomena to which these concepts correspond. That is the reason why the psychological generalizations featuring these terms are systematically false.

What is so extraordinary about the universal eliminativism mentioned above is that it suggests a complete elimination of certain terms and concepts, not only from scientific and philosophical discourses but even from the everyday discourse. The weakness of such a proposal consists not only in the fact that – because of the roles these concepts fulfill in the moral, legal and other types of human activities – such universal elimination could lead to “the greatest intellectual catastrophe in the history of our species”, as Jerry Fodor plausibly argues (Fodor, 1987). Its main weakness, however, is that it works only on two rather strong assumptions. The first assumption is that the best available empirical evidence points to the fact that there are no entities at the neurological level which fulfill the roles which define these concepts (Machery, 2009, p. 225). It should be noted that if terms such as ‘emotion’ are defined by the role of the corresponding concepts, this proposal generates only claims of a functional nature (Stich and Nichols, 2003). It has nothing to say about the ‘constituent nature’ of the class of emotion – be it neurological, or psychological, or any other. The second assumption consists in adopting a certain theory of reference (Machery, 2009, p. 225). Namely, if the empirical evidence points to the fact that there are no phenomena to which the term ‘emotion’ refers, the elimination of these terms and the corresponding concepts requires the adoption of the descriptivist theory of reference. According to this theory, if there are no entities which satisfy the definition of the term ‘emotion’ – provided by the role of the concept of emotion which features in the generalizations about emotions – this term does not refer and there is no such thing as emotion.

However, on the causal-historical theory of reference, the referent of 'emotion' is not specified by its definition, but by a causal-historical connection between this term and its referent. Therefore, if 'emotion' were linked in this way with a certain class of entities, it refers to this class even if the generalizations featuring the concept of emotion are false. For instance, the generalizations featuring the concept of the superlunary world proposed by Aristotle are false yet this does not mean that there are no superlunary objects.

However, concept eliminativism based on the denial that the concept does not have a natural kind referent is based on a different argumentative strategy. First of all, it does not rest on the assumption that there is nothing in the world which corresponds to the discussed concept. For instance, in the case of the concept of emotion one can argue that it refers to a class of feelings and still argue for the elimination of the corresponding concept of emotion by pointing to the fact that the class of feelings is not a natural kind. Or, somewhat more plausibly, one can argue that one should analyze this concept not by an appeal to its referents but by pointing to the role it plays in the discourse concerning, for instance, its relationship with practices of ascribing blame and responsibility. The difference between this type of eliminativism and the one proposed by Churchland is that the latter is of an all-encompassing character and the former applies only to the scientific discourse and to the discourse within the scientifically oriented philosophy.

The proposal that emotions do not form a homogenous category is certainly not a novelty. For instance, this proposal was extensively discussed by Gilbert Ryle (2009). Recently however, mainly due to the empirical findings from different areas of science of emotion, it has been given a strong justification by the appeal to the notion of a natural kind. The most comprehensive defense of the thesis that emotions do not form a natural kind was proposed by Paul Griffiths (1997).

The notion of a natural kind assumes that there two kinds of classes of objects. The first class consists of objects about which scientific generalizations can be formulated. Such generalizations are possible because objects belonging to this class share a large number of scientifically important properties which are logically unrelated and which are not used only to identify these objects. Typical natural kinds are chemical elements. Each of these elements possesses a large number of scientifically important properties about which generalizations can be formulated (such as electrical conductivity). The second class consists of objects about which no (or only a few) scientific generalizations can be formulated. A popular example of such a class is that of the already mentioned superlunary objects. Although objects of this kind possess a large number of properties, many of these properties are logically related and used only for the identification of these objects.

Here is how Griffiths understands the notion of a natural kind:

> The traditional requirement that natural kinds be the subjects of spatiotemporally universal and exceptionless laws of nature would leave few natural kinds in the biological and social sciences, where generalizations are often exception-ridden or only locally valid. Fortunately, it is easy to generalize the concept of a law of nature to the notion that statements are to varying degrees 'lawlike' (have counterfactual force). This allows a broader definition of a natural kind. A category is (minimally) natural if it would be reasonable to place some degree of reliance on some inductive predictions about unobserved instances. This, of course, is a very weak condition. Very many ways of classifying the world are minimally natural. The aim is to find categories that allow reliable predictions in a large domain of properties. The classic examples of natural kinds, chemical elements and biological species, meet these desiderata (Griffiths, 2004, p. 905).

This understanding of a natural kind rejects two hypotheses about the properties of the objects which belong to such a class. The first hypothesis states that all objects belonging to such a class possess fixed, precise, necessary and sufficient properties because of which they belong to the given natural kind. Although one can point to objects which possess such properties (for instance, chemical elements) this hypothesis is too strong because, if adopted, it would allow for only a few, if any, natural kinds which are interesting for such science as, for instance, biology (biological species certainly do not possess such properties). The second hypothesis is that the properties of the objects which belong to a natural kind allow one to formulate laws of nature which feature such a class. According to the standard understanding, laws of nature are supposed to be spatially and temporally unrestricted and they should support counterfactuals. However, as Griffiths plausibly argues, this hypothesis is too strong for the biological and social sciences. One can plausibly argue that such an understanding of the notion of a natural kind would leave few, if any, natural kinds relevant for these disciplines.

Instead, the natural kind theory adopted by Griffiths appeals to homeostatic property cluster theory proposed by Richard Boyd (1991). This theory is based on two assumptions:

> Assumption 1: Objects which belong to the natural kind category possess a cluster of typical properties.
> Assumption 2: This cluster of properties is supported by a causal homeostatic mechanism which is responsible for their occurrence.

It is because of the functioning of the causal homeostatic mechanisms that one can formulate generalizations featuring properties typical for objects belonging to a given natural kind. The proposal of the existence of the above-mentioned mechanisms is, therefore, what is crucial for the existence of a natural kind. If emotions are to form a natural kind, they should possess a cluster of properties which is generated by the appropriate mechanism.

What is more, in the context of the biological sciences one can distinguish between natural kinds based on homology and analogy (Griffiths, 1997). Homologues are essentially the same structures in different species which descended from a common ancestor. Analogies, on the other hand, are structures which share some resemblance because they were shaped by evolution to fulfill the same role. The importance of this distinction is based on the fact that similarities between different structures which are based on analogy are, as Griffiths describes them, 'shallow' (Griffiths, n.d.). For instance, if one was to compare two quite different species, such as dogs and snakes, in the context of one emotion such as anger, one would find only superficial similarities. It is because the causal mechanisms underlying these emotions are obviously quite different in those species. On the other hand, similarities between structures which are based on homology are 'deep' (Griffiths, n.d.). For instance, comparing two similar species, such as chimpanzees and humans, in the context of emotion of anger, one would find a large number of important similarities. The reason for this is that the mechanisms in these species share a common ancestry.

Griffiths argues that only emotions which fall into the category of 'affect programs' form a natural kind based on homology. Affect programs (the term is borrowed from Paul Ekman) are automatic patterns of response, homologues of which can be found among vertebrates, particularly primates (Griffiths, 1997). Phenomena which belong to this category are identified with the so-called basic emotions (according to the well-known categorization by Paul Ekman there are six basic emotions: fear, anger, disgust, happiness, sadness, and surprise) (Ekman, 1972). These phenomena are understood to be triggered automatically when the agent finds herself in an emotional eliciting situation, they are brief, and realized by the evolutionary 'older' parts of the brain which humans share with many other vertebrates (Griffiths, n.d.).

According to Griffiths, emotions can be divided into two other sub-categories which do not, however, form natural kinds. These categories are higher cognitive emotions and socially constructed emotions (Griffiths, 1997). Both of them are typical to hominids. The main characteristic of the former is that they involve complex evaluations (thoughts) and they are realized both by the evolutionary 'ancient' parts of the brain and evolutionary 'new' parts of the brain. Examples of this type of emotion include guilt or jealousy. The latter are also cognitively mediated, however, they differ among distinct cultures because they involve an "internalized cultural model of appropriate behavior" (Griffiths, n.d.). An example of this type of emotion is the emotion called *amae* which appears in Japanese culture and seems to lack a counterpart in Western cultures. It is a propensity to "depend or presume upon another's love" (Robinson, 2004). Both higher cognitive emotions and socially constructed emotions do not form distinct natural kinds because it is not possible, at least at the present moment, to formulate a large number of scientifically interesting generalizations about them.

These considerations result in doubts over the relevance of the concept of emotion – at least if one is interested in the scientific aspect of the phenomena discussed. However, the argument for the elimination of the concept of emotion appealing to the notion of a natural kind rests on a crucial assumption which should now be easy to notice. Namely, this argument assumes that in the context of scientific inquiry emotions are understood as a natural kind. If this assumption is absent – for instance because it is a too strong assumption pertaining to the phenomena to which the concept of emotion is supposed to refer – the eliminativist threat to the concept of emotion linked with the notion of a natural kind is groundless. Therefore, it is interesting to investigate whether this assumption is actually adopted by the scientists who study emotions.

3. Emotions as natural kinds and the science of emotions

It is quite clear that, at least among the philosophers of emotion, the prevalent understanding of emotion does not rest on the assumption that they form a natural kind. Not only Griffiths but also many other thinkers – who do not appeal to the results of scientific research on emotions as much he does – acknowledge that the concept of emotion refers to phenomena which often possess quite different properties (Ben-Ze'ev, 2000; de Sousa, 1987; Solomon, 1995).

Scientific research pertaining to emotion is difficult from a methodological point of view because emotions are usually correlated with feelings and the latter are of a subjective character. Because of their subjective character, feelings seem to escape scientific method. Furthermore, if feelings are important for understanding what constitutes an emotion, it follows that the concept of emotion also becomes inconvenient for scientific inquiry. However, the commonsense understanding of the link between emotion and feeling has been criticized since the beginning of modern psychology. According to the famous James-Lange theory of emotion formulated at the end of the 19th century, emotions do not precede a bodily reaction to the eliciting event. When we find ourselves in an emotion eliciting situation, our body reacts automatically and the subsequent emotion, such as fear in the above-mentioned example, is only a perception (feeling) of these bodily reactions (James, 1884).

Contemporary neuroscience has adopted this view and, for instance, Joseph LeDoux proposes a similar explanation of fear:

> The conscious fear that can come with fear conditioning in a human is not a cause of the fear response; it is one consequence (and not obligatory one) of activating the defense system in a brain that also has consciousness (LeDoux, 1998, p. 147).

This type of explanation of emotion is at odds with the commonsense explanation of emotion because according to the latter emotions cause action. However, according to the former, it is emotion which is caused by the perception of the prior bodily changes. The concept of emotion seems to be eliminated from the scope of the explanation of behavior as the emotions become epiphenomenal.

This type of understanding of emotions seems to be common among the scientists who study them. It objectifies emotional phenomena and, in turn, allows one to propose a large number of interesting empirical generalizations about them. The identification of emotions with a natural kind seems to be, therefore, an implicit assumption within this field of inquiry. As the psychologist Lisa Barrett describes it:

> Despite the differences in their (emotions – Ł.K.) surface features, many of the most prominent models (of emotions – Ł.K.) share a common set of beliefs about the nature of emotion: Emotions are categories with firm boundaries that can be observed in nature (meaning in the brain or body) and are therefore recognized, not constructed, by the human mind. As a natural kind of emotion, anger, for example, is assumed to be a package of behavioral and physiological changes that are produced by some causal mechanism (in the brain or the mind, again depending on the level of analysis) that is released under certain conditions. Researchers assume that they will know an instance of anger when they see it in the face, voice, or body of another person, or feel it in themselves (Barrett, 2006, p. 32).

This paradigm in the science of emotion proved to be very fruitful in the past. For instance, the discovery of several basic emotions which possess universal biological features among humans and other primates can be given as an example of this assumption.

On the basis of the hypothesis that emotions form a natural kind, Barrett generates two empirically testable hypotheses to which there is a significant amount of evidence which undermines it.

> Hypothesis 1: If certain categories of emotion are natural kinds, characterized by projectable property clusters, then it should be possible to characterize each kind of emotion in terms of a suite of distinctive, observable responses that are coordinated in time and correlated in intensity.

As Barrett goes on to observe, in most of the experiments the actual correlations found between experiential, behavioral and physiological measures of emotion proved to be a lot weaker than expected (Barrett, 2006, p. 33). Despite this discrepancy, psychologists and neuroscientists carry out their research on emotions as if each emotion possessed distinct characteristics described in terms of subjective experience, facial behavior, neuronal pattern of activation, and so on. One can even point to typical strategies of explaining the lack of correlation (for instance, social factors masking responses which would otherwise materialize; the inability to produce strong enough stimuli to elicit prototypical emotions in laboratory studies; or that the emotional responses vary across individuals) (Barrett, 2006, p. 33).

The second hypothesis derived by Barrett from the assumption that emotions form a natural kind states the following:

> Hypothesis 2: Different kinds of emotion have distinct causal mechanisms.

Also in relation to this hypothesis there is a large body of evidence which undermines the relation between emotions and psychological or neural mechanism which are connected with them. Barrett discusses four types of mechanisms which are commonly related with emotions: mechanisms producing subjective experience, mechanisms producing facial and vocal signals, mechanisms producing behavior, and neural mechanisms observed in the neuroimagining studies (Barrett, 2006, pp. 34–45). From these considerations it becomes evident that the correlations between distinct types of mechanisms and distinct types of emotion is a lot weaker than one could expect on the basis of the assumption that emotions form a natural kind. Her suggestion is more radical in comparison with Griffiths', as she proposes to give up the assumption of there being even natural kinds which correspond to basic emotions.

4. Replacing emotions with survival circuits

The wide-ranging critique of identifying emotion with a natural kind and the subsequent attempts at eliminating the concept of emotion from the scientific discourse seem to gain in significance in the scientific inquiry concerning emotion. An interesting proposal in this context has recently been made by Joseph LeDoux (2012). The idea underlying this proposal is to study key phenomena correlated with the concept of emotion without referring to the concept itself. The phenomena which are of interest to LeDoux in this context are:

> (…) responses that occur when an organism detects and responds to significant events in the course of surviving and/or maintaining well-being – for example, responses that occur when in danger or when in the presence of a potential mate or in the presence of food when hungry or drink when thirsty. These are fundamental phenomena that have always interested animal behavior scientists, and would be of interest even if the terms "emotion" and "feelings" never existed. The challenge for emotion researchers is to understand the relation of the phenomena to the field of emotion without redefining them as fundamentally emotional phenomena, and thus infusing the phenomena with confusing implications (LeDoux, 2012, pp. 653–654).

The organisms' responses which occur in the circumstances indicated by LeDoux include not only responses traditionally correlated with emotion (such as characteristic facial expressions in humans) but also responses connected with motivation, reinforcement and arousal (LeDoux, 2012, p. 654). As LeDoux claims, the latter three types of responses are strongly connected with emotion so it is only natural to treat them as components of same system.

There are at least two advantages of a research programme focused on studying systems of this type, which are described by LeDoux as survival circuits, over research focused on studying the neural mechanisms underlying emotion. The first advantage, which is made explicit by LeDoux, is that adopting this strategy allows one to bypass many traditional difficulties pertaining to the research on phenomena connected with emotions. Such difficulties particularly include investigating such phenomena on the basis of the outward expressions of emotional responses or verbal declarations by a person who feels a certain emotion. Survival circuits include at least circuits responsible for defense, maintenance of energy and nutritional supplies, fluid balance, thermoregulation, and reproduction (LeDoux, 2012, p. 655). They possess a cluster of typical properties which are directly accessible to the researcher, unlike subjective feelings often identified with occurrences of the appropriate emotion responses. The second advantage, which seems to be implicit in the survival circuit research programme, is that these circuits are present in many other species which are often not closely related from the evolutionary point of view. Furthermore, survival circuits are realized by biological mechanisms which are, to a large extent, of a similar structure between the species. Because of these features of survival circuits, one can formulate a large number of scientifically interesting generalizations about them, and, as a consequence, one can plausibly claim that they form natural kinds.

As LeDoux indicates, all complex animals, both invertebrates and vertebrates, have survival circuits (LeDoux, 2012, p. 655). However, the origins of these circuits can be found even in early life forms (LeDoux, 2012, p. 655). Even single-cell organisms can respond in a particular way when it detects and events significant to its survival or well-being. What is also interesting is that survival circuits generate an organism's responses whether they are innate or learned. It is not the case that these circuits are hard-wired but they allow for new responses to be formed on the basis of novel stimuli (LeDoux, 2012, p. 656).

LeDoux's proposal originates from the above-mentioned difficulties with formulating a large number of scientifically interesting generalizations featuring the concept of emotion. One can notice that perhaps the main reason for this is that the argument against identifying emotions with natural kinds can be directed even against the so-called basic emotions, as it was argued by Barrett. LeDoux agrees with this critique, emphasizing that "the foundation of support for the idea that basic emotions, as conventionally conceived, have dedicated neural circuits is weak" (LeDoux, 2012, p. 655). Because the causal mechanism which is supposed to bring about the phenomena possessing the cluster of typical properties for a given type of emotion is supposed to be of a neural character, such a statement essentially denies the existence of such a causal mechanism. As a consequence, it is also a denial of the ability to identify basic emotions with natural kinds. On the other hand, survival circuits seem to fall into the natural kind category. However, as it is clear from the considerations above, they do not correspond to even basic emotions. It seems, therefore, that even in the context of basic emotions which are, as far as affective science goes, emotions which are best understood, the prospect of any relevance of the concept of emotion looks grim indeed.

5. The concept of emotion as a theory-binding concept

The above discussion concerning the elimination of the concept of emotion rests on the assumption that only concepts which correspond to natural kinds are scientifically interesting. However, it seems that it is a too restrictive criterion for a concept to be useful for scientifically oriented inquiry. There are a large number of concepts connected with cognition, such as the concepts of emotion, perception, reasoning, or representation, which – despite the fact that they do not seem to correspond to natural kinds – remain very useful for scientifically oriented inquiry. Their usefulness stems from the fact that they help to unify inquiries carried out in different disciplines, including neuroscience and psychology.

Despite the fact that the concept of emotion fails to correspond to a natural kind (which could be interesting for neuroscience) it still seems to correspond to a kind of phenomena about which one can formulate interesting generalizations. For instance, emotions are usually conscious, they often involve characteristic bodily manifestations, or they generally arise in situations important for the well-being of the agent. These generalizations are, of course, very different from scientific laws which hold universally because the *ceteris paribus* clause connected with them is wide. Still, these fragile generalizations seem to be interesting enough to make the concept of emotions relevant for scientifically oriented inquiry as the concept unifying the research on emotions in different fields.

This unification is desirable because the explanation of emotions, similarly to other phenomena related to cognition, is hierarchical. The phenomena corresponding to the concept of emotion cannot be properly understood unless one explores a range of theories in which they feature – theories of not only neuroscientific character. The study of emotions can be undertaken from a variety of perspectives, utilizing different methods. The disciplines aiming at the explanation of emotions include anthropology, experimental psychology, neuroscience, evolutionary psychology and also philosophy with its method of conceptual analysis. It seems plausible to claim that none of these disciplines is the only relevant discipline to acquire knowledge about emotions. Each of them is suited for answering different questions about phenomena corresponding to the discussed concept. Neuroscience, as one can observe from the considerations above, is best suited for answering questions about the neural circuits involved in the production of emotions. Furthermore, the methods of neuroscience seems to be especially effective in investigating phenomena correlated with basic emotions, such as fear. Anthropology, on the other hand, aims at answering the questions of whether there are any differences in emotions between different cultures or within a culture. The method of this discipline, similarly to the method of experimental psychology, enables one to investigate the higher cognitive emotions such as guilt. In addition, evolutionary biology can provide a historical explanation of the presence of human emotions.

Arguably, a fully-fledged account of the phenomena corresponding to the concept of emotion should refer not only to the neuroscientific level of explanation, but also to other levels, such as psychological and anthropological. Furthermore, any account which undertakes to combine different level explanations of a given phenomenon should utilize not only the components of explanations present at each level (e.g. psychological or neural mechanisms) but also components allowing to bind the different levels of explanation. A useful conceptual tool in this context is proposed by Jose Bermudez (2005). Bermudez described the specific explanations present at each level of explanation as *horizontal explanations* whereas the components allowing the binding of different levels of explanation belong to the *vertical explanation* category.

The horizontal explanation of a particular event or state consists in indicating distinct events or states which are, at least usually, temporally antecedent from the events which are being explained. The next stage of this kind of explanation consists in a proposal of a relation between these events. The paradigm case of a horizontal explanation is causal explanation, where the relation holding between the two categories of events is described as causal (Bermudez, 2005, p. 32). For example, if we want to explain the firing of a neuron, we can refer to the events happening at the synaptic junctions between the neurons as the causes of the firing of the neuron. Such an explanation is a typical case of a horizontal explanation.

As Bermudez plausibly claims, in some situations it is not enough to propose only a horizontal explanation of a given phenomenon. In such cases, an explanation of the fact that the above-mentioned relations between the events hold is needed. In the case of the firings of the neurons, a very relevant question would be to ask how, in some cases, the events at the synaptic junctions happen. This question can be answered at the level lower than that of a whole neuron, for example by reference to the activity of neurotransmitters. According to Bermudez, explaining why horizontal relations between distinct events hold is the aim of *vertical explanation.* What is important here is that often when a vertical explanation is proposed, it consists in explaining the relations at the higher level of explanation by reference to the relations at the lower level of explanation. Therefore combining horizontal and vertical explanation usually leads to a hierarchical explanation of a given phenomenon.

