FACT AND VALUE IN EMOTION

Edited by
Louis C. Charland
Peter Zachar

John Benjamins Publishing Company
Fact and Value in Emotion
Consciousness & Emotion Book Series publishes original works on this topic, in philosophy, psychology and the neurosciences. The series emphasizes thoughtful analysis of the implications of both empirical and experiential (e.g., clinical psychological) approaches to emotion. It will include topical works by scientists who are interested in the implications of their empirical findings for an understanding of emotion and consciousness and their interrelations.

Editors
Ralph D. Ellis
Clark Atlanta University

Natika Newton
Nassau County Community College, NY

Editorial Board
Carl M. Anderson
McLean Hospital, Harvard University School of Medicine, Cambridge, MA

Maxim I. Stamenov
Bulgarian Academy of Sciences

Bill Faw
Brewton Parker College, Mt. Vernon, GA

Douglas F. Watt
Quincy Hospital, Boston, MA

Eugene T. Gendlin
University of Chicago

Peter Zachar
Auburn University, Montgomery, AL

Jaak Panksepp
Bowling Green State University, OH

Advisory Editors
Bernard J. Baars
Wright Institute, Berkeley, CA

Alfred R. Mele
Florida State University, Tallahassee, FL

Thomas C. Dalton
California Polytechnic Institute, San Luis Obispo, CA

Martin Peper
University of Freiburg, Freiburg, Germany

Nicholas Georgalis
East Carolina University, Greenville, NC

Edward Ragsdale
New York, NY

George Graham
Wake Forest University, Wake Forest, North Carolina

Howard Shevrin
University of Michigan, Ann Arbor, MI

Valerie Gray Hardcastle
University of Alabama, Birmingham, AL

Lynn Stephens

Nicholas Georgalis

University of Arizona, Tucson, AZ

Alfred W. Kaszniak

Kathleen Wider

University of Michigan, Dearborn, MI

Volume 4
Fact and Value in Emotion
Edited by Louis C. Charland and Peter Zachar
Fact and Value in Emotion

Edited by
Louis C. Charland
University of Western Ontario

Peter Zachar
Auburn University, Montgomery

John Benjamins Publishing Company
Amsterdam / Philadelphia
# Table of contents

Fact and value in emotion: An introduction and historical review  
*Peter Zachar*

A moral line in the sand: Alexander Crichton and Philippe Pinel on the psychopathology of the passions  
*Louis C. Charland*

How to evaluate the factual basis of emotional appraisals?  
*Mikko Salmela*

The problem with too much anger: A philosophical approach to understanding anger in borderline personality disordered patients  
*Nancy Nyquist Potter*

A confusion of pains: The sensory and affective components of pain, suffering, and hurt  
*Jennifer Radden*

Ethical implications of emotional impairment  
*Abraham Rudnick*

Facts and values in emotional plasticity  
*Luc Faucher and Christine Tappolet*

Attributing aberrant emotionality to others  
*Nick Haslam and Stephen Loughnan*

Emotion and the neural substrate of moral judgment  
*Anthony Landreth*
The phenomenology of alexithymia as a clue to the intentionality of emotion

Ralph D. Ellis

A phenomenologist’s view of the omnipresence of the evaluative in human experience: Knowledge as a founded mode and the primacy of care

Edwin L. Hersch

Index
Fact and value in emotion

An introduction and historical review

Peter Zachar

1. The Humean schism

David Hume famously claimed that one cannot derive an ‘ought’ from an ‘is,’ or in less poetic terms, one cannot derive a value from a fact. Hume was arguing that the imperatives given to our preferences, goals and actions are derived from emotion and not from logical reasoning. Emotion, therefore, had a prominent place in Hume’s thinking.

In the realm of moral philosophy, Hume’s philosophical descendants continued to insist on the primacy of the emotional. For example, utilitarians such as Bentham identified the good with pleasure and logical positivists such as A. J. Ayer (1936) held that claims about goodness merely verbalize emotional reactions of liking and claims about badness verbalize emotional reactions of disliking. Ayer believed that there are no moral facts.