The concept of emotion, although its relevance for the horizontal explanations in neuroscience or psychology seems to be undermined by the arguments based on the natural kind strategy – at least at the present moment of the development of the science of emotions – seems to be a useful concept for the vertical explanation of the phenomena corresponding to the discussed concept, binding together the different horizontal explanations of these phenomena. In that regard, the concept of emotion can be understood as a theory-binding concept of which one cannot provide a precise definition and which can be defined by the role it plays in different types of theories in which it features (for instance, theories of a neuroscientific, psychological, anthropological, or philosophical character). A somewhat similar idea of law-cluster concepts was proposed by Hilary Putnam in the context of the concept of energy:

> The concept ‘energy’ is a great example of a law-cluster concept. It enters into a great many laws. It plays a great many roles, and these laws and inference roles constitute its meaning collectively not individually. I want to suggest that most of the terms in highly developed science are law-cluster concepts, and that one should always be suspicious of the claim that a principle whose subject term is a law-cluster term is analytic. The reason it is difficult to have an analytical relationship among law-cluster concepts is that such a relationship would be one more law. But, in general, any one law can be abandoned without destroying the identity of the law-cluster concept involved, just as a man can be irrational from birth, or have a growth of feathers all over his body, without ceasing to be a man (Putnam, 1975, p. 52).

On this account, to explain the meaning of a law-cluster concept one needs to point to the laws in which this concept features. This idea seems to be plausible in the case of at least certain natural sciences, for instance physics. As Putnam points out, there are many laws which involve the concept of energy, similarly as in the case of other concepts used in physics. The situation is, however, different in the case of a large number of concepts related to the mind, an instance of which is the concept of emotion. The difference lies in the fact that it is difficult to point to generalizations of, for instance, a psychological or neuroscientific character, in which the concept of emotion features. Although that is a strong premise for the conclusion that this concept does not refer to a natural kind, the relevance of this concept can be defended on different grounds. Namely, the discussed concept can be viewed as the already mentioned theory-binding concept. The meaning of a theory-binding concept can be reconstructed with reference to different explanatory theories which feature this concept.

6. Conclusion

The relevance of emotions for scientifically oriented inquiry is at the present moment often undermined on the basis of the observation that this concept fails to correspond to a natural kind. This strategy of eliminating concepts is perhaps most persuasive when discussed in the context of the neuroscience of emotions, where one can find evidence supporting the claim that different emotions are produced by different kinds of neural mechanisms. One can even point to a persuasive proposal of replacing the concept of emotion with different concepts within the neuroscience of emotions. However, it was argued that this strategy is based on a limited view of the nature of scientific concepts. Not all of the concepts which are relevant for scientifically oriented inquiry should correspond to natural kinds. Namely, there is a type of concept – which can be usefully described as theory-binding as theory-binding concepts – which do not correspond to natural kinds and, at least at the given moment of the development of science, these concepts allow for a unification of the different levels of explanation of a certain type of phenomena. The concept of emotions is an example of a theory-binding concept.

References

Barrett, L. (2006). Are emotions natural kinds? Perspectives on Psychological Science *1(1)*, 28–58.

Ben-Ze'ev, A. (2000). The Subtlety of Emotions (Cambridge, MA: MIT Press).

Bermudez, J. (2005). Philosophy of Psychology (London: Routledge).

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. Philosophical Studies *61*, 127–148.

Churchland, P. (2002). Eliminative materialism and the propositional attitudes. In: Philosophy of Mind. Classical and Contemporary Readings, D. Chalmers, ed. (Oxford: Oxford University Press), pp. 568–580.

Ekman,P. (1972). Universals and cultural differences in facial expressions of emotions. In: Nebraska Symposium on Motivation 1971 (Lincoln, NE: University of Nebraska Press), pp. 207–283.

Fodor, J. (1987). Psychosemantics (Cambridge, MA: MIT Press).

Griffiths, P. (1997). What Emotions Really Are: The Problem of Psychological Categories (Chicago: The University of Chicago Press).

Griffiths, P. (2004). Emotions as natural and normative kinds. Philosophy of Science *71*, 901–911.

Griffiths, P., Emotion is (still) not a natural kind. http://www.philosophy.dept.shef.ac.uk/AHRB-Project/Papers/GriffithsPaper.pdf.

James, W. (1884). What is an Emotion? In: What is an Emotion?, C. Calhoun, R. Solomon eds. (New York: Oxford University Press), pp. 66–76.

LeDoux, J. (1998). The Emotional Brain: The Mysterious Underpinnings of Emotional Life (New York: Simon & Schuster).

LeDoux, J. (2012). Rethinking the emotional brain. Neuron *23*, 653–676. Reprinted in this volume.

Machery, E, (2009). Doing without Concepts (New York: Oxford University Press).

Putnam, H. (1975). The analytic and the synthetic. In: Mind, Language, Reality, H. Putnam, ed. (Cambridge, MA: Cambridge University Press).

Robinson, J. (2004). Emotion. Biological fact or social construction? In: Thinking About Feeling. Contemporary Philosophers on Emotions, R. Solomon, ed. (Oxford–New York: Oxford University Press).

Ryle, G. (2009). The Concept of Mind: 60th Anniversary Edition. (London: Routledge).

Solomon, R. (1995). Some notes on emotion, 'east and west'. Philosophy East and West *45(2)*, 171–202.

Sousa, R. de (1987). The Rationality of Emotion (Cambridge, MA: MIT Press).

Stich, S. and Nichols, S. (2003). Mindreading. An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds (Oxford: Oxford University Press).

Mateusz Hohol, Piotr Urbańczyk
The Pontifical University of John Paul II in Cracow,
Copernicus Center for Interdisciplinary Studies

Some Remarks on Embodied-Embedded Social Cognition[1]

1. Paradigm shifts

Is human behaviour mainly determined by our emotions or by reason? This is a question which has concerned many philosophers for some time and, during the last three decades, there have been a number of passionate debates on the relationship between emotion and cognition between psychologists and cognitive scientists. Perhaps the most important adversaries in these discussions were Robert Zajonc and Richard Lazarus (cf. Zajonc, 1980; Lazarus, 1984). They were both convinced that the systems of affect and cognition are separate but they disagreed, however, about the priority of these systems in their effect on behaviour. Zajonc believed that emotions are primary with respect to cognition, while Lazarus opted for the primacy of cognition. According to the former, emotions are always accompanied by cognition (thinking), but cognition (thinking) does not always have an impact on our emotions. The latter believed that Zajonc's view could be accepted only if cognition is understood very narrowly – without the processes of attention and the interpretation of the recorded stimuli. Since detailed arguments on both

---

[1] This contribution was made possible through the support of a grant "The Limits of Scientific Explanation" from the John Templeton Foundation.

sides are rather well known to psychologists, philosophers and cognitive scientists, in this work they will be omitted and we will focus on what this discussion has led to.

It seems that the final outcome of the debate between Zajonc and Lazarus was achieved thanks to research findings from neuroscience. In particular, Joseph LeDoux discovered and studied two pathways of inducing affect – the high (cortical) road and the low road (LeDoux, 1996). The latter is particularly important. It transpired that the phylogenetically older cognitive structures of our brains, such as the thalamus, are directly connected to the amygdala, which is one of the emotional centers of the brain. For this reason, the reaction to experienced stimuli is much faster in the low road than in the cortical pathway. In many cases the reaction is much faster than constructing a conscious representation.

It seems, therefore, that the above-mentioned discussion has ended with the adoption of Zajonc's approach. Nevertheless, this approach was modified or weakened since emotion can significantly affect cognition, but also – vice versa – cognition has an impact on emotion. Moreover, a considerable amount of other data, e.g. from lesion studies, indicates that prioritizing cognition or the emotions is rather a misguided strategy. This leads to a more general conceptual change: "cognitive processes" are not only traditional issues, such as memory and attention, but – at least implicitly – also concern emotions and decision-making. The study of their mutual influences and relationships serves to secure a better understanding of how the body works. Thus, one can certainly claim that we are dealing with a significant paradigm shift in this case.

Nonetheless, this was not the only paradigm shift that has taken place in inquiries on the nature of mind and cognition. From the 1980's, one can observe a move from a computational paradigm of cognitive science and cognitive neuroscience to the paradigm of the embodied-embedded mind. While the proponents of the computational paradigm treated the mind as software implemented in biological hardware, the followers of the embodied-embedded mind paradigm regard cognitive processes as products of motor processes and interactions, in which the individual enters the physical, social and cultural environment (Varela et al., 1993; Chemero, 2009; Johnson, 2007). In other words, the basic concepts "filling" our minds are generated on the basis of neural programs of motor control and express spatial relationships. More abstract concepts (e.g., concerning feelings and emotions) are created by "mental machinery" based on specific concepts and with the use of conceptual metaphors (as tools for understanding and action).

The embodied-embedded mind paradigm leads in turn to another revolution in the philosophy of mind, neurophilosophy and – last but not least – social cognition. This revolution concerns both cognition of one's own mind (self-knowledge) and the minds of other people. Generally speaking, the old view – dating back to Descartes – according to which knowledge of our own mental states is more primary than understanding the mental states of other people, gave way to an approach in which the cognition of other's minds is more primary than self-cognition and self-knowledge. We will try to explain what this paradigm shift is but will first present the scientific revolution which took place in social cognition through the development of a conceptual framework known as the embodied-embedded mind.

2. From brain maps to simulation

For a number of years Antonio Damasio has argued that one of the most important features of our brains is the ability to create maps (Damasio, 2010, chapter 3). In Damasio's view, maps are neural patterns created when the body *interacts* with the objects of the world (environment) as well as the minds of others. These maps are not only used in the unconscious self-regulatory activities of the body, but also play a major role in conscious cognition. The basic object of mapping is the body. As Damasio states:

> The human brain maps whatever object sits outside it, whatever action occurs outside it, and all the relationships that objects and actions assume in time and space, relative to each other and to the mother ship known as the organism, sole proprietor of our body, brain, and mind. The human brain is a born cartographer, and the cartography began with the mapping of the body inside which the brain sits (Damasio, 2010, p. 66).

The mechanism of creating brain maps explains how conscious body images, as well as body schemas related to sensory-motoric processes, are formed. Advocates of the embodied mind idea attach great importance to the latter (Gallagher, 2007).

It is not only cortical structures but also the phylogenetically older subcortical structures (such as the geniculate bodies, the colliculi, the nucleus tractus solitarius, and the parabrachial nucleus) which are able to create maps with especially the last two of these structures involved in body mapping (Damasio, 2010, chapter 3). A cerebral representation of what is happening at any given time to our body is important not only for motor activities undertaken in the physical environment. In the later part of this work we will show that this mechanism is involved in the formation of more complex content "inhabiting" our minds. The process of mapping the body is intricate, not only on the purely biological level – it also comprises all of the interactions between the body and the environment and, therefore, human activity in the physical and social setting. Damasio writes that

> Signals sent by sensors located throughout the body construct neural patterns that map the organism's interaction with the object. The neural patterns are formed transiently in the varied sensory and motor regions of the brain that normally receive signals coming from specific body regions. The assembling of the transient neural patterns is made from a selection of neuron circuits recruited by the interaction (Damasio, 2010, p. 74).

Nevertheless, generating maps of the current state of the body does not exhaust the possibilities of the brain. Indeed, the brain is able to transform created maps and simulate possible states. This brain-body communication is carried out in both directions. Cognitive processes and conscious representations can be causes of emotions that "engulf" the whole body. On the other hand, the body – in a particular emotional state – forms mental states. Moreover, in such mechanisms of mutual communication (referred to as resonant loops) Damasio sees the sources of self-consciousness:

> I envision these responses as initiating a tight two-way, resonant loop between body states and brain states. The brain mapping of the body state and the actual body state are never far apart. Their border is blurred. They become virtually fused. The sense that events are occurring in the flesh would arise from this arrangement. A wound that is mapped in the brain stem (within the parabrachial nucleus), and that is perceived as pain, unleashes multiple responses back to the body. The responses are initiated by the parabrachial nucleus and executed in the nearby periaqueductal gray nuclei. They cause an emotional reaction and a change in the processing of subsequent pain signals, which immediately alter the body state and, in turn, alter the next map that the brain will make of the body (Damasio, 2010, p. 105).

Let us return to the above-mentioned simulation. Damasio goes a step further by claiming that the brain is able to quickly generate a variety of alternative maps of the body. Thus, the maps of the states in which the body could find itself if it had been under the influence of a given emotion are also simulated. The simulation of an emotion can anticipate its actual appearance, or even completely replace the appearance of this emotion. A simulation of this type is called by Damasio 'as-if body loop' and involves both the structures responsible for the operation and somatosensory system structures. An example of the 'as-if body loop' is a connection of emotion and compassion centers, such as the amygdala and ventromedial prefrontal cortex, to the structures responsible for the state of the body processing, such as the insular cortex, SII, SI, and the somatosensory association cortices (Damasio, 2010, chapter 3–4). In addition (or, perhaps, above all), mechanisms of simulations are supported by the mirror neuron system. It allows us to feel, or to put our bodies in states in which we actually would find ourselves while experiencing a given emotion or performing a specific gesture. Damasio proposes an evolutionary scenario for explaining the formation of the mirror neuron system:

> I suspect that the system developed from an earlier as-if body loop system, which complex brains had long used to simulate their own body states. This would have had a clear and immediate advantage: rapid, energy-saving activation of the maps of certain body states, which were, in turn, associated with relevant past knowledge and cognitive strategies. Eventually the as-if system was applied to others and prevailed because of the equally obvious social advantages one could derive from knowing the body states of others, which are expressions of their mental states (Damasio, 2010, p. 109).

The issue of mirror neurons and their involvement in social cognition will be examined more closely in the following paragraphs.

3. Through the looking-glass: mirror neurons

Damasio's view indicates that the mapping of our bodily states by our brains is an important factor in the development of cognition. Our cognitive abilities depend largely on the motor actions of our bodies. Mapped bodily states are not in a vacuum, but are formed in response to the challenges posed to organisms by the surrounding environment – the natural, social and cultural environment.

A number of neuroscientists from Parma played an extremely important role in the study of the influence of motor skills on cognitive processes: Giacomo Rizzolatti, Giuseppe Di Pellegrino, Vittorio Gallese, Leonardo Fogassi, Marco Iacoboni and Luciano Fadiga. By studying the activity of neurons in the F5 motor cortex area of macaques, they obtained three surprising results about the action of understanding (Di Pellegrino et al., 1992; Fogassi and Gallese, 2002). It was known that nerve cells belonging to this cerebral structure coded the movement of a macaque's hand which accompanied the manipulation of objects (e.g. food). Firstly, it turned out that some neurons belonging to F5 are also activated when the macaque grasps food not only with his hand, but also directly with its mouth. The neuroscientists from Rizzolatti's team concluded that the neurons in the F5 area encode abstract representations of manipulation of objects, ignoring the part of the body which performs the action. Secondly, in this structure they indicated "canonical neurons" that respond only to objects with specific shapes and properties. Nevertheless, the true revolution came with the third result.

Studies have shown that some of the neurons in the F5 area are activated both when the motor system is used and in the case of utilizing visual representations of certain operations. The activity of nerve cells called "mirror neurons" can be registered in both cases – when a macaque manipulates an object (such as food) and when he sees that another macaque or experimenter executes such an operation. For example, some F5 area cells are activated both when a macaque grabs a peanut and when he sees that another individual perform a similar movement, subordinated to the same purpose (object manipulation). In summary, the experiments conducted by Rizzolatti and his team provided a strong basis to believe that mirror neurons encode not only the movements, but also the goals that these movements are meant to achieve. The F5 area is probably the only structure in the brain of macaques which is equipped with mirror neurons.

Nowadays they believe that there are many cerebral structures in which one can observe the phenomenon of the multi-modal resonance of neurons. While many neuroscientists accept mirror neurons as a "hard scientific fact" on which you can base hypotheses about human nature (Ramachandran, 2010; Gazzaniga, 2009), it should be noted that there is still an ongoing debate over the "nature" of mirror neurons. For example, it is still unclear whether a particular neuron is capable of resonance due to ultimative specialization and its structure or whether it is due to the role it plays in a particular network. In the latter case it can be expected that every – or virtually every – neuron in our brain is "potentially" capable of multi-modal resonance (Winkielman et al., 2009, p. 239).

Despite the problem mentioned above, the convergence of data collected from various sources suggests that there are also mirror neurons in human brains. It is believed that they are located, inter alia, in the 44 Brodmann area, which is the human equivalent of F5 area in the cerebral cortex of macaques. Although (mainly for ethical reasons) it is difficult to demonstrate their existence with single-cell recording, the hypothesis of the existence of a mirror system in the human cerebrum is supported by data obtained with EEG, MEG, TMS, PET and fMRI (Rizzolatti et al., 2002). For example, PET imaging indicates that in such structures of the human brain as the upper parieto-occipital sulcus, lower parietal lobe and a lower angular gyrus there are mirror neurons that are activated both during object manipulation and the observation of such operations performed by another individual (Rizzolatti et al., 1996). What is important for the hypotheses posed later in the article is that it is assumed that cells capable of multi-modal resonance are located in the emotional structures of the brain such as the insula, anterior cingulate, somato-sensory cortex, superior temporal sulcus, the extrastriate body area and the dentate of the cerebellum (Dacety and Jackson, 2004).

4. Embodied simulation and peripersonal space

The discovery of mirror neurons not only strengthened the conviction of a relationship between cognitive processes and emotions with the motor actions of the body, but it also led to the consolidation of the embodied-embedded mind paradigm. It became the basis for various theories constructed in the framework of this paradigm. The idea of *embodied simulation* formulated by Vittorio Gallese seems to be one of the most important (Gallese, 2005). The following quote captures its essence succinctly:

> I employed the term simulation as an automatic, unconscious, and pre-reflexive functional mechanism, whose function is the modeling of objects, agents, and events. Simulation (…) is therefore not necessarily the result of a willed and conscious cognitive effort, aimed at interpreting the intentions hidden in the overt behavior of others, but rather a basic functional mechanism of our brain. However, because it also generates representational content, this functional mechanism seems to play a major role in our epistemic approach to the world. It represents the outcome of a possible action, emotion, or sensation one could take or experience, and serves to attribute this outcome to another organism as a real goal-state it is trying to bring about, or as a real emotion or sensation it is experiencing. Successful perception requires the capacity of predicting upcoming sensory events. Similarly, successful action requires the capacity of predicting the expected consequences of action. As suggested by an impressive and coherent amount of neuroscientific data (…), both types of predictions seem to depend on the results of unconscious and automatically driven neural states, functionally describable as simulation processes.

Gallese also points out that

> (…) Simulation is not conceived of as being confined to the domain of motor control, but rather as a more general and basic endowment of our brain. It is mental because it has content, but it is sensory-motor because its function is realized by the sensory-motor system. I call it "embodied" – not only because it is neurally realized, but also because it uses a pre-existing body-model in the brain, and therefore involves a non-propositional form of self-representation (Gallese, 2005, pp. 41–42).

The mechanism of embodied simulation is based on the phenomenon of the multimodality of neurons. In particular, numerous studies indicate that the same neurons that are responsible for sensory information processing also play an important role in the motor control. Gallese carefully examines the properties of the cortical premotor-parietal network F4-VIP in macaques' brain. The F4 area occupies the posterior sector of the ventral premotor cortex, yet the structure of the VIP occupies the fundus of the intraparietal sulucus. This structure is involved in the control of purposeful movements of the head and forearm.

Single-cell recording shows that, apart from visual neurons, bimodal visual-tactile cells are located in the VIP structure. Cells belonging to this structure fire in response to both visual and tactile stimuli and the receptive field of these cells are related to the face and merged with the touch field creating peripersonal space. VIP area lesions result in the neglect of contralateral peripersonal space, which is one of the symptoms of hemispatial neglect (Darby and Walsh, 2005, chapter 6).

In the F4 area, apart from unimodal (sensory) neurons, there are also bimodal cells encoding both visual and somatosensory representations as well as trimodal cells which additionally encode auditory representations. The receptive fields of these neurons also participate in the formation of peripersonal space. The receptive fields of most F4 cells are independent of eyeball movements which, according to Gallese, "indicates that the visual responses of F4 do not signal positions on the retina, but positions in space relative to the observer" (Gallese, 2005, p. 25). In the embodied simulation theory, peripersonal space, which is an important element of a sense of corporeality, is closely associated with the motor processes. Research on brain damage indicates that the frontal lobe and inferior paritetal lobule play an important role in maintaining the body schema as well as identifying and raising awareness of the features of objects in peripersonal space.

5. Foundations of social cognition

The embodied simulation theory applies not only to the structure of physical space, but also – metaphorically speaking – to social space. In this way we come to the application of embodied simulation theory to embodied social cognition (Gallagher, 2008; Gallese, 2009). Gallese points out that social cognition has an intentional character and its aim is to predict the beliefs, feelings, desires, and intentions of other individuals. In his view:

> From a first-person perspective, our dynamic social environment appears to be populated by volitional agents capable of entertaining, similarly to us, an agentive intentional relation to the world. We experience other individuals as directed at certain target states or objects, similarly to how we experience ourselves when doing so (Gallese, 2005, p. 31).

In the philosophy of mind there is a clichéd scheme according to which either the problem of other minds (Searle, 1992) arises due to the inviolable first-person ontology of mental states, or those states are completely eliminated (Churchland, 1981). In contrast to this scheme, people constantly – and to a large extent independently of their will – engage in mind-reading. Nevertheless, scientists still argue about the mechanism responsible for this cognitive ability (Kurek, 2013). Proponents of the concept known as the theory-theory believe that the reading of other minds is possible through reasoning (Meltoff, 2005). More recent versions of the theory-theory differ from the classical "analogical" approach according to which people reason with the usage of the principle: "if I find myself in a particular state and behave thus and so, then I can suppose that someone else behaving in the same manner feels the way I do."

Simon Baron-Cohen identifies four stages of reading the minds of others (Baron-Cohen, 1997). The so-called Intentionality Detector works during the first period of ontogeny. It interprets the behavior of potential agents recorded in all modalities (visual, auditory, tactile) as intentional and volitional. At the second stage, the Eye-Direction Detector starts functioning. It comprises the detection of eyeballs, monitoring gaze direction and reasoning (based on experience) that if someone's gaze is focused on an object, she can see it. At the first two stages, dyadic relations between 'I' and an object are formed. In 9–18 month old infants the Shared-Attention Mechanism is developed. It allows one to construct triadic relations between 'I', 'you' and an object. Almost every play in which mother and child draw attention to the same object can serve as its example. At the fourth stage, healthy 3–4 year old children develop the Theory-of-Mind Mechanism. As Baron-Cohen writes, "ToMM is a system for inferring the full range of mental states from behavior – that is, for employing a 'theory of mind'" (Baron-Cohen, 1997, p. 51). It provides the possibility of representing epistemic states of mind such as thinking, knowledge, belief, imagining, guessing and cheating and combines volitional, perceptual and epistemic states into a coherent theory of how mental states and actions are linked. Thanks to this, children with a properly developed ToMM are able to pass a false belief task. As Baron-Cohen points out:

> (…) Children use their mentalistic knowledge in highly theory-like ways (…). Children probably could also affirm a long list of axioms that constitute the core of their theory of mind, though as yet only a fraction of these have been explicitly stated and tasted (such as 'seeing leads to knowing,' 'appearance is not necessarily the same as reality,' 'people are attracted to things they want,' and 'people think that things are where they last saw them') (Baron-Cohen, 1997, pp. 54–55).

Proponents of embodied simulation does not agree that mind reading consists of carrying out reasoning that results in the transition from "first-person" to "third-person" state. According to Gallese, reading other minds is immediate and intrusive:

> We are just *attuned to the intentional relation displayed by someone else* (…). We are not alienated from the actions, emotions and sensations of others, because we entertain a much richer and affectively nuanced perspective of what other individuals do, experience, and feel. What makes this possible is the fact that *we own* those same actions, emotions, and sensations (Gallese, 2005, p. 31).

Most advocates of the theory-theory believe that a certain type of reasoning serves as a "bridge" between one's own mental states and the cognition of mental states of others. The starting point of this approach is subjectivity, namely the 'I' perspective. The adoption of a 'you' perspective is something secondary. The embodied simulation assumes the existence of primary intersubjectivity, which can be called metaphorically the sphere of 'we'. Thanks to this, the minds of others are no less accessible than one's own mind. Moreover, Gallese tries to convince us that on both the phylogenetic (evolutionary) and ontogenetic level, primary intersubjectivity is prior to the ability to distinguish between the 'I' and 'other'. This idea is formulated in Gallese's *shared manifold hypothesis*:

> The shared blended space enables the social bootstrapping of cognitive and affective development. Once the crucial bonds with the world of others are established, this space carries over to the adult conceptual faculty of socially mapping sameness and difference ("I am a different subject") (…). The shared space provides an incredibly powerful tool for detecting and incorporating coherence, regularity, and predictability in the course of an individual's interactions with his or her environment. The shared space is progressively joined by perspectival spaces defined by the establishment of capacities to distinguish the self from others while self-control is developing. Within each of these perspectival spaces information can be further segregated in discrete channels (visual, somatosensory, etc.), making our perceptual view of the world more finely grained. The concurrent development of language probably contributes to further separating

> out of single characters or modalities of experience from the original multimodal perceptual world, but the shared intersubjective space does not disappear. It progressively acquires a different role: to provide our self with the capacity simultaneously to entertain self–other identity and difference (…). My proposal is that the "selfness" quality we readily attribute to others, the inner feeling of "being like me" triggered by our encounter with others, is the result of this preserved blended intersubjective space. Self–other physical and epistemic interactions are shaped and conditioned by the same body and environmental constraints (Gallese, 2005a).

Gallese argues that the neural substrate of tuning in to the mental states of others is the mirror neuron system described above. That description should now be supplemented with a few details, to which Gallese draws more attention (Umiltà et al., 2001). Later studies on mirror neurons show that those cells not only fire when macaques see the entire course of operation (the entire sequence of movements, which leads to the goal), but also when the ape observes an object and the initial movements, yet the final steps and the goal achievement are covered. This result strongly supports the hypothesis according to which the cells of the F5 area encode goals. Gallese adds that this experiment reveals the mechanism of embodied simulation. Covering the goal achievement induces a "perceptual gap" in macaque's mind. It is a simulation which is responsible for filling the gap. Although the achievement of the goal (e.g., a result of object manipulation) is not directly observed, the simulation assigns it to the general pattern, which under normal circumstances would comprise both observed and unseen events (Gärdenfors, 2006).

Another example relates to certain neurons in the F5 ventral premotor cortex, whose function is to encode the movements of the macaque's mouth (Ferrari et al., 2003). These neurons show activity in both cases – when the macaque itself performs some operations and when it observes grasping, biting, chewing or licking the food. Moreover, the F5 area also contains mirror neurons that are activated during acts of communication manifested in gestures and movements of the lips and tongue. This is important for social cognition because, as Gallese points out: “It is therefore plausible to propose that communicative mirror neurons might constitute a further instantiation of a simulation-based social heuristic” (Gallese, 2005, p. 34).

6. Embodied-embedded emotional mind

Proponents of applications of embodied simulation pay special attention to the role of emotion and empathy in social cognition. In the course of phylogeny, primary emotions mainly appeared in order to regulate the behaviour of individuals and to enhance the survival chances of their genes. Nevertheless, in socialized species such as Homo sapiens higher order emotions have evolved and they play an entirely new role. Joseph LeDoux indicates the difference between those two types of emotion:

> Basic emotions (fear, anger, joy, disgust, etc) are believed to be innately organized in humans, conserved to some degree across mammalian species, automatically or unconsciously elicited, expressed in characteristic ways in the body of all humans independent of culture, and mediated by distinct neural systems. Higher order emotions (e.g. empathy, jealousy, guilt) are often considered to be less conserved across species, cognitively mediated rather than automatically elicited, less rigidly expressed in the body, mainly due to learning and social factors, and can be different in different cultures (LeDoux, 2011).

In species with highly developed brains, emotions that are the holistic responses of the body to a stimulus may also have their internal aspects ("The bodily changes follow directly the perception of the exciting fact, an d that our feeling of the same changes as they occur is the emotion" [James, 1884, p. 189]). Typically, they are referred to as feelings (Damasio, 2010). Emotions may and may not be conscious. For example, we are sometimes unaware of the content of our emotions until somebody asks “why are you mad at me?”. Your partner can recognize your emotional state earlier than you and she makes it on the basis of the observation of states of the body, such as facial expressions.

Empathy, one of the higher order emotions, is a special case of compassion. According to Frans de Waal, it is one of the two main pillars of morality (the other being reciprocity and fairness) (de Waal, 2006, 2010, 2013). Michael Arbib and Jean-Marc Fellous argue that the mechanisms of empathy are associated with the activity of the mirror neuron system and embodied simulation:

> Clearly, human emotions are greatly influenced by our ability to empathize with the behavior of other people. Indeed, some have suggested that mirror neurons can contribute not only to ‘simulating’ other people’s actions as the basis for imitation, but also ‘simulating’ other people’s feelings as the basis for empathy (Arbib and Fellous, 2004).

The strong relationship between empathy and simulation is indicated by various studies on responses to pain. Single-cell recordings show that certain cells of the cingulate cortex fire when a person experiences a painful stimulus caused by a pinprick on herself as well as when she sees the other person suffering from a pinprick (Hutchison et al., 1999). The simulation of pain has also been confirmed by experiments with neuroimaging (fMRI). They both show that the first-person experience of pain and the observation of another person experiencing pain activate the anterior cingulate and insula (Singer et al., 2006).

It is widely known that the responses of the motor system not only accompany, but are embedded within the emotional states in which organisms find themselves. In this way Zajonc and Markus answer the persistent question of “why do people who are angry squint their eyes and scratch their shoulders?” (Zajonc and Markus, 1984). The face is the part of the body that reveals our emotions most expressively. Gallese, citing a study conducted by Carr et al. (2003) notes that an important argument for the relationship of embodied simulation and experiencing emotions are the data from neuroimaging studies using fMRI on people observing and imitating facial expressions specific for different emotions. These studies indicate that imitating emotions causes the activation of such brain structures as the ventral premotor cortex, insula and amygdala. The fact that those structures are involved in emotion processing is additionally confirmed by lesion studies. Damage to these areas results in a reduced ability to imitate and experience the emotions of others.

Gallese also refers to his own research with fMRI aiming to demonstrate a common neural architecture for emotions experienced from both the first-person and third-person perspective. His subject of interest was disgust, which is one of the basic emotions. The subjects' brains were scanned in two cases – when they inhaled an unpleasant scent arousing disgust and when they watched various faces expressing disgust. In both cases the left anterior insula was active (Wicker et al., 2003).

Another argument for the application of embodied simulation is the fact that the imitation of the facial expressions of others helps us to recognize their emotional state. In an experiment conducted by Niedenthal et al. subjects were asked to detect the offset of a sad facial expression changing into a happy expression and vice versa (Niedenthal et al., 2001). In doing this task one group of participants held a pen between their lips and teeth, which blocked the possibility of imitating the observed facial expressions. They found that people who were allowed to freely mimic the faces detected a change from a sad to a happy expression much faster than those who had fewer possibilities for imitation. In a more advanced exercise, this result was reaffirmed by Oberman, Winkielman and Ramachandran (2007).

Gallese points out, however, that the mere imitation of facial expressions does not always cause a conscious experience (feeling) associated with the imitated emotion (Gallese, 2005, p. 37). Nevertheless, the lack of a complete awareness of an imitated mental state does not necessarily mean that it is not utilized in the action or social cognition. This constitutes an argument for the thesis that some pre-reflective mechanism of simulation of state of the body which is common for a person experiencing the emotion and her observer/imitator works in the background and correlates with the neural structures mentioned above. This is an argument against the theory-theory – empathy does not involve any form of reasoning by analogy, but consists of entering into a direct resonance with another person. Piotr Winkielman, Paula M. Niedenthal, and Lindsay M. Oberman aptly sums up this thread:

> Theories of embodied cognition suggest that engagement of sensory-motor processes is part and parcel of the process of emotional perception, understanding, learning, and influence. On that account, the vicarious recreation of the other's state provides information about the stimulus meaning and can go beyond the previously established associations. If so, manipulation (inhibition or facilitation) of somatosensory resources should influence the perception and understanding of emotional stimuli. Evidence for this interpretation has been now obtained in multiple domains (Winkielman et al., 2009).

7. Conclusions: paradigms shifts again

At the beginning of this article we argued that in the last few decades there have been a few paradigm shifts in cognitive science research. It is argued that in the discussion between Lazarus and Zajonc on the primacy of cognition and emotion the view of the latter prevailed. It should be stressed, however, that relations between emotion and cognition are more complicated. We believe that the paradigm shift that has occurred thanks to the embodied-embedded mind approach consists of the adoption of the primacy of action. Both cognition and emotion – mutually connected on various levels of the complexity of brain and mind – are subordinated to generate a response that solves the problem facing the organism, and solves it in the most optimal way. This includes problems facing the organisms in real time, as well as problems in the evolutionary context. (One can assume that selection mechanisms favour the cognitive and affective systems being connected at different levels).

The second paradigm shift, which took place in the cognitive sciences in the 1980's, consisted of a transition from a computer-like vision of the mind to the embodied-embedded mind approach. One can claim that this vision, according to which the mind is software implanted in cerebral hardware, is close to the psychophysical dualism advanced by Descartes (Searle, 1992). In this interpretation "software" recalls *res cogitans*, while the "hardware" is similar to *res extensa*. Although computationism still plays an important role in some parts of cognitive science (e.g., in the thesis of Massive Mental Modularity accepted by evolutionary psychologists), the embodied--embedded mind approach introduces strongly non-dualistic vision. This vision is supported by various data sets from neuroscience and cognitive linguistics (Gallese and Lakoff, 2005; Lakoff and Johnson, 2009).

The resignation of a computational paradigm of cognitive science in favour of the embodied-embedded mind approach is connected with another (already mentioned) revolution. It consists of the purification of the 'mind' sciences of "Cartesian leftovers", which were still present in social cognition. Cartesianism assumed that we can only be sure of our own mental states and the cognition of other minds is secondary. Hence, we are involved in the other minds problem. In philosophy of mind this problem is solved by the idea of estimating the mental states of others by analogy to one's own mental states or by eliminating the entire realm of subjectivity. (The former account had been developed later in the theory-theory.) Embodied simulation breaks the scheme and seems to be much more credible solution supported by data from studies on the functioning of the brain.

Advocates of the embodied-embedded mind approach claim that the mind is created by the interactions with physical, social and cultural environment that the body enters. They assume that the mind of the individual can "resonate" with the minds of others. In brief, people are constantly reading the minds of others. The embodied simulation theory eliminates the difference between the subjectively received self and other minds. Moreover, according to the shared manifold hypothesis, the formation of the first-person perspective is secondary and subsequent to the prior intersubjectivity that our biological equipment provides. This is the genuine revolution, because one can say that in such a view the cognition of one's own mind seems to be at least as problematic as the old problem of other minds.

Embodied simulation allows us not only to predict the behaviour of others and feel empathy for them, but also to create an intersubjective community of language and culture. According to convincing studies conducted i.a. by Michael Tomasello and Merlin Donald, the human ability that underlies culture is the ability to imitate (Tomasello, 1999; Donald, 2005), which is based on a simulation mechanism. Imitation is defined as the precise replication of goals and the means which lead to their achievement and the precise replication of such patterns is possible thanks to the well-developed mirror neuron system. Thus, it seems concepts such as "cognition" and "emotion" and also "culture" and "nature" are not as contrasting as previously thought ("[…] imitation seems to be the key to understanding rule-following in particular, and the social practices which contribute to the development of culture in general. From the neuroscientific viewpoint it is the mirror neuron mechanism that lies at the heart of pattern-recognition and – indirectly – pattern-propagation. Evolutionary theory, on the other hand, suggests that we tend to imitate others due to our mutualism" [Brożek, 2013, p. 217]).

References

Arbib, M., and Fellous, J.-M. (2004). Emotions: From brain to robot. Trends in Cognitive Sciences *8(12)*, 558.

Baron-Cohen, S. (1997). Mindblindness: An Essay on Autism and Theory of Mind, chapter 4: Developing Mindreading: Four Steps (Boston: The MIT Press), pp. 31–58.

Brożek, B. (2013). Rule-Following. From Imitation to the Normative Mind (Kraków: Copernicus Center Press).

Carr, L., Iacoboni, M., Dubeau, M.C., Mazziotta, J.C., and Lenzi, G.L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. Proceedings of the National Academy of Sciences USA *100(9)*, 5497–5502.

Chemero, A. (2009). Radical Embodied Cognitive Science (Cambridge–London: The MIT Press).

Churchland, P.M. (1981). Eliminative materialism and the propositional attitudes. Journal of Philosophy *78*, 67–90.

Dacety, J., and Jackson, P.L. (2004). The functional architecture of human empathy. Behavioral and Cognitive Neuroscience Reviews *3*, 71–100.

Damasio, A. (2010). Self Comes to Mind. Constructing the Conscious Brain (New York: Pantheon Books).

Darby, D., and Walsh, K.W. (2005). Walsh's Neuropsychology: A Clinical Approach (Edinburgh: Elsevier).

de Waal, F. (2006). Primates and Philosophers. How Morality Evolved (Princeton–Oxford: Princeton University Press).

de Waal, F. (2010). The Age of Empathy (New York: Harmony Books).

de Waal, F. (2013). The Bonobo and the Atheist. In Search of Humanism Among the Primates (New York: W.W. Norton & Company).

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G. (1992). Understanding Motor Events: A Neurophysiological Study. Experimental Brain Research *91(1)*, 176–180.

Donald, M. (2005). Imitation and mimesis. In: Perspectives on Imitation, vol. II, S. Hurley and N. Chater, eds. (Cambridge, MA: The MIT Press), pp. 283–300.

Ferrari, P.K., Gallese, V., Rizzolatti, G., and Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. European Journal of Neuroscience *17*, 1703–1714.

Fogassi, L., and Gallese, V. (2002). The neural correlates of action understanding in non-human primates. In: Mirror Neurons and the Evolution of Brain and Language, M.I. Stamenov and V. Gallese, eds. (Amsterdam–Philadelphia: John Benjamins Publishing Company), pp. 13–35.

Gallagher, S. (2007). Phenomenological and experimental contributions to understanding embodied experience. In: Body, Language and Mind. Vol. 1, T. Ziemke, J. Zlatev, R. Frank, and R. Dirven, eds. (Berlin: Mouton de Gruyter), pp. 241–263.

Gallagher, S. (2008). Understanding others: Embodied social cognition. In: Handbook of Cognitive Science. An Embodied Approach, P. Calvo and T. Gomila, eds. (San Diego–Oxford–Amsterdam: Elsevier), pp. 439–452.

Gallese, V. (2005). 'Being like me': Self-other identity, mirror neurons, and empathy. In: Perspectives on Imitation, vol. 1: Mechanisms of Imitation and Imitation in Animals, S. Hurley and N. Chater, eds. (Cambridge, MA: The MIT Press), p. 106.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. Phenomenology and Cognitive Sciences *4*, 23–48.

Gallese, V. (2009). Mirror neurons and the neural exploitation hypothesis: From embodied simulation to social cognition. In: Mirror Neuron Systems, J.A. Pineda, ed. (New York: Humana Press).

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. Cognitive Neuropsychology *21*, 455–479.

Gärdenfors, P. (2006). How Homo Became Sapiens. On the Evolution of Thinking (Oxford: Oxford University Press).

Gazzaniga, M.S. (2009). Human: The Science Behind what Makes Us Unique (New York: Harper Perennial).

Hutchison, W.D., Davis, K.D., Lozano, A.M., Tasker, R.R., and Dostrovsky, J.O. (1999). Pain related neurons in the human cingulate cortex. Nature Neuroscience *2*, 403–405.

James, W. (1884). What is an emotion? Mind *9*, pp. 188–205.

Johnson, M. (2007). The Meaning of the Body. Aesthetics of Human Understanding (Chicago: The University of Chicago Press).

Kurek, Ł. (2013). Philosophical issues in the neuroscience of mindreading: The case of belief. In: Philosophy in Neuroscience, J. Stelmach, B. Brożek, and Ł. Kurek, eds. (Kraków: Copernicus Center Press), pp. 175–200.

Lakoff, G., and Johnson, M. (1999). Philosophy in the Flesh. The Embodied Mind and its Challenge to Western Though (New York: Basic Books).

Lazarus, R. (1984). On the primacy of cognition. American Psychologist *39(2)*, 124–129.

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

LeDoux, J. (2011). Emotional coloration of consciousness: How feelings come about. In: Philosophy in Science. Methods and Applications, B. Brożek, J. Mączka, and W.P. Grygiel, eds. (Kraków: Copernicus Center Press), pp. 190–191.

Meltoff, A. (2005). Imitation and other minds: The 'like me' hypoothesis. In: Perspectives on Imitation, vol. 2: Imitation, Human Development, and Culture, S. Hurley and N. Chater, eds. (Cambridge, MA: The MIT Press), pp. 55–77.

Niedenthal, P.M., Brauer, M., Halberstadt, J.B., and Innes-Ker, A. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. Cognition and Emotion *15*, 853–864.

Oberman, L.M., Winkielman, P., and Ramachandran, V.S. (2007). Face to face: Blocking expression-specific muscles can selectively impair recognition of emotional faces. Social Neuroscience *12*, 167–178.

Ramachandran, V.S. (2010). The Tell-Tale Brain, A Neuroscientist's Quest for What Makes Us Human (New York: W.W. Norton & Company).

Rizzolatti, G., Craighero, L., and Fadiga, L. (2002). The mirror system in humans. In: Mirror Neurons and the Evolution of Brain and Language. M.I. Stamenov and V. Gallese, eds. (Amsterdam–Philadelphia: John Benjamins Publishing Company), pp. 37–59.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettenardi, V., Paulesu, E., Perani, D., and Fazio, F. (1996). Localization of grasp representations in humans by positron emmision tomography: I. Observation versus execution. Experimental Brain Research *111(2)*, 246–252.

Searle, J.R. (1992). The Rediscovery of the Mind. (A Bradford Book).

Singer, T., Seymour, B., O'Doherty, J., Klaas, E.S., Dolan, J.D., and Frith, C. (2006). Empathic neural responses are modulated by the perceived fairness of others. Nature *439*, 466–469.

Tomasello, M. (1999). The Cultural Origins of Human Cognition (Cambridge: Harvard University Press).

Umiltà, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., and Rizzolatti, G. (2001). I know what you are doing. A neurophysiological study. Neuron *31*, 155–165.

Varela F., Thompson, E., and Rosch, E. (1993). The Embodied Mind. Cognitive Science and Human Experience (Cambridge: The MIT Press).

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. Neuron *40*, 655–664.

Winkielman, P., Niedenthal, P., and Oberman, L.M. (2009). Embodied perspective on emotion-cognition interactions. In: Mirror Neuron Systems, J. Pineda, ed. (New York: Humana Press), pp. 239–240.

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist *35(2)*, 151–175.

Zajonc, R., and Markus, H. (1984). Affect and cognition: The hard interface. In: Emotions, Cognition and Behavior, C. Izard, J. Kagan, and R.B. Zajonc, eds. (Cambridge: Cambridge University Press), p. 74.

Wojciech Załuski
Jagiellonian University
Copernicus Center for Interdisciplinary Studies

Emotions and Rationality[1]

1. The vagueness of the concept of the 'rationality of emotions'

The traditional way of conceptualizing relations between emotions and rationality is in terms of a conflict or tension. This is quite understandable given that one can provide numerous examples of situations in which emotions interfere with rational decision-making. For instance, emotions may affect the present behaviour by virtue of the associated arousal that makes one act against one's better judgment, i.e., emotions may bring about *akratic* behaviour (e.g., they may move the agent to engage in an extramarital love affair even though she judges marital infidelity to be a worse option than the option of remaining a faithful wife); emotions may affect behaviour by causing a drastic shortening of the time perspective (e.g., they may induce an agent to take revenge immediately when it would be more rational – because involving a smaller risk – to postpone it); or emotions may narrowly focus our attention only on one aspect of a situation making us forget about its other aspects – in other words: they may distort our judgment of a given situation (e.g., anger at a given person for the wrong

---

[1] The paper, in a slightly different form, was originally published as an appendix of W. Zaluski's book *Game Theory in Jurisprudence* (Copernicus Center Press, Krakow 2013)

action she did to us may make us forget about all her earlier benevolent actions toward us). Clearly, this is not an exhaustive list of possible conflicts between emotions and rationality. In the present article, however, we do not propose to deal with such conflicts but, rather, we propose to show that there may be a harmony between emotions and rationality, i.e., that emotions may also be rational. We shall not attempt at deciding a rather uninteresting (and arguably undecidable) question of whether emotions are more often irrational or rational; our intention is more modest: we shall attempt at clarifying the very concept of the 'rationality of emotions'.