There is, however, an argument to be made that, with respect to science, the role of emotion was diminished as a result of the Humean perspective. Likely aided by the modern distinction between the subject and the object that is often attributed to Descartes, the Humean schism between fact and value led to a model of science in which matters of fact and existence, or objective, observable information were the only legitimate kind of data. Thinkers who emphasized the factual side of the Humean schism in the scientific domain included Bertrand Russell, the Wittgenstein of the Tractatus, and the logical positivists.

In psychology, an important role for emotions was always acknowledged by those who were influenced by romantic thinkers such as Rousseau and Nietzsche. The humanists and the various psychodynamic schools come to mind in this respect. Philosophical romantics have consistently been associated with antagonism toward science, viewing the scientific perspective on human nature as sterile and
shallow. The romantics tended to proclaim their outsider status with respect to science, which, for psychological and sociological reasons, limited how receptive some scientists were to their arguments. In the middle of the 20th century, however, skeptical analyses of the Humean schism were put forward by thinkers who were not associated with the romantic critique. One can even speculate that the fact-value distinction became a more legitimate target of philosophical critique once another of Hume’s dichotomies, the analytic-synthetic distinction, was undermined by Willard Quine (1961/1953).

Tracing how a consideration of emotional factors in science became dissociated from the romantic critique would require more than a short introductory chapter, with many thinkers deserving attention. The short story is that a loose collection of philosophical trends, unrelated to emotion per se, combined to create an atmosphere in which the relationship between fact and value became interesting again. Two developments that I will briefly explore are the introduction of post-positivist theories among philosophers of science and the diminishment of the behaviorist paradigm in scientific psychology.

In the philosophy of science, the rejection of the theory-observation distinction by thinkers such as Norwood Hanson (1958) served as an important impetus to the re-introduction of values into science. One of the points made by Hanson was that facts are not simply observed; rather scientists adopt theories and models that prescribe what counts as a good fact. Certain kinds of facts only appear when a scientist has a theory to guide her or his observations. Part of learning to be a scientist is learning to see the factual work in a specific way. Sometimes then, the perception of fact is tied to acquired intellectual preferences. Later, Kuhn (1962, 1977) would write about epistemic values, or criteria for what counts as a good theory, such as coherence, comprehensiveness, and simplicity. Later still, feminist philosophers of science such as Keller (1985), Harding (1986) and Longino (1990) introduced the consideration of political and cultural values, further undermining the Humean schism.

The attenuation of radical behaviorism appears to have occurred in proportion to the amplification of cognitive psychology. In concert with psychologists’ new respect for mental states such as attention and memory came a renewed interest in studying both motivation and emotion. The work of Schachter and Singer (1962), Ekman (1972), Lazuras (1982), Zajonc (1980) and Averill (1983) were important in re-establishing the scientific study of emotion. Interest in emotion among neuroscientists such as Panksepp (1998) and LeDoux (1996) has seemed to have even more gravitas, and has provided the study of emotions widespread scientific respectability. Currently, emotion is an important topic throughout the discipline of psychology.
With respectability in science comes respectability in philosophy. Phenomenologists have long had an interest in emotion. Even so, only twenty years ago, intensive work on emotion was limited to a small number of philosophers, such as Irving Thalberg (1977), Robert Solomon (1980) and Amelie Rorty (1980). For various and sundry reasons an interest in emotion began to steadily grow in the philosophy of mind – as exemplified by this very book series and the journal that preceded it. The philosophically-oriented writing of neurologist Antonio Damasio in *Descartes’ Error* (1994) has been particularly influential in philosophy. In arguing that practical reasoning requires quickly eliminating possible response options as ‘bad,’ Damasio makes emotional information processing necessary for rationality, including scientific rationality. Philosophical interest in emotion now crosses several specialty areas, and is becoming something of a specialization itself.

Currently, the field of *affective science* parallels cognitive science as an interdisciplinary endeavor spanning a variety of subfields in psychology, philosophy, the neurosciences, ethology, sociology and anthropology. As in cognitive science, philosophical issues are considered to be respectable, making critical examination of Hume’s famous distinction relevant. Mirroring the interdisciplinary nature of affective science, the topic of fact and value in emotion can, therefore, be approached from a variety of perspectives. Louis Charland and I wrote our call for papers broadly, suggesting topics in several areas including psychopathology, psychology, neuroscience, cognitive science and philosophy. As a perusal of both the chapter titles and the list of contributors indicate, we received papers that addressed all of these areas, and more.

2. Themes of the book

Like everyone else, scholars tend to self-organize into communities and those communities are loosely reflected in the composition of edited books. The community that ended up serving as a magnet for this book, sometimes directly and sometimes indirectly, was the philosophy of psychiatry. Readers will find psychiatric issues discussed throughout the book, although many chapters mention psychiatry not at all.

Because of the wide variety of perspectives contained in what follows, I would like to describe some of themes that appear across the various chapters. Following this analysis, I offer a summary of the each chapter.
A. An adequate understanding of psychology must take emotion into account, and doing so requires considering values.

A corollary of this theme is that, contrary to positivism, a scientific psychology cannot be limited to consideration of facts, and scientists must make normative assumptions in order to proceed with their work. This position is clearly articulated by Charland in the introductory chapter with respect to the different views of Pinel and Crichton in the early days of clinical psychiatry. Using a Heideggerian framework, Hersch also argues that the emotional dimension of experience precedes the experience of objects, and any one who wants to understand consciousness cannot ignore the primacy of emotion. Rudnick claims that psychiatric evaluations of whether or not a patient is competent to make treatment decisions is often based on an assessment of their cognitive capacities when in fact, an assessment of their emotional capacities is equally or even more important. Finally, Landreth explores the sentimental approach to moral psychology, a view that moral intuitions require emotional processing.

B. Claims about emotion are also subject to factual evaluation.

Although an emotionless psychology would be somewhat empty, and values are an important part of understanding emotion, the domain of emotion is not mystical or fictional. Facts about emotion are subject to scientific discovery. Landreth examines the worth of embodied and appraisal theories of emotion with respect to the evidence of neuroimaging data. Faucher and Tappelet draw on research in developmental psychology to critically analyze those theories that view emotions as being primarily hard wired and biologically-based. Ellis also explores possible neurophysiologic processes that may underlie different types of emotion blindness or ‘alexithymia,’ thereby helping validate the types.

C. Emotion has both factual and evaluative aspects. Although neither can be eliminated, it is still important to try to make some distinction between the two.

To say that values influence the perception of facts and facts influence the articulation and experience of value does not mean that facts and values can be merged into a single concept. Distinctions are important. Salmela explores the complicated question of when facts warrant a particular emotional reaction. He argues that non-normative facts are more specific to individual situations and are associated with emotional authenticity. Other situations have a communal aspect and are associated with emotional truth – which has a more normative dimension and is less
subject to re-evaluation in the light of new facts. In exploring dehumanization, Haslam and Loughnan note that one kind of dehumanization involves making factual claims about another group that are empirically false. In these cases, the perception of facts are influenced by biases about an outgroup. Potter explores the moralistic undertone to judgments of inappropriate emotion, highlighting the subtle ways in which beliefs about what counts as emotionally normal are gendered. She argues that gender-laden evaluations shouldn’t be taken to be merely observable facts in some value-neutral descriptive psychopathology.

**D. Philosophical analysis can contribute to a better understanding of the role of emotional factors in defining the nature of psychiatric disorders.**

Jennifer Radden and Nancy Potter explore the emotional dimension of psychiatric disorders, with Radden examining emotional pain as might occur in depression and Potter examining anger associated with Borderline Personality Disorder. They both note that ambiguity in the definition of these symptoms creates a parallel ambiguity in how mental disorders are conceptualized. Rudnick focuses on the effect that emotional impairment has on the ability to choose for those with depression, narcissistic personality disorder and schizophrenia. His exploration has implications for the construct of impairment itself. Salmela examines the important distinction between considering an emotion appropriate and considering it maladaptive. Both Potter and Salmela claim that rather than relying on general norms, consideration of a client’s personal history is important in making evaluations about emotional appropriateness. Ralph Ellis attends to philosophical issues involved in the psychiatric condition of alexithymia. He argues that a proper understanding of consciousness, particularly understanding differences between the intensity and intentionality of emotional consciousness, will help clarify some of the mysteries of this puzzling condition.