At the start of our analysis let us note that when one can speaks about emotions one can mean either *types of emotions* or *token (occurrent) emotions of a given type*. Now, our basic claims to be developed and substantiated throughout this paper are the following. (1) The proper (in the sense: aptly reflecting our common practice of speaking about rationality of emotions and at the same time honing this practice) explication of rationality with regard to types of emotions is in terms of *adaptive rationality*. Thus, to say that a given type of emotion is rational means that it is a biological adaptation – a 'Darwinian algorithm', i.e., a computational process that is triggered by certain evolutionarily relevant inputs and yields certain outputs (behaviour). (2) The proper explication of rationality with regard to token emotions is in terms of *intrinsic* and *instrumental* rationality (we shall see, though, that these two notions of rationality can be used derivatively or secondarily also with regard to types of emotions). Now, a token emotion is *intrinsically rational* iff the elements of its structure satisfy certain conditions of rationality; and a token emotion of agent *A* is *instrumentally rational* iff it does not affect this agent's decision-making processes negatively. In the next part of this paper we shall make these two notions of rationality of emotions more precise and analyze the relations between them. In the final part, we shall return to the notion of adaptative rationality of emotions.

2. The intrinsic rationality of emotions

2.1. Token emotions

The above definition of intrinsic rationality of token emotions makes reference to the 'structure' of emotions. Therefore there arises a question about components of this structure. Following Robert Nozick, we assume that each emotion has three components: belief, evaluation, and feeling (Nozick, 1990). A belief may be assessed (depending on the assumed epistemological conception) as true or false, or justified or unjustified; an evaluation (on a scale from 'maximum badness' to 'maximum goodness') may be assessed as correct or incorrect; and feeling (psycho-physiological arousal) may be assessed as proportionate or disproportionate (to an evaluation). From the above it follows that a given emotion can be defective, inappropriate, i.e., *intrinsically irrational* (our, not Nozick's, term[2]), in three ways: belief may be false or unjustified, evaluation may be incorrect, and feeling may be disproportionate to evaluation (exceedingly intense or not sufficiently intense). The condition of rationality our definition of intrinsic rationality of emotions speaks about are therefore the following three: truthfulness (or justification), correctness, and proportionality. Nozick noticed that an agent whose emotion is appropriate (intrinsically rational) 'responds' to reality in a proper way, i.e., her response is consistent with the 'reality principle': such an agent has an adequate picture of reality encoded in her emotion and her emotion enables her to be closely connected to – 'cling to' – this reality (Nozick also stressed that such a response is itself a value, which he calls 'a second-order value'). It is worth noting that some emotions (e.g., aesthetic ones) can arguably be assessed only in terms of their intrinsic rationality because they are not accompanied by action tendencies, i.e., do not motivate an agent to engage in a certain behaviour, and thereby can-

---

[2] Also Elster (1999) uses this term and but in his definition of it he does not appeal to the three-component account of emotions.

not be assessed in terms of instrumental rationality. Let us give now some examples of intrinsically rational and irrational token emotions.

(1) When can, for example, compassion, be said to be intrinsically rational? Aristotle (in *Rhetoric*) defined compassion (ἔλεος) as a painful emotion directed at another person's misfortune or suffering. Now, for compassion to be intrinsically rational, three conditions must be fulfilled: (a) seriousness of suffering, (b) un-deservedness of suffering, (c) similarity (more precisely: (a) and (b) seem to be the conditions of the intrinsic rationality of compassion, whereas (c) seems to be an empirical condition of experiencing compassion). Thus, *B*'s compassion with *A*'s suffering is intrinsically rational iff *B*'s *belief* is: "*A* suffers", and the belief is true or justified, *B*'s *evaluation* is: "*A*'s suffering is serious and undeserved *and is a bad state of affairs to a degree x*", and the evaluation is correct, and *B* 's *feeling* is proportionate to *B*'s evaluation that the considered state of affairs is bad to a degree x.

(2) Emotions triggered by a truncated cognition, i.e., cognition focused "on a favorable or unfavorable outcome without any concern for its causal history (Elster, 1999, p. 312)" are intrinsically irrational because beliefs they are based on are incorrect, or, rather – in this case – incomplete. Jon Elster gives the following examples of such emotions: anger at those who frustrate our goals even though it was impossible for them to foresee the consequences of their actions; guilt for emotions, thoughts, or events outside our control; shame for emotions, thoughts, events, or character traits outside our control; pride for events outside our control; contempt for features of a target individual that are outside her control; admiration for features of a target individual that are outside her control[3].

---

[3] Other intrinsically irrational emotions mentioned by Elster, are, inter alia, hate to those we have harmed; hate to those who have helped us; hate to those who did not reciprocate our love; phobias without cognitive justification; love to those whom we have helped and flowing only from the fact that we have helped them.

2.2. Types of emotions

The category of ‘intrinsic irrationality’ can also be used, though secondarily, with regard to types of emotions. But in this case (unlike the case of token emotions) it is more apt to oppose it not to intrinsic rationality but to what may be called ‘indeterminacy in respect of intrinsic rationality’ (because it is difficult to find a type of emotion that would be intrinsically rational, i.e., such a type of emotion that all token emotions belonging to this type would be intrinsically rational). Now, a given type of emotion is intrinsically irrational if all token emotions that belong to this type of emotion are intrinsically irrational; and a given type of emotion is indeterminate in respect of intrinsic rationality if token emotions that belong to this type of emotion may be intrinsically irrational or intrinsically rational. Some additional comments are in order here. *First*, the assessment of the intrinsic rationality of a token emotion of a given type may be different in two different axiological systems even if these systems concur in assuming that this type of emotion is not intrinsically irrational (i.e., is indeterminate in respect of intrinsic rationality). For instance, for the ancient Greeks the following token emotions, which are regarded contemporarily as irrational, were not irrational: guilt for emotions, thoughts, or events outside our control; contempt for features of a target individual that are outside her control; admiration for features of a target individual that are outside her control. *Second*, some types of emotions (e.g., envy, conceit) appear to be regarded as intrinsically irrational on the grounds of all axiological systems (these emotions may, though, still be biologically adaptive). *Third*, some types of emotions are regarded as intrinsically irrational on the grounds of some axiological systems, and as indeterminate in respect of intrinsic rationality on the grounds of the other. For instance, according to some currents of the Christian ethics (e.g., Jansenism) pride (not only conceit, but also ‘moderate’ pride) is always irrational because it is based on an incorrect value judgment of oneself – it is a manifestation of one’s failure to understand the fundamental misery of the human condition (a different evaluation of pride was assumed, e.g., in the Homeric ethics). According to, for example, Spinoza, La Rochefoucauld, and Nietzsche, compassion is always irrational; a different evaluation of compassion is made, for example, within the Christian, Buddhist, or Islamic ethics. *Fourth*, on the grounds of the Stoic ethical system, no state of affairs that "does not fully depend on us" has a value (this Stoic tenet was especially emphasized by Epictetus), and thereby almost all types of emotions are intrinsically irrational (because almost all types of emotions are based on beliefs and evaluations regarding events that do not fully depend on us).

3. Instrumental rationality of emotions

One can distinguish three general views of role of emotions in decision-making. According to the *traditional view*, one chooses rationally only if one does not choose under the influence of the emotions (emotions are viewed here as blind, causal forces interfering with rational choice). According to what one may call the *revisionist view*, one chooses rationally only if one chooses under the influence of the emotions. According to the *'compromise' view* emotions sometimes facilitate and sometimes hinder rational choice. The 'compromise' view seems most plausible: it is clear that emotions sometimes hinder rational choice but it can also be plausibly argued that emotions may facilitate rational choice; the following considerations can be viewed as developing this view.

Let us start by distinguishing between two ways in which emotions may affect the choice: (a) emotions may affect the parameters of choice, i.e., they may enter the choice as costs and benefits associated with the available options; they may therefore change the structure of payoffs; (b) emotions may affect the very mechanisms of decision-making. Having distinguished these two ways, we can propose the following definition of the instrumental rationality of an emotion: an agent's emotion is *instrumentally rational* iff it affects the parameters of choice in a rational way (*Condition 1*) and does not prevent the agent from choosing the option that maximizes her expected utility (*Condition 2*). In the remainder of this point we shall analyze relations between intrinsic and instrumental rationality of emotions; especially, we shall reflect on whether only intrinsically rational emotions can be simultaneously instrumentally rational or also intrinsically irrational emotions can be instrumentally rational.

In order to analyze the role of intrinsically rational and intrinsically irrational emotions in decision-making one has to make a distinction between two types of decision problems: decision-problems related to a given intrinsically rational or irrational emotion and decision-problems unrelated to a given intrinsically rational or irrational emotion.

(1) *Decision-problems related to a given intrinsically rational or irrational emotion*.

In such decision problems intrinsically rational emotions always (by definition) satisfy *Condition 1* of instrumental rationality (i.e., they modify their structure of payoffs in a rational way). But these emotions may fail to satisfy *Condition 2* because psycho-physiological arousal caused by them may be an obstacle for choosing the utility-maximizing option. By contrast, in this type of decision problems intrinsically irrational emotions always (by definition) fail to satisfy *Condition 1* of instrumental rationality. They may also fail to satisfy *Condition 2* because psycho-physiological arousal caused by them may be an obstacle for choosing the utility-maximizing option; this type of influence of emotions will, of course, practically matter (i.e., may lead to a failure in choosing a rational – utility-maximizing – option) if the rational option in a decision problem with the improperly determined payoffs (i.e., determined under the influence of intrinsically irrational emotions) is the same as in an analogous decision problem with the properly determined payoffs (i.e., determined under the influence of intrinsically rational emotions). To make our distinction between properly and improperly determined payoffs clearer, we shall illustrate it with a game-theoretic example. Consider the following four games (in the tables below the underlined results are Nash equilibria in pure strategies):

Player 1/Player 2	*C*	*D*
C	6, 6	4, 6
D	3, 2	3, 2

Figure 1. Game 1

Player 1/Player 2	*C*	*D*
C	6, 6	4, 6
D	3 + 4, 2	3 + 4, 2

Figure 2. Game 2

Player 1/ Player 2	*C*	*D*
C	6, 6	4, 6
D	3 + 6, 2	3 + 6, 2

Figure 3. Game 3

Player 1/ Player 2	*C*	*D*
C	6, 6	4, 6
D	3 + 1, 2	3 + 1, 2

Figure 4. Game 4

The initial game is Game 1. In this game each player has two strategies: Cooperation (*C*) and Defection (*D*). In this game Player 2 plays *D* and thereby elicits Player 1's anger (because his choice can be seen as a manifestation of malice: by choosing *D* rather than *C* Player 2 does not increase his own payoffs but decreases Player 1's payoffs). Now, Player 1's anger can be intrinsically rational or intrinsically irrational. Assume that intrinsically rational anger's strength is 4 and that the anger manifests itself in increasing the value of playing *D* (and thereby causing a loss to Player 2) to an 'angry' Player 1. In Game 2 the strength of this anger is 4, in Game 3 it is 6, and in Game 4 it is 1. Thus, Player 1'a anger is intrinsically rational in Game 2 and intrinsically irrational in Game 3 and Game 4. As we can see, Game 3 has the same Nash equilibria as Game 2; thus, intrinsically irrational emotions (in our example: anger) may generate the same results (i.e., the same combinations of strategies as Nash equilibria) as intrinsically rational emotions. But, of course, intrinsically irrational emotions may lead to different results than those generated by rational emotions (Game 4 has different Nash equilibria than Game 2).

(2) *Decision-problems unrelated to a given intrinsically rational or irrational emotion*.

In such problems both intrinsically rational and intrinsically irrational emotions always (trivially) satisfy *Condition 1* of instrumental rationality (i.e., they do not modify at all the structure of payoffs of these decision problems). But these emotions may fail to satisfy *Condition 2* because the psycho-physiological arousal caused by them may be an obstacle for choosing the utility-maximizing option.

The above analysis shows that intrinsically rational emotions may fail to be instrumentally rational (though, arguably, they usually are instrumentally rational), and that intrinsically irrational emotions may lead to rational decisions (even though they modify the structure of payoffs in an improper way).

At the end of this point, by way of digression, one may notice that Antonio Damasio's famous 'somatic markers hypothesis' can be interpreted as providing a precise (though controversial) description of the way in which emotions affect the very mechanisms of decision-making (Damasio, 1994). According to Damasio, the basic role of emotions consists in eliminating the danger of overly long deliberations that may hinder making a choice. When choosing between any range of possible options, one cannot consider all those options and their consequences in detail but, in order to make a decision, one must eliminate some of these options *a limine.* This task, according to the 'somatic markers hypothesis', is realized by emotions: somatic markers – which are a sort of gut feelings – force our attention onto negative outcomes of available options and serve as warning signals that enable the brain to reject the options generating these outcomes and to choose only from among the remaining options. One can therefore say that emotions enhance the agent's ability to take account of long-term consequences by acting as current cognitive signals. As is well known, Damasio's hypothesis is based on his neurobiological research which has shown that patients who sustained injuries to the prefrontal and somatosensory cortices of the brain have a diminished capacity to experience emotions *and* to make decisions. Damasio's (controversial) interpretation of this research is that there is a *causal link* between the inability to experience emotions and the inability to make decisions.

4. Adaptive rationality of emotions

The first claim made in point 1 says that types of emotions can be assessed in terms of adaptive rationality, i.e., in terms of their fitness-maximizing character in the ancestral environments in which they had emerged[4]. We shall now develop the notion of adaptive rationality of emotions at somewhat greater length.

It bears emphasizing that adaptive rationality is in fact a special case of instrumental rationality. Let us recall a *general definition of the instrumental rationality of emotions* formulated in point 3: An agent's emotion is *instrumentally rational* iff it affects the parameters of choice in a rational way (C*ondition 1*) and does not prevent the agent from choosing the option that maximizes her expected utility (C*ondition 2*). Now we can formulate a variant of this definition which is relevant for the discussions about the evolutionary origins of emotions: A given type of emotion is *evolutionarily instrumentally rational* (i.e., adaptively rational) iff it tends to affect the parameters of choice and the very processes of decision making in a way that had tended to lead to fitness-maximizing decisions in the ancestral environments. An interesting question that arises in the context of discussion of the adaptive rationality of emotions is *whether there are any types of emotions that are not adaptations*. It seems that even for types of emotions that feature among *peccata capitalia (gula*, *lussuria*, *avaritia*, *invidia*, *ira*, *acedia*, *superbia*), i.e., emotions that might seems especially immoral and/or irrational

---

[4] An in-depth discussion of the problem of the evolutionary origins of emotions can be found, for example, in LeDoux 1999.

one can find a plausible evolutionary explanation (with the possible exception of *acedia*). In other words, it seems that for virtually all type of emotions one can find a plausible evolutionary justification, i.e., one can consider them to be adaptively rational. In the remaining part of this point we would like to present two insightful hypotheses regarding the evolutionary origins of a certain type of emotions (the so-called 'moral emotions').

According to the first hypothesis, *moral emotions evolved because they were efficacious in implementing the Tit-for-Tat strategy in an iterated version of the Prisoner's Dilemma* (Trivers, 1971; Axelrod, 1981; Axelrod, 1984; Axelrod and Hamilton, 1981). Such emotions as gratitude, kindness, forgiveness, and anger can serve as proxies for the Tit-for-Tat strategy: agents who act on these emotions act *as if* they played the Tit-for-Tat strategy. Therefore these emotions are especially effective in supporting the relationships of reciprocal altruism. According to the second hypothesis, *moral emotions evolved because they helped overcome impulse-control problems (i.e., the temptations to choose options serving our short-term rather than long-term interests) and commitment problems (i.e., problems arising "when it is in a person's interest to make a binding commitment to behave in a way that will later seem contrary to self-interest (Frank 1988, p. 47)")* (Schelling, 1978; Hirshleifer, 1987; Frank, 1988)[5]. As for impulse-control problems: moral emotions help solve them by providing an agent with additional incentives to choose options that serve her long-term self-interest; they therefore enhance the agent's ability to take account of long-term consequences by acting as current motivational proxies. Let us present this insight in a more precise way. In the case of *an agent A with no moral emotions*, at time *t* (*the moment of making a choice*), the expected utility of option *x* serving *A*'s short-term interest is greater than the expected utility of option *y* serving *A*'s long-term interest.

---

[5] It should be noticed, however, non-moral emotions are one of the main causes of impulsive, akratic behaviour.

But in the case of *an agent A with moral emotions*, at time *t (the moment of making a choice)*, the sum of expected utility of option *x* serving *A*'s short-term interest and disutility caused by moral emotions activated by the choice of *x* is smaller than the expected utility of option *y* serving *A*'s long-term interest. Moral emotions can also provide an advantage in strategic interactions, i.e., they help solve commitment problems. For instance, a player's reputation for being inclined to experience moral emotions (e.g., guilt, shame, moral indignation) is likely to make her promises and threats credible and thereby to serve her long-term interests. Thus, for example, a player's *A* disposition to have the feelings of guilt if she plays non-cooperatively is likely to make her promise to player *P* to play cooperatively credible and as a consequence is likely to induce *P* to play cooperatively in a game with *A*. Let us illustrate this last point with the game of agency. In this game there are two players: an agent and a principal. An agent can increase (by 500%) the amount of money possessed by the principal (say, 100 zlotys). An agent promises to the principal that he will give him back one-half of the increased sum (i.e., 250 zlotys). The principal may trust the agent (and give him the money) or not trust him (and not give him the money). Assuming that agent is *homo oeconomicus* (and thereby is not prone to experiencing feelings of guilt) the result of the game (marked by a bold line and established by means of reasoning called in game-theoretic parlance 'backward induction) will be (100, 0): the principal will not trust the agent's promise. But if the agent is prone to feelings of guilt and the principal is aware of this fact, then the result of the game (marked by a dotted line and established by means of 'backward induction') will be (250, 250) (it is assumed that guilt generates costs to the agent which is equivalent to the loss of 300 zlotys).

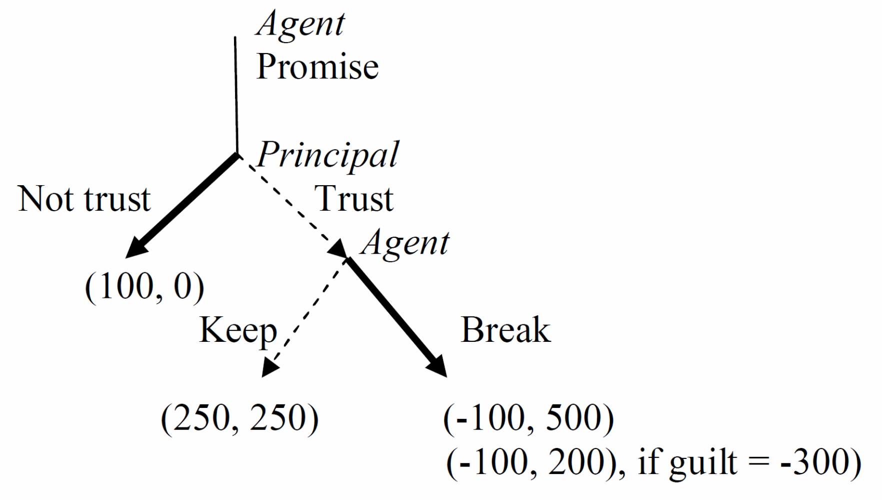


Fig. 5. Game of agency

Thus, the result of the game with the agent prone to experience feelings of guilt is a Pareto-improvement on the result of the game in which the agent is *homo oeconomicus*.[6] It is therefore in the material interest of the agent to feel moral emotions, and thereby *not to be homo oeconomicus*, i.e., not to be narrowly focused on pursuing his material interest.

5. Summary

In order to explicate the concept of the 'rationality of emotions' we have proposed a conceptual framework based on the following pairs of distinctions between various types of rationality: adaptive rationality vs. adaptive irrationality; intrinsic irrationality vs. indeterminacy in respect of intrinsic rationality; intrinsic rationality vs. intrinsic irrationality; and instrumental rationality vs. instrumental irrationality.

---

[6] To give another example: "…a person who is known to "dislike" an unfair bargain can credibly threaten to walk away from one, even when it is in her narrow interest to accept it. By virtue of being known to have this preference she becomes a more effective negotiator (Frank 1988, p. 5)".

We have argued that adaptive rationality (irrationality) is a special case of instrumental rationality (irrationality). Our main conclusions can be summarized in two points. First, types of emotions can be assessed by means of the following two pairs of distinctions: adaptive rationality vs. adaptive irrationality (though, we have suggested that it is difficult to provide an example of a type of emotion for which one could not construct a plausible evolutionary justification, i.e., which could not be plausibly viewed as adaptively rational), and intrinsic irrationality vs. indeterminacy in respect of intrinsic rationality. Second, token emotions can be assessed in terms of the other two pairs of distinctions: intrinsic rationality vs. intrinsic irrationality; and instrumental rationality vs. instrumental irrationality.

References

Axelrod, R. (1981). The emergence of cooperation among egoists. American Political Review *75 (2)*, 306–318.

Axelrod, R. (1984). The Evolution of Cooperation (New York: Basic Books).

Axelrod, R., and Hamilton W.D. (1981). The evolution of cooperation in biological systems. Science *211*, 1390–1396.

Damasio, A. (1994). Descartes' Error (New York: Grosset Putnam).

Elster, J. (1999). Alchemies of the Mind. Rationality and the Emotions (Cambridge: Cambridge University Press).

Elster, J. (2000). Ulysses Unbound. Studies in Rationality, Precommitment and Constraints (Cambridge: Cambridge University Press).

Frank, R. (1988). Passions within Reason: The Strategic Role of the Emotions (New York, London: W.W. Norton and Company).

Frank, R. (1992). Emotion and the costs of altruism. In: The Sense of Justice: Biological Foundations of Law, R.D. Masters and M. Gruter, eds. (Newbury Park, London, New Delhi: Sage Publications), pp. 47–48.

Hirshleifer, J. (1987). On the emotions as guarantors of threats and promises. In: Latest on the Best: Essays on Evolution and Optimality, J. Dupré, ed. (Cambridge MA: The MIT Press).

LeDoux, J. (1999). The Emotional Brain (New York: Phoenix Paperbacks).

Nozick, R. (1990). The Examined Life: Philosophical Meditations (New York: Simon and Schuster).

Schelling, T. (1978). Altruism, meanness, and other potentially strategic behaviours. American Economic Review *68 (1)*, 229–230.

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

Bartosz Brożek
Jagiellonian University
Copernicus Center for Interdisciplinary Studies

Morality, Law and Emotions

The goal of this essay is to question the existence of moral (and legal) emotions. I begin by outlining the theory of law and morals proposed by Leon Petrażycki at the beginning of the 20th century. Petrażycki saw emotions as central to our normative practices and attempted to define law and morality as two separate mental phenomena. After highlighting some of the weaknesses of his account, I present and criticize Jonathan Haidt's view of moral emotions, one which posits the existence of four families of emotions that are the guardians of our moral practices. I suggest that Haidt's conception is too limited, as it fails to recognize that our emotional mechanisms condition the existence of the entire culture, not only its morals. At the end of the essay, I argue for a broader conception, which sees emotions such as shame, guilt or anger as instrumental in any collaborative social interaction.

1. Petrażycki's legacy

Leon Petrażycki is considered one of the most eminent – if not the preeminent – Polish legal theorist, but also one of the first Polish psychologists. In a number of works written between 1899 and 1908 (Petrażycki, 2002; Petrażycki, 1959; Petrażycki, 1960), he developed an intriguing theory of law and morality. More precisely, he set out to lay new foundations for jurisprudence and claimed that this task requires, first and foremost, an answer to the question of *what is law*. He observed that without such a definition any legal-philosophical and legal-theoretic considerations are carried out in a vacuum: "This is a principal and prejudicial issue, one which conditions the very possibility of the science of law" (Petrażycki, 1959, p. 41).

The answer to the question 'what is law' demands, according to Petrażycki, the application of an adequate method. The problem is, however, that the methodological tools Petrażycki used were outdated, if not anachronistic – even from the perspective of the philosophy of science of the 19th century. Petrażycki claims that the goal of jurisprudence – understood as "a science in the correct sense of the word" (Petrażycki, 1959, p. 13) – is to look for the *essence of law* (Petrażycki, 1959, p. 39). In effect, one should try to construct a classical definition of law, i.e. a definition *per genus et differentiam: "*In order to divide the law scientifically into kinds and determine the differences between them, one should know the *genus* of law (…). It is thus necessary to recognize to which higher, more general category of phenomena law belongs" (Petrażycki, 1959, pp. 31, 35). Here, in a nutshell, one can find all the ingredients of the Aristotelian view of science. It assumes, first of all, that in the world there exist *essences*; the aim of science is, thus, to capture those essences in definitions which serve to build a table of essential definitions (Porphyry's tree), one that classifies all entities unequivocally. Petrażycki's method presupposes, then, a very strong metaphysical view. It also encapsulates a static conception of science. Both those features are inconsistent with the practice of contemporary science.

At the same time, it is noteworthy that some of Petrażycki's claims open the door to a revision of the Aristotelian orthodoxy. He says, for example: "the foundation of scientific legal policy should be the examination of the causal features and mechanisms of law in general, and of its different kinds and elements in particular" (Petrażycki, 1959, p. 14). This passage indicates that jurisprudence should confine itself to the considerations of causal connections, while in the Aristotelian tradition the teleological connections – particularly in relation to law and morality – play an eminent role. Moreover, Petrażycki also notes:

> Especially in science, where each and every theory needs to overcome attempts at rejection and modification in order to be considered acceptable (…), where one has to do with a 'struggle for life' and only the fittest doctrines survive, one should expect that with the passage of time there should survive objectively sound theories (Petrażycki, 1959, pp. 205–206).

Such a declaration fits well with the conceptions of Popper or Lakatos. This is evidence that Petrażycki had many original insights connected with the question of what is science. Unfortunately, he combined them with the anachronistic ideas of Aristotle and this led, in turn, to the development of an incoherent methodology. It would not be too devastating if the elements of the essentialist ideology remained only at the surface level, but Petrażycki applied those methodological rules meticulously, making his theory of law – one conceived in original sin – an unacceptable one.

Jurisprudence for Petrażycki, together with other subjects in the humanities and the social sciences, needs a foundation which is to be found in a more basic science: psychology. Petrażycki was unhappy with the psychology of his time, in particular disagreeing with the Kantian heritage it accepted. Kant divided mental phenomena into three categories: sensations, feelings and will. Petrażycki considered this division incomplete, claiming that one should add to it a fourth category:

> One should distinguish not three, but four basic forms of inner experiences and four classes of mental elements: (1) emotions, i.e. impulsions (two-sided mental experiences), 2) and 3) sensations and feelings (passive one-sided experiences), 4) processes of the will (one-sided active experiences) (Petrażycki, 2002, p. 14).

He also provides the following description of the relationship between the four classes:

> Precisely speaking, one should distinguish: (1) emotions, as a foundation and the basic form of the (sensitive-impulsive) mental life; and (2) sensations, feelings and will, as the posterior products of the differentiation of prior emotions (sensations and feelings as differentiated ramifications of the sensitive side of emotions, and will – of their impulsive side) (…) (Petrażycki, 2002, p. 15).

This passage, underscoring the primacy of emotions in human mental life, is quite striking, given it was written in 1899 or 1900.

Petrażycki claims that emotions are two-sided, sensitive (in the sense that they are generated by experiencing something) and impulsive (in the sense that they have motivational force), while sensations, feelings and will are one-sided – purely sensitive (passive) in the case of the former two, and exclusively impulsive (active) in the case of the latter. Petrażycki stresses further that the failure to see the existence and role of emotions had been the major reason behind the futility of 19th century psychology: "mental life is two-sided, passive-active, sensitive-impulsive, (…) and as such cannot be explained in terms of uniquely passive or uniquely active 'basic elements'" (Petrażycki, 2002, pp. 17–18).

Furthermore, Petrażycki provides us with a classification of emotions. From our perspective, the most interesting are the ethical emotions (emotions of duty), which "are experienced as an inner limitation of freedom" (Petrażycki, 2002, p. 27). Ethical emotions could be further divided into moral and legal; the former are exclusively imperative (i.e., the actor feels she should do something), while the latter are imperative-attributive (the actor feels she should do something, but also accepts that someone else may expect her to do it). This distinction leads Petrażycki to his famous definition of law: "Law, as a separate class of real phenomena, should be understood as such mental experiences whose emotions are of the attributive character (…).

All the other ethical experiences, i.e. experiences of exclusively imperative emotions, should be deemed moral phenomena" (Petrażycki, 1960, pp. 72–73, 123). One would be mistaken, however, if they insisted that Petrażycki identifies legal norms with certain emotions. In order to clarify this issue, one should distinguish between such notions as emotion, representation, norm and duty.

Emotion, as noted above, is one of the four basic mental experiences. People, according to Petrażycki, also have the power to imagine certain situations or behaviour. Such an imagined representation – together with the emotion it causes – generates *a motive* for action. In *On the motives for action and on the essence of morality and law* Petrażycki says: "We are interested, especially, in one particular type of motivation, the one in which there is a connection of representations of various acts with very peculiar emotions, which we deem ethical emotions or emotions of duty" (Petrażycki, 2002, p. 25). For instance, if I imagined taking part in a fraud, "I would experience a mental state similar to that which I experience while considering eating a piece of rotten meat, touching a spider or a snake; in normal circumstances I would experience repulsive emotions" (Petrażycki, 2002, p. 21).

Furthermore, one should distinguish between emotions, representations and motives on the one hand, and legal and moral norms and duties on the other. Petrażycki defines legal (moral) norms as the *content* of the ethical (moral or legal) convictions (Petrażycki, 2002, p. 33). Thus, norms are not emotions or motives but they are of an intellectual character. They may be described as specific representations or propositions, which can be grasped (contemplated). It is worth noting that 'representation' (or 'proposition') serves as a primitive term in Petrażycki's psychology, similar to the notion of *idea* in Hobbes, Descartes or Locke, or the notion of *concept* in Kant.

The notion of duty is defined in a similar way. Petrażycki says: "Duties are ideal projections, which originate in our minds. Such projections are connected to the described emotions and representations, and not to some things or phenomena in the outer world" (Petrażycki, 2002, p. 34). It seems, therefore, that duties should also be called representations or propositions, which are graspable by the human mind. One can find the confirmation of those conceptual distinctions in the following passages:

> The explained difference of *genus* between one-sided imperative (moral) and two-sided duty-imposing (legal) norms and duties is based on the adequate *genus* differences among the emotional-intellectual complex phenomena which are, as we demonstrated, the real base for ethical duties and norms (Petrażycki, 2002, p. 49).

> The basic motivation, which consists in connecting representations of actions with the above characterized repulsive or impulsive emotions, we should deem ethical motivation and the corresponding principles of behaviour – ethical principles or norms (Petrażycki, 2002, p. 28).

Therefore, law and morality exist in the minds of people. Law cannot be identified with norms. It would be a mistake, however, to identify it with certain ethical emotions. When Petrażycki says that "law, as a separate class of real phenomena should be understood as such mental experiences whose emotions are of the attributive character", he claims that law is a *complex* mental phenomenon, one that consists of adequate emotions, norms and duties.

Still, one more problem should be addressed: what is the relationship between the law (in Petrażycki's sense) and the provisions of legal acts? Petrażycki's reply is the following:

> The representations of legal provisions or biblical commandments shall be deemed the representations of "normative facts". Ethical convictions, to which such representations belong, shall be called positive ethical convictions, and their contents – positive norms. Ethical convictions, which lack such representations of normative facts, are intuitive ethical convictions, and the corresponding norms – intuitive norms (Petrażycki, 2002, p. 33).

The general mental mechanism proposed by Petrażycki appears as follows: people have the capacity to imagine certain situations, patterns of behaviour etc. There are a plethora of sources of such imagined representations: legal acts, the Bible, or any other ‘normative fact’, as well as one’s own intuition. Those representations cause the corresponding legal or moral emotion, and together they serve as motives for action. In light of the above, one may say that Petrażycki presents us with a peculiar ontology of law and morality. He believes that norms are certain representations or propositions and claims, moreover, that law (or morality) cannot be identified with the set of legal (or moral) norms. Law (morality) consists of complex mental states, which include representations (propositions), and emotions, together generating motives for action.

Petrażycki’s insights were quite novel – if not unprecedented – at the turn of the 20th century. He clearly stressed the fundamental role of emotions in human (and, more generally, animal) mental life and opposed those who saw will as the major motivational force, replacing it with the motive-generating power of emotions. Furthermore, his insistence on the role of sensitive-impulsive emotions in our legal and moral decision-making stands in marked contrast to previous attempts at founding morality (and law) on the feelings of pleasure and displeasure. In fact, Petrażycki explicitly rejects such an approach, as when he says:

> According to the received view, the motives of human action can always be traced back to pleasure or displeasure (…): it is the pursuit of pleasure and happiness, and the avoidance of suffering, that form the bedrock of all human behaviour (…) According to my view, our actions are never guided by those factors; their motives are always emotions (…). Regarding the representation of pleasure resulting from some contemplated, possible action (…), it plays no role whatsoever in the motivational process, save when it causes an emotional impulse. (…) Only in those cases, when appulsive (attractive) emotions are ge-

> nerated, a drive towards pleasure exists. But even under such circumstances, the impulse for action is not the imagined hedonistic goal, but the appulsive (attractive) emotion (Petrażycki, 2002, pp. 17–19).

Petrażycki's dismissal of pleasure and displeasure as the driving forces behind human action has been confirmed by the subsequent developments in psychology and, later, in cognitive science. It does not mean, however, that Petrażycki's theory does not face serious objections. First, he uses an anachronistic methodology (Aristotle's essentialism), which carries with it serious metaphysical baggage. By accepting it, Petrażycki is forced to look for the *essence* of law (morality), an ephemeral entity that is nowhere to be found. Moreover, misled by Aristotelianism, Petrażycki looks for a foundational answer to the question 'what is law' (morality). This foundationalism has two faces: firstly, it requires us to look for a science that is more basic than jurisprudence (moral science); secondly, it launches a search for some basic phenomena which, taken together, 'produce' law (morality).

But Petrażycki's account of law and morality is troublesome also regarding more specific problems. A case in point is his claim that among 'ethical' emotions one can sharply distinguish moral and legal ones: the former are imperative, while the latter – imperative-attributive. In this way, one can identify the proper subject-matter of both disciplines – law and morality – and it turns out that the domains in question are disjointed: the legal and moral sciences deal with different phenomena. This stance is problematic for a number of reasons. First, law and morality clearly overlap, as many moral rules are *at the same time* legal rules. Second, there exist seemingly imperative-attributive emotions (such as gratitude or guilt), which are highly relevant for morality. Third, it is not always clear whether a given emotion is merely imperative, or imperative-attributive. Let us consider sympathy: there are cases in which we sympathize with someone, who – say – has had bad luck in his life, but we do not feel that there is some action they are entitled to expect of us; however, there are also situations – such as when we witness a car accident – in which our sympathy leads to the belief that the injured person may rightly expect some action from us. This shows that the distinction between merely imperative and imperative-attributive emotions is not a sharp one. Fourth and finally, the very definition of ethical (legal and moral) emotions is troublesome. It suggests that both moral and legal emotions are experienced as "inner limitations of freedom", which lead to the recognition that one *should* do something. This is problematic, since the emotions in question are defined with the use of a *normative* concept ("should"), which *prima facie* belongs to some other cognitive *niveau*. It is as if Petrażycki moved too easily from facts (the experience of a certain emotion) to norms.

Altogether, both his definition of ethical emotions, as well as the criterion for distinguishing moral and legal emotions, seem unacceptable. It does not mean, however, that the idea of identifying a cluster of emotions 'responsible' for our moral life is ill-stated. Indeed, the category of moral emotions is a subject of debate in contemporary psychology; in what follows, I will present and criticize one of the more important contributions to the debate.

2. Are there moral emotions?

In his seminal paper "The Moral Emotions" (Haidt, 2003a), Jonathan Haidt makes an attempt to identify a class of emotions which are the key to understanding human moral behaviour. He begins with a historical note, recalling that neither classical philosophy nor psychology considered emotions as central to morality; it was only the 'affective revolution' of the 1980s and 1990s that placed emotions among the building-blocks of human moral life[1].

The key question Haidt tries to answer is whether it is possible to precisely define moral emotions. He adopts a preliminary definition, according to which moral emotions are those "that are linked to

---

[1] Which is not fully accurate, given Petrażycki's conception described above.

the interests or welfare either of society as a whole or at least of persons other than the judge or agent" (Haidt, 2003a, p. 853). To provide a more precise characterisation, he claims that there are several features of each and every emotion: an eliciting element, facial expression, physiological change, phenomenological experience, and motivation (action tendency). One can distinguish moral from other kinds of emotions by examining two features: (1) moral emotions have *disinterested elicitors* (i.e., they are brought about by situations which do not affect self-interest, as when one sees a photograph of suffering people); and (2) are connected to *pro-social action tendencies* (i.e., they are likely to lead to actions which benefit others, not oneself).

Haidt realizes that there are no emotions which, always and under any circumstances, can be characterized by disinterested elicitors and pro-social action tendencies. For example, anger – which may be triggered by seeing someone being beaten and lead us to help them, and so constitute a prototype of a moral emotion – may also be a result of sexual frustration and generate no pro-social behaviour. More generally, Haidt believes that 'prototypical' instantiations of various emotions can be placed on a diagram, in which the *x*-axis represents the degree of the disinterestedness of elicitors, while the *y*-axis – the level of pro-sociality of the action tendencies associated with the given emotion (see Fig. 1). The emotions placed near the up-

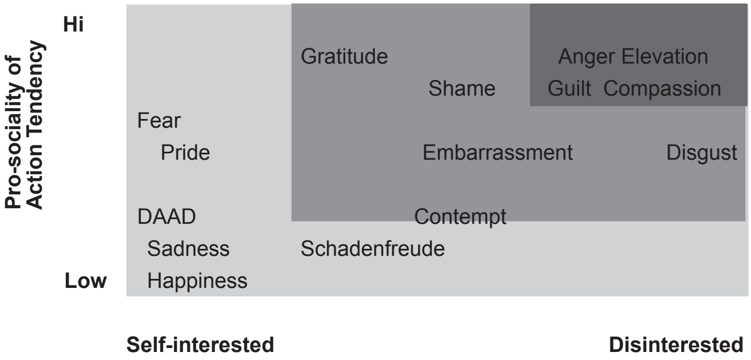


Figure 1. Disinterestedness of Elicitors

per-right corner of the diagram (anger, elevation, guilt, compassion, shame, embarrassment, disgust, gratitude, contempt) are characterized by a high degree of disinterestedness and a high level of pro-sociality, and hence may be deemed moral emotions; those placed near the lower-left corner (sadness, happiness, fear, pride, *Schadenfreude*) are non-moral emotions.

Next, Haidt suggests dividing moral emotions into four families. The two *large* families are the 'other-condemning' emotions (contempt, anger, disgust) and the 'self-conscious' emotions (shame, embarrassment, guilt). The two *small* families are the 'other-suffering' emotions (compassion) and the 'other-praising' emotions (gratitude, elevation). Further, Haidt stresses three aspects of moral emotions. First, they all have simpler, but non-moral forms, which are present in other animals and our evolutionary ancestors. Second, all the mentioned emotions are *panhuman*, but there exist cultural differences in how they are shaped. For example, it is assumed that most Asian cultures do not distinguish between shame and embarrassment. Third, Haidt's classification shows that there is more to morality than altruism and niceness; indeed, our moral experience would be impossible without 'the guardians of morality' (contempt, anger, disgust) or 'the guardians of moral self' (shame, embarrassment, guilt).

It is interesting to compare Haidt's approach with Petrażycki's. Given the hundred year distance between the two theories, the similarities between them are striking. In particular, both Petrażycki and Haidt underline the special role of emotions as the driving force behind our moral life, although – of course – the details of their accounts are different. Unlike Petrażycki, Haidt does not claim that moral emotions can be *sharply* defined: it is rather a matter-of-degree distinction. Moreover, Haidt stresses that some instantiations of typically moral emotions may have neither disinterested elicitors nor pro-social action tendencies. These differences notwithstanding, both theories point to the emotional roots of morality, thus placing themselves on the same side of the barricade in the controversy over the role of reason and emotions in moral decision-making.

In this context, Haidt offers his account of moral judgement. In his famous essay "The Emotional Dog And Its Rational Tail" (Haidt, 2001), he argues that one should reject the rationalistic paradigm in moral psychology, i.e. an approach "that moral knowledge and moral judgement are reached primarily by a process of reasoning and reflection" (Haidt, 2001). Instead, he proposes the *Social Intuitionist Model*, based on the thesis that moral judgement is a result of a quick moral intuition, which is sometimes – albeit not always – followed by relatively slower *ex post* reasoning. Haidt explains that by 'moral judgement' he understands "evaluations (good versus bad) of the actions or character of a person that are made with respect to a set of virtues held by a culture or subculture to be obligatory" (Haidt, 2001). 'Moral reasoning', on the other hand, is "a conscious mental activity that consists of transforming given information about people in order to reach a moral judgement" (Haidt, 2001). Finally, moral intuition constitutes "the sudden appearance in consciousness of a moral judgement, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion" (Haidt, 2001).

Haidt's model is composed of six theses (Haidt, 2001). First, he claims that moral judgement always appears in consciousness in an automatic, effortless way, as a result of the workings of moral intuition. Second, he believes that moral reasoning is a process tied up with effort and takes place usually after the decision has been reached, supporting it only *ex post*. Third, the goal of the *ex post* arguments and communicating them is to justify intuitive moral judgements to others. Fourth, moral judgements uttered by others (in particular: family members, acquaintances or friends) exercise a significant influence on moral decision-making. Fifth, Haidt admits that it may occasionally transpire that moral reasoning leads to a moral judgement; however, it is possible only in situations, in which the initial intuition is weak, and the circumstances allow for the carrying out of time- and energy-consuming reasoning processes. Sixth, and finally, during the reasoning process a new intuition may be activated, one incompatible with the initial intuition.

Haitd cites a number of neuroscientific, psychological and primatological findings which support his model. In particular, he highlights the role of somatic markers. The somatic marker hypothesis, formulated by Antonio Damasio, posits that in a normally functioning brain, when a complex decision process must be undertaken, somatic states are activated which 'categorize' possible courses of action as good or bad (Damasio et al., 1991). Importantly, the responses provided by the somatic markers are learned through experience, lead to quick, unconscious, but quite accurate intuitive decisions and do not preclude the use of reasoning to decide in the given situation; rather, their output precedes any conscious deliberations (Damasio et al., 1991).

Damasio's hypothesis constitutes a perfect complement to Haidt's claim that moral intuition is a result of the functioning of an automatic, unconscious set of 'moral concepts'. On the one hand, any somatic marker is a mechanism based on emotional responses. The process of the unconscious categorization of behaviour as good or bad is connected to the functioning of the ventromedial prefrontal cortex, which mediates between the regions of brain responsible for emotional processing and the areas that control action. This explains why damage to the ventromedial prefrontal cortex results in patients whose emotional states have ceased to play a role in influencing their behaviour (Damasio et al., 1991). On the other hand, 'moral concepts' are formed in the course of social interactions. According to Haidt, the development of moral intuition is tied up with a number of processes (Haidt, 2001). In this context, a special role is played by the unconscious mechanism of cultural transmission. Numerous experiments and observations suggest that only a relatively small part of our cultural knowledge is learned consciously – children acquire it mainly through the imitation of the behaviour of older children and adults. Haidt points out that the recent findings regarding the neural foundations of intuition underscore the importance of practice and repetition for the proper training of cultural intuitions. He claims further that the process of social learning depends on the activity of the basal ganglia's circuits, which are also instrumental to motor learning: because of that many social skills are rapid and automatic, just like well-learned motor sequences are. Social skills and judgement processes, learned in a gradual and implicit way, are experienced in the consciousness as arising from nowhere (Haidt, 2001). "The implication of these findings for moral psychology is that moral intuitions are developed and shaped as children behave, imitate, and otherwise take part in the practices and custom complexes of their culture. Participation in custom complexes in this way provides a cultural 'front end' for Damasio's somatic marker hypothesis, and for Lakoff's embodied cognition. Even though people in all cultures have more or less the same bodies, they have different embodiments, and therefore end up with different minds" (Haidt, 2001, p. 835).

Haidt's theory has been the subject of many critiques. For instance, it has been pointed out that quick and automatic moral intuitions are usually 'educated' by prior reasoning; in addition, in cases of genuine moral dilemmas people tend to rely on deep reflection (Pizarro and Bloom, 2003). It has also been observed that although the processes leading to the formation of a moral judgement are usually irrational, it does not necessarily mean that one should consider the judgement itself irrational, as it may be subject to discussion (Levy, 2006). Finally, it has been claimed that people make deliberative corrections to their moral judgements more often than Haidt seems to suggest (Fine, 2006). Irrespective of whether those critical remarks are sound (Haidt, 2003b), there is little doubt that a significant part of our morally relevant actions are automatic, and only rationalized *ex post*. Moreover, they are not 'innate' reactions, as although they are certainly based on some biologically determined tendencies, in most cases they are learned, repeated patterns of conduct.

However, exactly this aspect of Haidt's theory underscores an important weakness of his approach. The role of emotions which Haidt calls 'moral' (i.e. those which are characterized by disinterested elicitors and pro-social action tendencies) is *not* limited to providing a foundation for our moral behaviour; rather, the emotions in question are important in all social interactions, and as such are instrumental in bringing about the *entire* culture, not only its moral dimension. In order to substantiate this thesis, let us inspect in detail the evolutionary scenario recently proposed by Michael Tomasello.

3. Emotions and cultural transmission

Michael Tomasello, in his two books, *The Cultural Origins of Human Cognition* (Tomasello, 1999) and *Why We Cooperate* (Tomasello, 2009), sketches an evolutionary scenario in which the ability to imitate plays a pivotal role in the creation and preservation of culture. He begins with an observation that the genetic difference between the human species and other animals is not so significant (approx. 1–1,2%). This constitutes an argument for the thesis that the biological adaptation enabling the flourishing of human culture must be relatively 'small'. In other words, Tomasello claims that it is impossible to account for the development of various aspects of culture (language, morality, science, etc.) by recourse to a large number of biological adaptations. In particular, he believes that

> the 6 million years that separates human beings from other great apes is a very short time evolutionarily, with modern humans and chimpanzees sharing something on the order of 99 percent of their genetic material – the same degree of relatedness as that of other sister *genera* such as lions and tigers, horses and zebras, and rats and mice. Our problem is thus one of time. The fact is, there simply has not been enough time for normal processes of biological evolution involving genetic variation and natural selection to have created, one by one, each of the cognitive skills necessary for modern humans to invent and maintain complex tool-use industries and technologies, complex forms of symbolic communication and representation, and complex social organizations and institutions (Tomasello, 1999, p. 2).

Tomasello claims that "there is only one possible solution to this puzzle. That is, there is only one known biological mechanism that could bring about these kinds of changes in behaviour and cognition in so short a time (…). This biological mechanism is social or cultural transmission, which works on time scales many orders of magnitude faster than those of organic evolution" (Tomasello, 1999, p. 4). This, in turn, is made possible by three forms of learning: imitative, instructed and collaborative. This observation suggests that the *ability to imitate* is one of the most crucial adaptations in the evolutionary history of humankind. However, the mere ability to imitate is not sufficient to account for the enormous cultural advantage of humans over other species, since the imitation skills of apes and monkeys are also considerable. The question, therefore, is where does the human *tendency* to imitate come from.

Tomasello's answer singles out the ability to cooperate. In *Why We Cooperate* he addresses the problem of the human capacity to act together, and stresses that our cooperation has two peculiar features: it leads to cumulative cultural evolution and is connected with the tendency to imitate others within our social group, to 'be like them':

> To an unprecedented degree, *homo sapiens* are adapted for acting and thinking cooperatively in cultural groups, and indeed all of humans' most impressive cognitive achievements – from complex technologies to linguistic and mathematical symbols to intricate social institutions – are the products not of individuals acting alone, but of individuals interacting (Tomasello, 2009, pp. XV–XVI).

Tomasello's analysis is carried out in two stages. Firstly, he considers our tendency for altruistic behaviour, and asks whether it emerges spontaneously (and so has purely biological foundations), or rather the development of altruistic propensities is reinforced by culture. He stresses that altruism is not a single adaptation, but rather a bundle of adaptations, in particular in such areas of human activity as helping others, sharing with them and informing them (Tomasello, 2009).

According to Tomasello, the fact that people tend to help each other is deeply rooted in our biology. The innate character of helping is backed by five arguments. Firstly, the tendency to help emerges early (14-18 months) in ontogeny. Secondly, many experiments in developmental psychology clearly show that the system of parental punishments and rewards does not strengthen, but rather weakens the willingness to help. For example: in the experiments of Tomasello and Warnecke, children who had been rewarded for helping others were less eager to do so in the second stage of the experiment than children who had not been rewarded during the initial stage. Thirdly, helping others is not a uniquely human trait: chimpanzees are also inclined to help, although to a lesser degree than humans. Fourthly, children raised in traditional cultures, with little contact with adults, have a similar tendency to help as children raised in the Western culture. This observation leads to the thesis that the form of culture, and in particular the scope and intensity of parental care, does not influence the tendency to help. Fifthly, helping is strictly connected to biologically conditioned empathic abilities. Thus, it seems that the tendency and willingness to help others, already present in small children, does not depend on culture so much as on our genetic underpinning[2].

Humans, in contrast to other primates, are also altruistic in regard to sharing with others and informing them. Primatologists agree, for instance, that apes are not given to sharing food, while even small children do so regularly. Children are also ready to share information in order to help others to complete a task (e.g., they say where an object was hidden which an adult is looking for). Importantly, this ability does not depend on acquiring linguistic skills – even nine-month-old children help in this way, while other primates do not. In particular, it is argued that alarm calls – e.g. in the presence of a predator – do not serve as a warning to others as such calls are made even in situations when other members of the group spot the danger (Tomasello, 2009).

---

[2] A different opinion is defended by C.S. Dwecke in a comment in Tomasello, 2009, pp. 125–136.

Tomasello claims that the three above described forms of altruism are inscribed into our biological constitution. They are a *default program* with which we are born[3]. It does not mean, of course, that toddlers are pure altruists: in many contexts they behave egoistically. However, in comparison with other primates, human innate altruism is extensive and non-discriminatory: children have the tendency to help, inform and share with anyone, irrespective of whether they know the person or not. This is the case up to the point when they reach their third year, when they begin to display reciprocal altruism, as well as being influenced by social norms (Tomasello, 2009).

Tomasello underscores that the ability to be guided by norms is typically human. In this context he recalls the famous experiment with capuchin monkeys which was interpreted as showing that the monkeys possess a kind of primitive sense of fairness. In the experiment, a monkey was given cucumber, which she accepted. However, when another monkey was simultaneously offered grapes (a food preferred by the capuchins), the first monkey rejected the cucumber. It was speculated that the reason for this behaviour is that the monkey considered such food distribution 'unfair'. However, Tomasello claims that there is no social comparison at work here. Experiments carried out at several laboratories confirmed that both capuchins and chimpanzees reject cucumber only because the presence of the preferred food (grapes) makes the cucumber far less attractive (even if no other monkeys are present). Similar conclusions are warranted by another experiment of Tomasello's, in which a version of the ultimatum game for chimpanzees was designed. The rules of the game are simple: two players are given a certain sum, say $100. The task of one of the players is to propose a division of the sum between both, while the other may accept or reject it. If accepted, the players receive their shares according to the proposed division; if the proposal is rejected, they receive nothing. Experiments show that humans do not act in a

---

[3] In a comment, Dwecke claims that the scope of altruistic behaviour in children depends on social interactions to a much higher degree than suggested by Tomasello.

purely rational way in the ultimatum game setting: they are prone to accept only such divisions which seem ‘fair’ (75–25% at worst), when any division which gives the second player some money (even 1%) should rationally be accepted. It turns out that apes – under analogous conditions – behave much more rationally, accepting also the ‘unfair’ distributions (Tomasello, 2009).

Tomasello claims further that humans, already at the age of 3–4, follow two kinds of norms: the norms of cooperation and the norms of conformity. The first type includes, but is not limited to, moral norms; these are all the rules strengthening and coordinating the collaborative activities. The second type – the norms of conformity – do not influence cooperation directly, but they constitute ‘social identity’; a norm of this sort is, for example, the pure conventional rule calling for the hanging of coats on a rack:

> Importantly, children do not just follow norms as they encounter them, but in new situations they actively seek out what they are supposed to do – what the social norms and rules are in the situation – so that they can behave accordingly. On their first day in a new classroom, for example, children want to know what they are supposed to do with their coats. When they learn that we hang our coats on the rack before sitting down at our desks each morning, they understand this as the way ‘things are done’ here, and they want to do it this way too (Tomasello, 2009, p. 35).

According to classical Piagetan theory, there are two reasons for children following various types of social norms: the authority of adults and the reciprocity in relation with coequals. Tomasello acknowledges the role of those factors, but observes that experiments also point to other sources of rule-complying behaviour. In particular, even very young children (3-4 years-old) actively participate in the process of enforcing norms (e.g., within the framework of various games organized by adults). Tomasello believes that in order to explain this fact one needs to postulate that already small children have the capacity for *shared intentionality*, i.e. they adopt the group-perspective (*we* are doing something), and not the individual perspective (*I* am doing something) (Tomasello, 2009). It is the group-perspective, in addition to authority and reciprocity, that enables the full development of human altruistic tendencies. The development is strengthened by emotions of guilt and shame, characteristic of humans, which 'presuppose' some kind of social norms or, at least, social judgement. To sum up:

> The development of altruistic tendencies in young children is clearly shaped by socialization. They arrive at the process with a predisposition for helpfulness and cooperation. But then they learn to be selective about whom to help, inform, and share with, and they also learn to manage the impression they make on others – their public reputation and self – as a way of influencing the actions of those others toward themselves. In addition, they learn the social norms that characterize the cultural world in which they live, and they actively attempt to learn what these are and to follow them. They even begin to participate in the enforcement process by reminding others of the norms (…) and punishing themselves through guilt and shame when they do not live up to them. All of this reflects not only humans' special sensitivity to social pressure of various kinds, but also a kind of group identity and social rationality that is inherent in all activities involving a shared, 'we' intentionality (Tomasello, 2009, pp. 43–44).

Having presented the *form* of human altruistic behaviour, Tomasello inquires as to its *sources*. He puts forward a controversial thesis that altruism is only a manifestation of our tendency to cooperate, which results from a more basic biological adaptation: *mutualism* (Tomasello, 2009). He observes that cooperation among apes is usually based on kinship or reciprocity. The evolutionary passage from these two forms of behaviour to human cooperative capacities was conditioned by three processes: the development of social cognition and motivational mechanisms, enabling the coordination of cooperative activities and complex communication; the increase of tolerance and trust in relation to others, primarily in the context of acquiring food; and the development of group institutional practices, based on social norms[4] (Tomasello, 2009).

As stressed by Tomasello, children become capable of cooperative activities, which crucially depend on the ability to react to the actions of others, quite quickly and form joint goals and divide roles within a common undertaking. The experiments show that very young children (14-20-months-old) begin to understand the joint goal of an activity; for instance, when an adult, with no apparent reason, withdraws from a game, they try to persuade them to rejoin the play. Tomasello speculates that this ability constitutes the first step towards the emergence of joint attention, common knowledge, intersubjectivity, the ability of complex communication, etc. "Human cooperative communication thus evolved first within the bounds of collaborative activities because these activities provided the needed common ground for establishing joint topics, and because they generated the cooperative motives" (Tomasello, 2009, p. 73). Tomasello believes, further, that the key role in this process was played by joint efforts in gathering food (Tomasello, 2009).

Mutualism also requires an appropriate attitude towards others: tolerance and trust. Tomasello underscores the fact that animals display neither of them, as clearly illustrated with food-sharing. Experiments and observations show that animals, even if hunting together, do not share the prey in a 'fair' manner. For example, chimpanzees engage in hunts for red colobus monkeys. Usually, it is the chimpanzee who makes the kill that becomes the 'possessor' of meat; however, the second biggest share does not go to the chimpanzees that helped most, but to those that beg and harass the possessor most. Humans tend to be much 'fairer' when it comes to sharing food. For example: in an experiment in which one of a group of children was

---

[4] J. Silk claims that altruism is prior to mutualism; see her comment in Tomasello, 2009., pp. 111–124.

given all of the sweets, the protests of the others led to a fairer share (Tomasello, 2009). Thus, one may say that evolution has equipped humans with unique emotional and motivational mechanisms which promote tolerance and trust. There exist various evolutionary scenarios pertaining to the emergence of those emotions and motivations. According to one conception, tolerance and trust have been selected because of the advantage they give in the context of gathering food (those individuals that were able to hunt together and exercise the fair sharing of food had an evolutionary advantage over more egotistical individuals). A competing theory has it that it was due to the egalitarian character of human hunter-gatherer groups that more aggressive and egotistical individuals were ostracized, and hence their survival chances diminished. Yet another factor that could have contributed to the development of tolerance and trust was the fact of human cooperative childcare, as it is quite surprising that when in other species it is the mother that is almost uniquely responsible for childcare, in human societies her share is around 50% (Tomasello, 2009). Irrespective of which of those scenarios is the most persuasive,

> the important point is simply that there was some initial step in human evolution away from great apes, involving the emotional and motivational side of experience, that propelled humans into a new adaptive space in which complex skills and motivations for collaborative activities and shared intentionality could be selected (Tomasello, 2009, p. 85).

Finally, mutualism manifests itself in various human practices based on social norms. As I remarked above, Tomasello distinguishes between norms of cooperation and of conformity. He speculates that the norms of cooperation evolved within the context of mutually beneficial behaviour. However, mere mutual benefit does not seem to be sufficient to ground general or context-independent norms of cooperation. To this end, an understanding of the roles in joint activities is necessary, which in turn requires the ability to perceive the situation from a common perspective (*we*-intentionality) and to form joint goals. The norms of conformity, in turn, must have evolved under the pressure to conform with the group and, in particular, through the imitation of the behaviour of others. Such norms are constitutive of intra-group homogeneity and enable cultural evolution. Tomasello also underscores the essential role of the ability for symbolic communication (language) in the process of the emergence of both the norms of cooperation and of conformity[5] (Tomasello, 2009).

In sum, Tomasello sketches an intriguing picture of the development of culture, one which is conditioned by human mutualism. Mutualism requires specific motivational mechanisms, and so Tomasello emphasises its emotional aspect: in order 'to do things together' one needs special motivation, which is generated by such emotions as shame or guilt. In other words, Haidt's moral emotions turn out to be not only 'guardians of morality', but, more generally, 'guardians of culture'. Shame, guilt, embarrassment and gratitude, as well as contempt, anger and distrust play an eminent role in cementing social identity: the former motivate us to conform to social norms and mimic others, the latter lead to punishing those who are rule-violators. Importantly, the norms in question are not exclusively moral – they are all kinds of norms constituting culture. Thus, ultimately, the emotions identified by Haidt are as moral as they are mathematical. Of course, their indispensability is more clearly visible in the context of moral judgement; however, without them, there would be no language, physics or mathematics.

4. Against 'emotional foundationalism'

The above considerations may serve as a background for rethinking the relationship between emotions and morality (or law). The existing

---

[5] A different opinion is expressed by E.S. Spelke in her comment in Tomasello, 2009, pp. 159–173.

accounts of moral emotions are *foundational*. For instance, Petrażycki believed that there exist ontologically separate classes or moral and legal emotions. Only by identifying them may one find oneself in a position to define morality and law in a methodologically correct way. To put it differently: Petrażycki claimed that moral (and legal) emotions constitute the foundations for the practice of morality (and law). This foundational structure should be mimicked by our *theories* of morality and law, which are to be embedded in our knowledge concerning the relevant psychological mechanisms: it is psychology that provides the foundations for moral and legal sciences.

Haidt's approach is also foundational, although the foundations in question are not as sharply defined as in the case of Petrażycki's theory. Haidt points out that one can identify moral emotions through their characteristic features (disinterested elicitors, pro-social action tendencies), but underlines that only paradigmatic instantiations of relevant emotions do indeed have those features. It does not change the fact that it is the emotional mechanism that constitutes the bedrock of our moral practices. It is clearly visible once one considers Haidt's conception of moral judgement, in which intuitive, emotionally-conditioned, unconscious decisions are the backbone of morality. This leads Haidt to play down the role of reasoning in our moral behaviour.

In all fairness, Haidt's account is not purely foundational. He stresses that moral emotions are – to a certain extent – 'informed' by culture (e.g., in East Asian societies there is no strict distinction between shame and embarrassment). This coincides with his claim that our 'intuitive', unconscious moral decisions are largely trained: they are 'applications' of socially shared patterns of conduct. Thus, morality is not exclusively a result of the functioning of our emotional selves, as culture plays an active role in shaping our moral practices. However, on this view, emotions remain the foundation of moral action – they are like a musical score that may be played in various ways, depending on some particular cultural constraints; they are the 'moral invariance' between different cultures, explaining why there are so many panhuman moral traits.

I posit that this is not the only way of looking at things. Let us consider the problem of distinguishing between moral and conventional rules. Experimental research pertaining to this issue dates back to the 1970s, when Elliot Turiel carried out a series of experiments in the hope of answering the question of whether the distinction is purely conceptual, or if it has some biological underpinning (Turiel, 1983). He proposed characterizing moral rules as:

- having an objective prescriptive force, i.e. being independent of the authority of a person or an institution;
- holding universally, i.e. being applicable at any place and time;
- their violations typically involve a victim who has been harmed, whose rights have been violated, or who has been subject to an injustice;
- their violations are typically more serious than violations of conventional rules.

Conventional rules, on the other hand, are:

- arbitrary, i.e. they depend on some authority;
- often local, i.e. they are ‘binding’ only at a particular time and place;
- their violations do not involve a victim who has been harmed;
- their violations are typically less serious than violations of moral rules.

These assumed differences between moral and conventional rules provided the theoretical background for developing questionnaires, where – after presenting a scenario – the following questions were posed: (a) in the described circumstances, has any rule been violated? (b) was the violation of the rule wrong (and to what degree)?; (c) does the violated rule hold universally?; and (d) is the rule justified by the willingness to avoid harm, violation of rights or the requirements of justice? The results obtained by Turiel show that, indeed, people react in different ways to the violations of moral and conventional rules. Thus, one can speak of two different signature pattern responses: the violated rules are either considered universally binding, independent of authority, and protecting from harm; or local, dependent on authority and having nothing to do with the requirements of justice. Interestingly, Turiel noted that the distinction is manifested in the behaviour of 3.5-4-year-old children, but not in that of psychopaths. Thus, he concluded that distinguishing between moral and conventional rules is pancultural in character and develops early in ontogeny. Turiel claimed that it is a result of early cultural training, but others have suggested it has genetic sources, i.e., is innate.

In the past 20 years, a number of experiments have raised some doubts concerning Turiel's research. For instance, Haidt showed that in the USA and Brazil, the violations of some rules which are clearly conventional, as they are connected to no harm, are treated as serious offences (e.g., the use of the national flag to clean the toilet in the privacy of one's home) (Haidt et al., 1993). Similar observations are reported by Nisan, who shows that in Arabic countries addressing teachers by their first names is considered a violation of moral norms (Nisan, 1987). Thus, it seems that under specific circumstances violations of rules which are connected to no harm may be treated as serious. On the other hand, other experiments have shown that there are actions causing harm to others, but regarded as violations of rules, which depend on authority and are only locally binding (Kelly et al., 2007). Therefore, there is strong evidence that the distinction between moral and conventional rules is not a sharp one, let alone innate.

There are two interpretations of these observations *vis a vis* the debate over the existence of moral emotions. The first, which I hazard would be Haidt's opinion, is that our morality is based on emotions, but culture may modify *some* of our behavioural responses (e.g., giving rise to 'moral response' to the violation of a purely conventional rule, which bans speaking to teachers on a first-name basis). This remains a *foundational* view of the role of emotions in morality, positing the existence of an *emotional core* of moral behaviour, only altered by cultural training.

The second interpretation would be that emotions and culture are so intertwined that there is no *emotional core* of morality; rather, emotions play an eminent role in the development of culture *tout court*, and our cultural tools, such as language, morality or conventions, decisively influence our perception of emotions and the way they operate. On this view, the distinction between conventional and moral rules may be rendered in various ways, depending on the cultural context. Similarly, the classification of emotions may differ from culture to culture. Haidt stresses that all of his ‘moral emotions’ have simpler, non-moral evolutionary predecessors. For instance, he cites Fessler’s work on shame, in which two forms of the emotion are distinguished: a phylogenetically earlier and simpler proto-shame, usually caused by the presence of one’s superiors, and a “more cognitively complex form of shame that is triggered by violating a norm and knowing that someone else knows about the violation” (Haidt, 2003a, p. 859). It clearly shows that the route from protoshame to shame leads through culture, and it is not surprising that different cultures (e.g., Western as opposed to East Asian) ‘see’ shame in different ways. At the same time, it is no wonder there are so many moral similarities between different cultures, since they all have started on the same rung of the evolutionary ladder, being equipped with the same set of simple emotions.

One may be justified in asking whether there is any *real* difference between the two sketched conceptions: one positing the existence of the *emotional core* of morality, shaped differently by various cultures, and the other dispensing with this notion and suggesting a strict interplay between emotions and culture in bringing about our moral practices. At the phenomenological level there seems to be no significant difference: the fact, that the distinction between moral and conventional rules is not sharp, is explainable within both approaches. Similarly, that East Asian cultures seem not to embrace two separate emotions, shame and guilt, while Western societies clearly distinguish between the two, may be either a result of culture merely shaping the ‘emotional foundations of morality’, or of a deeper interplay of emotions and cultural factors. However, once we consider the problem from other perspectives, the difference becomes much more important.

First, at the meta-theoretical level, the *emotion-culture interplay* account seems quite different from the *emotional core* account: the former is much more comprehensive than the latter, underlining the role of emotions in our various cognitive skills. Therefore, it leads to a more *coherent* view of the functioning of the human mind. Second, both theories seem to be essentially different *heuristic tools*: while the *emotional core* account suggests concentrating on moral practices, and looking for biologically conditioned mental mechanisms which enable them, as well as the ways in which ‘cultural overlay’ modifies ‘moral instincts’, the *emotion-culture interplay* account sees morality as embedded in a larger class of cultural phenomena. This may lead to diametrically different ways of setting experiments and conducting observations, as well as interpreting them. For instance, the question concerning the co-evolution of language and morality, as seen from the *emotional core* perspective, may generate questions pertaining to the role of language in the development of moral practices, while considered from the *interplay* position, leads to the problem of the common evolutionary precursors of *both* language and morality. Third, there is a philosophical or methodological difference: the *emotional core* account is clearly driven by *reductionism*, i.e. by looking for what stands behind or constitutes the building blocks of morality; the *interplay* theory, on the other hand, sees moral practice as intimately linked with other social practices, and thus predicts a failure of any attempt to reduce morality to simpler phenomena.

To sum up: it is not my claim that there are *no moral emotions* in the sense of motivational mechanisms behind our moral practices. One cannot deny that morality is possible in large part due to the evolution of relevant emotional mechanisms. However, I believe that concentrating on moral emotions as pertaining exclusively – or at least primarily – to moral judgement and action may lead us astray, both in science and philosophy. Emotions *are* the guardians of morality, but not only: they guard morals because they guard the larger realm of culture.

References

Damasio, A.R., Tranel, D., and Damasio, H. (1991). Somatic markers and the guidance of behaviour: theory and preliminary testing. In: Frontal Lobe Function and Dysfunction, H.S. Levin, H.M. Eisenberg, and A.L. Benton, eds. (New York: Oxford University Press), pp. 217–229.

Fine, C. (2006). Is the emotional dog wagging its rational tail, or chasing it? Unleashing reason in Haidt's social intuitionist model of moral judgment. Philosophical Explanations *9*, 83–98.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgement. Psychological Review *108*, 814–834.

Haidt, J. (2003a). The moral emotions. In: Handbook of Affective Sciences, R.J. Davidson, K.R. Scherer, and H.H. Goldsmith, eds. (Oxford: Oxford University Press), pp. 852–870.

Haidt, J. (2003b). The emotional dog does learn new tricks: a reply to Pizarro and Bloom. Philosophical Review *110*, 197–198.

Haidt, J., Koller, S., and Dias, M. (1993). Affect, culture and morality, or is it wrong to eat your dog? Journal of Personality and Social Psychology *65*.

Kelly, D., Stich, S., Haley, K.J., Eng, S.J., and Fessler, D.M.T. (2007). Harm, affect and the moral/conventional distinction. Mind and Language *22(2)*.

Levy, N. (2006). The wisdom of the pack. Philosophical Explanations *9*, 99–103.

Nisan, M. (1987). Moral norms and social conventions. A cross-cultural comparision. Developmental Psychology *23*.

Petrażycki, L. (1959). Wstęp do nauki o prawie i moralności (An Introduction to the Science of Law and Morality) (Warszawa: PWN).

Petrażycki, L. (1960). Teoria państwa i prawa (Theory of State and Law), vol. I and II (Warszawa: PWN).

Petrażycki, L. (2002). O pobudkach postępowania i o istocie moralności i prawa (On the Motives for Action and on the Essence of Morality and Law) (Warszawa: Oficyna Naukowa).

Pizarro, D.A., and Bloom, P. (2003). The intelligence of the moral intuitions: A reply to Haidt. Psychological Review *110*, 193–196.

Tomasello, M. (1999). The Cultural Origins of Human Cognition (Cambridge, Mass.: Harvard University Press).

Tomasello, M. (2009). Why We Cooperate (Cambridge, Mass.: MIT Press).

Turiel, E. (1983). The Development of Social Knowledge (Cambridge: Cambridge University Press).

Dominika Dudek

Madness or Brain Disorders. From Dualism towards Integration

1. Historical views on "mad" people

In the past, there was no such term as mental illness. The expression that was used for a very long time in order to describe this condition was 'madness' and its treatment was often indistinguishable from torture or murder. The attitude towards mentally ill people derived from fear – fear of something incomprehensible, bizarre and horrible. Both madness and the fear of madness were not understood in any scientific manner until the 20th century. It worth noting that understanding biological causes of mental disorders and neurobiology of emotions have always gone hand in hand.

What is interesting is that some behavioral abnormalities, related to human psychopathology have been found in non-human great apes. As Brune et al. stated, many captive great apes show gross behavioural abnormalities such as stereotypes, self-mutilation, inappropriate aggression, fear or withdrawal, which impede attempts to integrate these animals in existing or new social groups. These abnormal behaviours resemble symptoms associated with psychiatric disorders in humans such as depression, anxiety disorders, eating disorders, and post-traumatic stress disorder (Brüne et al., 2006). Monkeys have a complex social system, and they form relationships with one another on an individual basis. They suffer from stress, much like we do, and often it seems to relate to social problems. Primates, and in particular great apes, are candidates for highly developed capabilities for empathy and theories of mind and it seems that emotions play a basic role in social behavior not only in humans, but also in primates.

It is, however, very difficult to judge the existence or nature of mental disorder prior to written records due to the very limited evidence. Although mental disorders are often considered to be caused by changes in the human environment, especially by the move from a natural environment towards modern conditions, evolutionary psychology suggests that some of the underlying genetic dispositions, psychological mechanisms and social demands have always been present. What is equally important, evolutionary psychology also explains why natural selection has not eliminated the genetic variation that results in diseased behavior. Generally, by giving a solid framework for understanding how behaviors are regulated to accomplish many of the conflicting tasks of life (e.g. getting food, surviving, finding mates, protecting children) (Nesse, 2005), evolutionary psychology gives an interesting perspective on the causes of mental disorders.

There is another reason which makes scientists believe that madness has a history as long as mankind. It is the evidence from Neolithic times of the practice of trepanning. There may have been many reasons for performing a procedure which involved cutting large holes into the skull, one suggesting that the purpose was to release a demon. This magical-medical approach has been observed in some primitive peoples yet trepanation might have also been an attempt to cure ailments including mental disorders, which would prove the thesis of evolutionary psychologists.

Although generally, “mad” people were treated as “possessed by devils”, it is also important to mention that in some primitive or ancient societies bizarre people could be respected as prophets or shamans. In ancient Israel, it was widely believed that mental or emotional disturbances were caused by supernatural forces or an angry God as a punishment for sin or a failure to follow the commandments. On the other hand, such disturbances were sometimes believed to provide one with a special power. The Old Testament contains numerous references to kings suffering from disturbances that could be attributed to various forms of madness. The Jewish prophets were thought to be psychologically abnormal because they acted in strange ways. They also departed significantly from the norm in appearance. However, the fact that they foretold future events that few understood was both proof of their abnormality and a reason to be respected. Apart from the Israelites, Socrates himself considered there to be some of the positive aspects of "madness". He included as such: prophesying (a 'manic art'), mystical initiations and rituals, poetic inspiration; and the madness of lovers.

Nonetheless, in the main part of human history, the social position of insane peoplehas been similar to that of the leprous. People with mental disorders were isolated, imprisoned, chained to walls and tortured. "Deviants" or "mad people" were confided to hospitals and asylums, where they stayed for many years, often until the end of their lives. Asylums were famous for their prison-like conditions. Active treatment included restrictive cribs, hunger cures, bloodletting, cold water dunking and other painful and ineffective methods (Shorter, 1997).

Such an attitude cannot be explained by the "darkness of the Middle Ages" – it is enough to mention Nazi Germany, where the institutionalized mentally ill were among the earliest targets of sterilization campaigns and covert "euthanasia" programs. The so-called "euthanasia" program entailed the killing of patients with the use of gas and lethal injections. The former was used in special hospitals between 1939 and 1941 and the latter somewhat later, between 1942 and 1945, in psychiatric hospitals. In this period, patients were killed not only with lethal injections but also through the introduction of a starvation diet. It has been estimated that approximately 200,000 individuals with mental disorders of all kinds were put to death, although their mass murder has received relatively little historical attention (Von Cranach, 2003; Strous, 2007).

Until now, the problem of stigma has been considerable and has influenced patient-doctor relationships as well as compliance and treatment programs. Many people hold negative stereotypes of those who suffer from mental illness and mentally ill patients are often considered to be less human and less worthy. Quite often they are treated as if they deserved less help or care in comparison to those with other kinds of illness (one example is that while suffering from comorbid medical disease, they are provided with less medication or procedures than "normal" patients). There are countless ways that the mentally ill people have been degraded in film, TV, literature, and in ordinary life. For example, in 1996 Philo found that in 66% of items concerning mental illness on UK television the focus was on violence (Philo, 1996). One of the important aims of a big branch of psychiatry – social psychiatry – is to combat stigma in order to reunite patients with their society. In this context, the contemporary approximation between psychiatry and neuroscience may be very helpful. By including more aspects of the neuroscientists' concentration on brain disorders, it is possible to change the way in which we think about mental disorders.

2. Historical views on the origins of mental disorders

In the evolution of psychiatry, the titanic human struggle to understand abnormal behavior – in the context of either the mind or of the body – has played a central role. In the course of searching for the causes of mental disorders, the idea of the integration between the forces of the mind (soul) and brain (body) was long in coming, changing throughout the ages and building up a fascinating story (Gerdenio, 2005).

The beginnings of the history of medicine are characterized by a supernatural and magical approach to both mental and physical illnesses. For much of our recorded history, deviant behavior was considered a reflection of the battle between good and evil. The great cultures of the past, such as those of Egypt and Mesopotamia, fluctuated between naturalistic and supernatural explanations of diseases (Ackerknecht, 1959).

Although the ancient Greeks felt that the gods controlled health and illness, they looked beyond supernatural influences and explored biological, psychological and social causes of illness. In the classical Greek era, attempts were made to explain physical and psychological phenomena with more scientific approaches. The Greeks believed that the mind and the body were closely interconnected. The famous Greek physician Hippocrates (460–377 BC) – the father of medicine – thought that disease resulted from an imbalance in four body fluids or humors, rather than from spiritual factors. He applied humoral theory to mental illness and was insistent that all mental disorder should be explained on the basis of natural causes. Unpleasant dreams and anxiety were seen as being caused by a sudden flow of bile to the brain, melancholia was thought to be brought on by an excess of black bile, and exaltation by a predominance of warmth and dampness in the brain. Temperament was thought to be choleric, phlegmatic, sanguine or melancholic depending on the dominant humor. According to Hippocrates, the imbalances in body fluids might originate in the patient's environment and in modern psychiatry we also believe that temperamental and personality factors may predispose one to certain mental disorders, and that the development of personality is highly influenced by the environment of the child. Hippocrates was also sensitive to interpersonal, psychological and stress factors that contribute to disturbed behavior. This early biopsychosocial perspective was further championed by Plato, Aristotle and Galen, until its demise in the Middle Ages (Ackerknecht, 1959).

Plato (427–347 BC) reintroduced a mystical element. He believed in two types of madness, the first was divinely inspired and gave the recipient prophetic powers, while the second was caused by disease. However, he also believed in the strong connection between mind and body. He wrote: "As it is not proper to cure the eyes without the head, nor the head without the body, so neither it is proper to cure the body without the soul" (Gerdenio, 2005). On the other hand, for him, as for the greatest thinkers throughout recorded history, reason has reigned supreme. The traditional paradigm has been one of a dichotomy where refined and uniquely human *reason* pitches an ongoing battle for control over animalistic and lustful *emotions*. This Platonic dichotomy remains a pillar of Western thought (Lehrer, 2009).

Aristotle (384–322 BC) believed that because reason was immortal it must be immune to illness, so all kinds of illnesses, either mental or otherwise, must be rooted in man's physical structure. He felt that certain distinct emotional states, including joy, anger, fear and courage, impacted on the functioning of the human body. The treatment of mental problems should include talking and using logic to influence the soul and psyche. Nowadays, the use of logic and reason to influence emotional and behavioral problems is one of the major principles behind cognitive-therapy (CBT) (Ackerknecht, 1959).

Cicero (106-43 BC) – a Roman philosopher - rejected Hippocrates' bile theory, stating that emotional factors could cause physical illness and "perturbations of the mind may proceed from a neglect of reason". Man could help with his own cure through "philosophy", which would nowadays be known as psychotherapy (Zilboorg and Henry, 1941).

It is also worth mentioning Galen (130–200 AD), who felt that the brain was the rational soul and the centre of sensation and reason and who developed a holistic program of medical practice.

3. The supernaturalism of the Middle Ages

Probably in response to the highly turbulent, frightening and stressful times during the Black Death and numerous wars, the focus of supernatural influences to explain events including illness and health, became common (Gerdenio, 2005). The Christian Church of the early Middle Ages was very much concerned with the life hereafter and not on earth. It also greatly stressed the healing powers of religious symbols.

Insanity was believed to be caused by spiritual matters such as the influence of demons, witches and sin. The phenomenon of mental disturbance troubled the early Christian authorities. The Devil could not always be blamed because the content of the madness seemed to have a religious significance. Therefore, it was difficult for the Church authorities to decide whether the mad were communicating with the Devil or whether they were saints. However, in the early 7th century the Devil was accepted as a culprit for all types of deviant behavior and Demonology became the "psychiatry" of the day. Treatment was a spiritual issue: atonement for sins, exorcisms, prayer. Some insane people were chained to church walls in order to benefit from prayers while others were tortured (Gerdenio, 2005).

In 1484 Pope Innocent VIII approved the persecution of "witches". Among them many bizarre, insane people – mainly women - were executed. It was assumed that women stimulated men's carnal desire, so the blame for sinful erotic behavior was laid here; women tempted men so they must be the Devil's agents. Psychotic women who openly acted erotically were easy targets. The *Malleus Maleficarum*, the treatise on witches, stated that "all witchcraft comes from carnal lust which is in women insatiable" (Johnstone et al., 1998).Nowadays, the majority of those poor creatures – tortured and killed - would probably be diagnosed with psychosis, or histrionic disorders (e.g. some examples of stigmata diaboli). In this climate, nearly all natural thinking about mental illness disappeared.

However, not everyone believed in supernatural causes for mental disorders. For example, St Thomas Aquinas (1225–1274) reasoned that the soul was unable to become sick and, therefore, mental illness must have a physical cause.

During the Renaissance a renewed interest in the physical and medical worlds emerged. The interest in mind and soul was considered unscientific and thus relegated to the domain of philosophers and priests. As a result, the emphasis on scientific observation and experimentation provided a model for future research and teaching and the biological side of the integrative biopsychosocial perspective was emphasized.

4. Dualism: the somatic and psychological school

René Descartes (1596–1650), argued that the mind and the body were separate. He believed in *duality*, that the mind and body were two distinct entities (Robinson, 2003). No philosophical concept has been as widely influential in the field of psychiatry. This dualism of mind and body became the basis for Western medicine until recently. Descartes clearly identified mind with consciousness and with self-awareness while distinguishing it from the brain as the seat of intelligence. Hence, he was the first to formulate the mind-body problem in the form in which it exists today. No philosophical concept has been as widely influential in the field of psychiatry. This dualism of mind and body became the basis for Western medicine until recently. Descartes clearly identified mind with consciousness and with self-awareness while distinguishing it from the brain as the seat of intelligence. Hence, he was the first to formulate the mind-body problem in the form in which it exists today. In his famous book *Descartes' Error: Emotion, Reason, and the Human Brain* António Damásio argues that René Descartes' "error" was the dualist separation of mind and body, rationality and emotion (Damasio, 2005).

Not all philosophers shared those ideas. David Hume, the eighteenth-century Scottish philosopher, declared that reason was the "the slave of the passions" (Hume, 2000). Hume also contended that reason alone is insufficient to motivate human action. Reason must excite passion and emotion if it is to have any influence on the motives or actions of the will. He claimed that reason can prevent the expression of a passion or emotion only by exciting a contrary passion or emotion. Jonah Lehrer, in his book, *How We Decide* discusses this very issue and notes that:

> The crucial importance of our emotions – the fact that we can't make decisions without them – contradicts the conventional view of human nature, with its ancient philosophical roots. (…) The expansion of the frontal cortex during human evolution did not turn us into purely ra-

> tional creatures, able to ignore our impulses. In fact, neuroscience now knows that the opposite is true: a significant part of our frontal cortex is involved with emotion (Lehrer, 2009).

5. Dualism: neurology-psychiatry

A clear distinction between neurology and psychiatry existed for many decades. Neurologists treated organic illnesses (like strokes, brain tumours etc.) whose medical cause was well recognized: damage of the brain's structure. In response to the common fear of psychiatry, they also treated the so-called "nervous disorders", using refreshment, massage and treatment in sanatoriums. Nowadays, such "nervous disorder" are diagnosed as different kinds of neurotic problems. The field of psychiatry was restricted to treating severe psychotic diseases (like schizophrenia or manic-depressive illness), in which the structural background was unknown.

The belief in the distinction between "neurological" and "psychiatric" disease is still often held by a significant number of the general public as well as by the wider medical community, although it arises from a stigmatized view of psychiatry. Some people find it useful to be able to consider "real" neurological disease arising from brain pathology as distinct from psychiatric disturbances arising from weak moral fibre or bad breeding (Ring, 2002).

Nevertheless, it has been proven that purely "neurological" disorders cause behavioral and emotional disturbances (e.g. Huntington's disease, Pick disease, epilepsy, the damage of certain brain regions, e.g. frontal lobes, etc.).

One of the first and most exciting examples that proved the above statement was the case of Phineas Gage. He was a railroad construction foreman who survived an accident (1848) in which a large iron rod was driven completely through his head, destroying much of his brain's left frontal lobe. He is now remembered mostly for its reported effects on his personality and behaviour. These effects were so profound that his friends saw him as “no longer Gage.” Before the accident he was known as diligent, even-tempered, nice gentleman. After the accident, he became quick-tempered, unable to pursue a defined goal, rude, deprived of empathy. Dr John Harlow (the physician responsible for Gage) described him as being ‘intellectually feeble’, ‘exceedingly capricious’, ‘childish’, coarse, very impulsive, and indifferent towards others. “A child in his intellectual capacity and manifestations, he has the animal passions of a strong man” – Harlow noted (Harlow, 1868). Such a transformation of a man once admired for his professional and social virtues, finally led to his being fired from his job and subsequent social deterioration.

Gage’s case is cited as one of the earliest pieces of evidence that frontal lobe damage can influence one’s personality and social interactions. Earlier it had been widely believed that this part of brain does not play any role in the creation of human personality. Gage developed a mental state that could be named as acquired psychopathy - such a kind of personality disturbance that involves the triad of characteristic dimensions: an arrogant and deceitful style of interpersonal functioning, a deficiency in experiencing, understanding and expressing emotions, and highly impulsive behavior. From the functional point of view, psychopathy seems to be related to an inappropriate pattern of activation in the frontal-temporal circuit. Impulsivity and impaired moral judgment in psychopathic patients are probably related to functional and/or anatomical deficits of the structures responsible for decision-making, behavioral steering and emotional control (i.e. the orbitofrontal, ventromedial prefrontal, and the cingulate cortex). On the other hand, shallow affect and a lack of empathy seem to be the outcomes of dysfunctions of medial temporal regions (particularly the hippocampus and amygdala), as they are the key elements of the ‘emotional processing system’. The amygdala is the crucial structure in determining responses to signals indicating distress in others; therefore its dysfunction leads to antisocial behavior, as well as seeming to determine susceptibility towards solving personal moral dilemmas in a utilitarian and unemotional manner.

Three decades following Gage's accident, David Ferrier presented the earliest experimental evidence (derived from studies performed on monkeys), suggesting that consecutive areas of the cerebral cortex differ between each other in terms of function, and the destruction of prefrontal lobes – while exerting virtually no impact on neurological performance – causes "very decided alterations in the animal's character and behavior (…) while not actually deprived of intelligence, they had lost, to all appearance, the faculty of the attentive and intelligence observation" (Ferrier, 1878).

Contemporarily, much of researchers' attention is being paid to the issue of personality changes (including those relevant for the emotions) taking place in subjects with certain neuropsychiatric diseases. For example, in 1939, Heinrich Klüver and Paul Bucy reported that the surgical removal of both temporal lobes (including the amygdalae) in monkeys produced a dramatic behavioral condition now referred to as the Klüver-Bucy syndrome. After surgery, the monkeys, who had previously feared humans, no longer showed such fear. They also demonstrated a number of other behavioral changes, including hyperorality (a compulsion to examine objects by way of the mouth), hypersexuality (excessive sexual behavior), hypermetamorphosis (excessive tendency to react to visual stimuli), and visual agnosia (inability to recognize familiar objects).

For over two decades, researchers at the University of Iowa have been studying a woman (patient SM), who acquired damage to both amygdalae (due to a rare congenital genetic condition known as Urbach-Wiethe disease). SM did not show fear in any threatening situations and her ability to detect dangerous situations was completely impaired. As SM was capable of experiencing other emotions normally, she was not emotionless, but rather fearless, contributing greatly to her high incidence of life-threatening experiences and influencing her behavior and decisions (Feinstein at all., 2011).

Possibly the most convincing data on emotions, personality and brain damage derive from the studies regarding subjects with frontotemporal dementia (FTD), and its various subtypes (including Pick's disease [Hodges at al., 2004]). FTD belongs to the group of neurodegenerative disorders, characterized by chronic clinical course, with progressive deterioration of cognitive functioning, altered behavior, and personality changes, leading to severe disability, and – ultimately – death. In subjects with FTD neuropathological lesions are typically located in frontal lobes and anterior temporal regions, especially ventromedial prefrontal cortex (vmPFC) and orbitofrontal cortex (OFC) (Rosen et al., 2002).

Compared to patients with Alzheimer's disease, subjects suffering from FTD tend to present more prominent social and personality distortions, while cognitive and neuropsychological deficits are usually milder (during early stages of the illness, at least). Probably the most typical behavioral motifs observed in course of FTD are related to difficulties in modulating social conduct. Notably, the following symptomatic triad: transgression of social norms (reflecting an impairment of the ability to modulate social behaviour), 'emotional blunting' (amounting to a loss of empathy), and loss of insight with regards the subject's behavior and their consequences constitute the core clinical features of the illness. As a result, as the intensity of basic drives remain intact, lack of empathy often leads to overtly sociopathic acts, such as sexual indiscretions, physical aggression, stealing, breaking into other people's properties, and paraphilias. The severity of those symptoms is even greater in subjects whose OFC is affected, as the presence of orbitofrontal lesions leads to the loss of impulse control.

Thus, neurological disorders or damage to certain regions of the brain result in profound emotional and behavioral disturbances, leading to psychopathological states

On the other hand, there is more and more data on structural changes in the brains of the patients with "purely" psychiatric disorders, such as schizophrenia or major depression. It has been suspected for over a century that schizophrenia and manic-depressive disorder (bipolar disorder) are disorders of the brain. Thanks to advanced neuroimaging methods, there is strong evidence confirming that patients with severe psychotic disorders (e.g. significant loss of brain gray matter, enlarged ventricles, enlarged amygdalia) have important structural brain abnormalities.

Moreover, the absence of detectable structural changes does not rule out the possibility that some more subtle, but nonetheless important biological changes are occurring. These changes may simply be below the level of detection due to the still-limited nature of the techniques available today.

Thus, the dichotomy of "neurological" and "psychiatric" cannot be valid anymore. It is worth citing Joseph B. Martin, former Dean of Harvard Medical School and a neurologist by training, who summarized the argument for the reunion of the two fields: "the separation of the two categories is arbitrary, often influenced by beliefs rather than proven scientific observations. And the fact that the brain and mind are one makes the separation artificial anyway" (Martin, 2002). In the near future we will witness the triumph of neuropsychiatry, a medical division that may unite psychiatry and neurology.

6. Dualism in psychiatry: "somatic school" and "psychosocial school"

In the 19th century, the advancement in the understanding of mental and physical illness led to the development of biological psychiatry. The milestones for biological psychiatry were the discoveries of Rudolf Virchow (1821–1902) and Louis Pasteur (1822–1895). These scientists proved that disease could be attributed to dysfunctions_at the cellular level or to germs. It transpired, for example, that neurosyphilis – one of the predominant causes of hospitalization in psychiatric institutions in the 19th century - was caused by microorganisms entering the brain following sexual activity. The treatment of neurosyphilis with malaric fever was the first efficient method in the struggle against psychopathology (resulting in a Nobel prize in 1927 for Julius Wagner-Jauregg) (Gerdenio, 2005).

Also Emil Kraepelin (1856–1926), a German psychiatrist who is identified as the founder of modern scientific psychiatry, as well as of psychopharmacology and psychiatric genetics, believed the chief origin of psychiatric disease to be biological and genetic malfunction. He asserted that mental disorders were brain disorders, and mental illness could be classified as arising from either exogenous or endogenous influences.

In the middle of the 20th century, psychiatry underwent a transition that is often referred to as the "psychopharmacology revolution." The 1950s witnessed the development in pharmacological treatment of different kinds of disorders. Chlorpromazine was prescribed for psychoses, lithium carbonate for mania. Then, in rapid succession, there was the development of tricyclic antidepressants, monoamine oxidase inhibitors, and benzodiazepines with their anxiolytic properties. It opened the way to a better understanding of the neurobiology of mental disorders.

On the other hand, it was also during the 19th century that clinical psychology was born as a distinct specialty. The symbolic date associated with the birth of modern, scientific psychology is 1879 when the opening of the first psychological laboratory by Wilhelm Wundt took place. Initially, academic psychologists were not concerned with serious forms of mental illnesses as this area was already being addressed by the developing fields of psychiatry and neurology within the asylum (big psychiatric institutions) movement (Ludy, 2007). This trend changed towards the end of the 19th century, at the time Sigmund Freud first developed his "talking cure". With Freud, the first scientifically clinical application of psychology began. He went on to develop theories about the unconscious mind and the mechanism of repression, and established the field of verbal psychotherapy by creating psychoanalysis, a clinical method for treating psychopathology through dialogue between a patient and a psychoanalyst. Psychoanalysis has helped inspire the development of many other forms of psychotherapy, some diverging from Freud's original ideas and approach. However, one of the unfortunate results of promoting psychotherapy was the fact that psychiatry in post-World War II America became so greatly influenced by psychoanalysis that it was transformed from a medical discipline into a practicing therapeutic art, surprisingly unconcerned with the brain as an organ of mental activity (Shorter, 1997).

A nowadays very influential psychotherapeutic school, cognitive-behavioral therapy (CBT), struggles with another important dichotomy within psychiatry: cognitions and emotions. CBT holds that most of our emotions and behaviors are the result of what we think or believe about ourselves, other people, and the world. These cognitions shape how we interpret and evaluate what happens to us, influence how we feel about it, how are we emotionally aroused and provide a guide as to how we should respond. Unfortunately, sometimes our interpretations, evaluations, and underlying beliefs thoughts contain distortions, errors, or biases, or are not very useful or helpful. This results in unnecessary suffering and often causes us to react in ways that are not in our best interests. Cognitive restructuring is a set of techniques for becoming more aware of our thoughts and for modifying them when they are distorted or not useful. It uses reason and evidence to replace distorted thought patterns with more accurate, believable, and functional ones, in turn resulting in positive changes in emotions. Thus, cognitions and emotions are strongly interrelated (Popiel and Prągłowska, 2008).

7. Towards integration

It is amazing to notice that in 1895 nobody else but Freud – the father of psychoanalysis – wrote “A project for a scientific psychology”, in which he tried to adopt a neural model of behaviour in an attempt to develop a scientific psychology. In this work Freud stated that what we experience as the conscious or unconscious processing of information is reflected in the neuronal architecture of the brain and the nervous system in general. He concluded that ‘speaking therapy’ can rearrange neuronal connections and change the character of psychological experiences (Freud, 1953; Centonze et al., 2004). Regarded as radical, those ideas were rejected by Freud's contemporaries and, because of the immaturity of brain science at the time, he abandoned this biological model for a purely mentalistic one, based on verbal reports of subjective experiences.

The most important steps toward breaking down the mind-brain dichotomy took place thanks to the progress in psychiatry in the 1960s, 1970s and 1980s. The milestones of this progress were:

1. The development of efficient treatments;
2. New clinically validated and objective criteria established for diagnosing mental illness;
3. Renewed interest in the biology of mental illness, especially in genetics;
4. Major developments in brain sciences, in particular the analysis of how different aspects of mental functioning are represented by different regions of the brain. The conclusion that specific lesions of the brain produce specific alterations in behaviour, and specific alterations in behaviour are reflected in characteristic functional changes in the brain (Kandel, 1998).

Neuroscience has started to identify the neural correlates not only of mental disorders, but also of therapeutic changes. It transpired that procedural re-learning in psychotherapy can influence the structure and functions of the brain by altering synaptic plasticity and gene-expression. There is growing evidence for a modification of gene expression by emotional experience and signals from the environment (so called epigenetic modification) (Kandel, 1979, 1998). For example, several studies on monkeys have been performed, with Suomi et al's study of gene-environment interaction in rhesus monkeys being particularly important. Approximately 5-10% of rhesus monkeys growing up in the wild consistently exhibit impulsive and/or inappropriately aggressive responses to mildly stressful situations throughout development; those same individuals also show chronic deficits in their central serotonin metabolism. Laboratory studies have demonstrated that although these characteristics are highly heritable, they are also subject to major modification by specific early experiences, particularly those involving early social attachment relationships. For example, a specific polymorphism in the serotonin transporter gene is associated with deficits in early neurobehavioral functioning and serotonin metabolism, extreme aggression, and excessive alcohol consumption can be observed among monkeys who experienced insecure early attachment relationships, but not in monkeys who developed secure attachment relationships with their mothers during infancy. Because daughters tend to develop the same type of attachment relationships with their own offspring that they experienced with their mothers early in life, such early experiences provide a possible non-genetic mechanism for transmitting these patterns to subsequent generations (Suomi, 2003).

Further evidence for the modification of gene expression by emotional experience and signals from the environment is the report which summarizes the results of the research on the effects of increased pup licking and grooming and arched-back nursing by rat mothers. It seems that these activities/practices altered the offspring epigenome at a glucocorticoid receptor (GR) gene promoter in the hippocampus, through alteration in DNA methylation. This resulted in better stress responses in the offspring throughout their whole life (Weaver et al., 2004).

Another important field that concerns close connections between neurobiology and psychology are the studies on explicit and implicit memory which are associated with different brain structures. Numerous studies have shown that patients who lack explicit memory are capable of learning tasks that do not require conscious awareness. A good example is the famous case study of the patient HM. He suffered from refractory epilepsy, with numerous seizures each day. In 1957, when he was 27 years old, he underwent radical surgical procedure: the removal of epileptic foci located in the temporal lobes with bilateral removal of temporal cortex, amygala and 2/3 of hippocampus.

After the surgery, partial retrograde amnesia and total anterograde amnesia (inability to memorize and learn) were observed. However, he was able to develop unconscious emotional associations (connecting stimuli with emotions). For example, although he did not remember the fact that he had visited his mother in hospital, he had an 'unclear feeling' that something bad had happened to her. What is also interesting, Brenda Milner, the neuropsychologist who was observing his case and examining him regularly, one day pricked him with a needle when shaking hands. The next day, although he did not remember either the person nor the prick, he retracted his hand when greeting. Those studies imply the neural basis for a set of unconscious mental processes (Milner, 2005; Milner et al., 1989). People might unconsciously associate certain situations or important objects (other people) with positive or negative feelings. As a consequence, they might unconsciously react to signals, avoid or search for situations or people who subconsciously fit the patterns of previous experiences; at the same time they are unaware of memories that have left such emotional traits. This is in accordance with the knowledge that the main changes in the course of analytic therapy concern implicit memory.

Nowadays, the neurobiological basis of psychotherapy is obvious, as is the statement that psychotherapy changes the brain. PET studies of patients with different disorders (OCD, phobia, depression) support the hypothesis that disorders show more or less characteristic changes of regional brain activity that may be normalized by psychotherapy (Baxter et al., 1992; Brody et al., 2001; Furmark et al., 2002; Rok-Bujko, 2009). Nancy Andreasen, the previous editor in chief of the "American Journal of Psychiatry", declared: "Psychotherapy – sometimes humiliated as a 'verbosity' – in a way is as 'biological' as pharmacotherapy."

All of this has created a new intellectual framework for psychiatry: "All mental processes are biological, and therefore any alteration in those processes is necessarily organic". It is worth citing Kandel's famous manifesto here (Kandel, 1998):

Principle 1. All mental processes, even the most complex psychological processes, derive from operations of the brain.

Principle 2. Genes and their protein products are important determinants of the pattern of interconnections between neurons in the brain and the details of their functioning. Genes, and specifically combinations of genes, therefore exert a significant control over behaviour. As a corollary, one component contributing to the development of major mental illnesses is genetic.

Principle 3. Altered genes do not, by themselves, explain all of the variance of a given major mental illness. Social or developmental factors also make a very important contribution. Behaviour and social factors exert actions on the brain by feeding back upon it to modify the expression of genes and thus the function of nerve cells. Learning produces alterations in gene expression.

Principle 4. Alterations in gene expression induced by learning give rise to changes in patterns of neuronal connections.

Principle 5. Insofar as psychotherapy or counseling is effective and produces long-term changes in behaviour, it presumably does so through learning, by producing changes in gene expression that alter the strength of synaptic connections and structural changes that alter the anatomical pattern of interconnections between nerve cells of the brain.

However, the biological hegemony over psychological approaches may imply some kind of “biological reductionism”, which holds that all psychological states are really brain states. According to this view, our minds are just a function of our brains and mental states are constituted solely by brain states. Mental disorders should then be regarded as nothing other than chemical imbalances, and psychiatrists should not treat individuals, but their brains (Fuchs, 2004; Glannon, 2009). The most influential and prominent scientists have stated that “we are our synapses” (LeDoux, 2002), or “we are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules” (Crick, 1994).

However, although as psychiatrists we fully agree with some bioethical statements and feel that understanding the functions of the brain is necessary, we also emphasize the fact that it is insufficient to account for all the physiological and psychological properties that make each of us a unique person. The mind is not solely based on brain structure and functions, but also on the continuous interaction of the brain with the body and with the external world (Glannon, 2009).

The body shapes one's perception and psychological response to the world. For a different body, this perception and response would be different. As the environment shapes our experience of interacting with the world and others, living in a different environment would result in creating a different content and meaning of a person's experience and memory of the interactions. Thus transplanting a brain into a different body – if it were ever possible – would not preserve the identity of the person whose brain was transplanted (Fuchs, 2004; Glannon, 2009).

Kenneth S. Kendler – an American psychiatrist best known for pioneering research in psychiatric genetics, particularly the genetic causes of schizophrenia, who is one of the most oft cited psychiatry researchers, postulated: "Stop searching for Big Simple Explanations (…). We have hunted for big, simple neuropathological explanations for psychiatric disorders and have not found them. We have hunted for big, simple neurochemical explanations for psychiatric disorders and have not found them. We have hunted for big, simple genetic explanations for psychiatric disorders and have not found them" (Kendler, 2005).

In conclusion, all major psychiatric disorders are complex and multifactoral. Thus, in contemporary psychiatry, the predominant model of disorders is the biopsychosocial one. This perspective utilizes a multidimentional and interactive approach. The biological, psychological and social aspects of health and illness influence each other, creating ethological networks. This concept has a practical impact – promoting combined therapies which match biological treatment with psychosocial interventions. Many studies have confirmed the efficacy of such an approach and the superiority of combined treatment of most disorders over unimodal treatments.

References

Ackerknecht, E.H. (1959). A Short History of Psychiatry (New York: Hafner).

Baxter, L.R., Schwartz, J.M., Bergman, K.S., Szuba, M.P., Guze, B.H., and Mazziotta, J. (1992). Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive-compulsive disorder. Archives of General Psychiatry *49*, 681–689.

Brody, A.L., Saxena, S., Stoessel, P., Gilles, L.A., Fairbanks, L.A., and Alborzin, S. (2001). Regional brain metabolic changes in patients with major depression treated with either paroxetine or interpersonal therapy: Preliminary findings. Archives of General Psychiatry *58*, 631–640.

Brüne, M., Brüne-Cohrs, U., McGrew, WC., and Preuschoft, S. (2006). Psychopathology in great apes: concepts, treatment options and possible homologies to human psychiatric disorders. Neuroscience and Biobehavioral Reviews *30 (8)*, 1246–1259.

Centonze, D., Siracusano, A., Calabresi, P., and Bernardi, G. (2004). The Project for a Scientific Psychology (1895): A Freudian anticipation of LTP-memory connection theory. Brain Research Reviews *46*, 310–314.

Crick, F. (1994). The Astonishing Hypothesis: The Scientific Search for the Soul (New York: Scribner's).

Damasio, A. (2005). Descartes' Error: Emotion, Reason, and the Human Brain (New York: Penguin Books).

Feinstein, J.S., Adolphs, R., Damasio, A., and Tranel D. (2011). The human amygdala and the induction and experience of fear. Current Biology CB *21 (1)*, 34–38.

Ferrier, D. (1878). The Goulstonian Lecutres on the localisation of cerebral diseases. British Medical Journal *1*, 443–447.

Freud, S. (1953). Project for a scientific psychology (1895). The Standard Edition of the Complete Psychological Works of Sigmund Freud, vol. 1 (London: Hogarth Press), pp. 283–397.

Furmark, T., Tillfors, M., Merteinsdottir, I., Fisher, H., Pissiota, A., and Angstrom, B. (2002). Common changes in cerebral blood flow in patients with social phobia treated with citalopram or cognitive-behavioral therapy. Archives of General Psychiatry *59*, 425–433.

Fuchs, T. (2004). Neurobiology and psychotherapy: and emerging dialogue. Current Opinion in Psychiatry *17*, 479–485.

Gerdenio, M. (2005). Foundations and early history of clinical psychology. In: Contemporary Clinical Psychology, T.G. Plante, ed. (Hoboken, NJ: John Wiley and Sons), pp. 33–49.

Glannon, W. (2009). Our brains are not us. Bioethics *23(6)*, 321–329.

Harlow, J.M. (1868). Recovery from the passage of an iron bar through the head. Publ. Massachusetts Medical Society *2*, 327–347.

Hodges, J.R., Rhys-Davies, R., Xuereb, J.H., Casey, B., Broe, M., Bak, T.H., Krill, J.J., and Halliday, G.M. (2004). Clinicopathological correlates in frontotemporal dementia. Annals of Neurology *56*, 399–406.

Hume, D. (2000). A Treatise of Human Nature, D.F. Norton and M.J. Norton, eds. (Oxford: Oxford University Press), 2.3.3.4

Johnstone, E.C., Freeman, C.P.L., and Zealey, A.K. (1998). Companion to psychiatric studies (6th rev ed.) (Edinburgh: Churchil Livingstone).

Kandel, E.R. (1979). Psychotherapy and the single synapse. The impact of psychiatric thought on neurobiologic research. New England Journal of Medicine *301*, 19, 1028–1037.

Kandel, E.R. (1998). A new intellectual framework for psychiatry. American Journal of Psychiatry *155(4)*, 457–469.

Kendler, K.S. (2005). Toward a philosophical structure for psychiatry. American Journal of Psychiatry *162*, 433–440.

LeDoux, J. (2002). Synaptic Self: How Our Brains Become Who We Are (New York: Viking).

Lehrer, J. (2009). How We Decide (New York: Houghton Mifflin Harcourt).

Ludy Jr., B., (2007). A Brief History of Modern Psychology (Malden, MA: Blackwell Publishing).

Martin, J.B. (2002). The integration of neurology, psychiatry, and neuroscience in the 21st century. American Journal of Psychiatry *159 (5)*, 695–704.

Milner, B. (2005). The medial temporal-lobe amnestic syndrome. Psychiatric Clinics of North America *28(3)*, 599–611.

Milner, B., Corkin, S., and Teuber, H.L. (1989). Further analysis of the hippocampal amnestic syndrome: Fourteen year follow-up study of H.M. Neuropsychologia *6*, 215–234.

Nesse, R.M. (2005). Evolutionary psychology and mental health. In: The Handbook of Evolutionary Psychology, D. Buss, ed. (Hoboken, NJ: John Wiley and Sons), pp. 903–929.

Philo, G. (1996). Media and Mental Distress (New York: Addison Wesley Longman).

Popiel, A., and Prągłowska, E. (2008). Psychoterapia poznawczo-behawioralna. Teoria i praktyka (Warszawa: Wydawnictwo Paradygmat).

Ring, H.A. (2002). Wall between neurology and psychiatry. Neuropsychiatry is alive and well. British Medical Journal *325*, 778.

Robinson, H. (2003). Dualism. In: The Stanford Encyclopedia of Philosophy, E.N. Zalta, ed. (Fall 2003 Edition), http://plato.stanford.edu/entries/dualism/

Rok-Bujko, P. (2009). Neurobiologiczne podstawy psychoterapii (Neurobiological Basis of Psychotherapy). In: Od neurobiologii do psychoterapii (From Neurobiology toward Psychotherapy), S. Murawiec, and C. Żechowski, eds. (Warszawa: Instytut Psychiatrii i Neurologii), pp. 235–266.

Rosen, H.J., Gorno-Tempini, M.L., Goldman, W.P., Perry, R.J., Schnuff, N., Weiner, M., Feiwell, R., Kramer, J.H., and Miller, B.L. (2002). Patterns of brain atrophy in frontotemporal dementia and semantic dementia. Neurology *58*, 198–208.

Shorter, E. (1997). A History of Psychiatry. From the Era of Asylum to the Age of Prozac (Hoboken, NJ: John Wiley and Sons).

Strous, R.D. (2007). Psychiatry during the Nazi era: ethical lessons for the modern professional. Annals of General Psychiatry *6 (1)*, 8.

Suomi, S.J. (2003). Gene-environment interaction and the neurobiology of social conflicts. Annals of the New York Academy of Sciences *1008*, 132–139.

Von Cranach, M. (2003). The killing of psychiatric patients in Nazi Germany between 1939–1945. Israel Journal of Psychiatry and Related Sciences *40(1)*, 8–18.

Weaver, I.C., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., and Meaney, M.J. (2004). Epigenetic programming by maternal behavior. Nature Neuroscience *Aug. 7(8)*, 847–854. Epub 2004 Jun 27.

Zilboorg, G., and Henry, G.W. (1941). A History of Medical Psychology (New York: W.W. Norton and Company).