**3. Chapter summaries**

In Chapter 1, Louis Charland explores historical issues related to the conceptualization of psychiatric disorders that have an emotional dimension – which he calls affective psychopathology. One way to introduce Charland’s position is to note that an important ethical principle in the mental health field is that professionals should not impose their values on those they serve. This principle is most evident in psychotherapy, but has it roots in medical ideas about illnesses being objective dysfunctions – matters of fact rather than value. Charland considers the dismissal of values in the name of science to be a mistake.
The late 18th century debates about ‘moral treatment’ are conventionally understood to be debates about adding a psychological dimension to the nascent field of psychiatry. Charland claims that this historical reconstruction does not accurately reflect the nature of these debates. In the conventional history, Philippe Pinel is usually hero of the moral treatment story. Pinel is Charland’s hero as well, but Charland argues that many of the more interesting nuances of Pinel’s position have been forgotten due to English translations of his work being abridged, the failure of translators to address how the English term *moral* has multiple meanings in French, and lack of attention to all the major players. Charland contends that the deeper problems that concerned Pinel are also important problems today. Only by ignoring them has psychiatry been able to believe that it solved the vexing problems of affective psychopathology early in its history. Nor are the vexing problems independent of philosophy, in the late 18th century or now. For example, Hume famously argued that reason is a slave to the passions. For Hume, then, ordering the mind required ordering the passions. The passions had both psychological and moral aspects, and ‘order’ was particularly moral (i.e., normative or concerned with ‘ought’). Likewise, Charland claims that Pinel’s moral treatment also had both psychological and ethical dimensions.

Charland points out that the moral treatment debate’s center of gravity was the philosophy-inspired insight that emotions have an etiological role in the development of ‘mental alienation.’ If addressing questions of value are essential for understanding emotions, then the early psychiatrists had a choice of embracing moral concepts in order to grasp mental alienation, or remain purely scientific by settling for a partial understanding. Charland states that Pinel embraced the option of attending to physiology, psychology and morality. A second option, attributed to Pinel in the conventional histories, focuses on physiological and psychological factors only. A third option, adopted by Alexander Crichton, holds that there is no room for either values or psychological factors in the scientific study of emotional psychopathology.

Charland notes that even in the original French, Pinel himself obscured his moral treatment project in order to distance himself from the religious treatment of William Tuke at the York Retreat. It is also likely that, in addition to guild issues related to psychiatry’s desire to be ‘scientific, the sheer difficulty of the vexing problem contributed to its being de-emphasized. Considering moral issues may also open the door to moralism, which is discouraged in professional ethics. Charland concludes that Pinel’s concern with the practical aspects of treatment led him to attend to both fact and value in affective psychopathology, and Crichton’s lack of concern for treatment allowed him to advocate a partial psychiatry, and to believe that with respect to mental illness, facts are all there is.
Mikko Salmela confronts the engaging problem of what counts as valid situation-specific emotions in Chapter 2. ‘Valid’ refers to evaluations that are justified by the facts. Clinical psychologists tend to define validity in terms of behavioral adaptiveness (or functionality). For example, it is adaptive for an alcoholic to be afraid to drink red wine with dinner, but not for an otherwise healthy adult. Some philosophers, on the other hand, are more interested in determining when an emotional reaction is rationally appropriate. To illustrate, it is appropriate for a two-year old to cry if she is separated from her mother in a public place, but not a thirty-year old. What reasons warrant the child crying but not the adult? Salmela notes that clinical psychologists look at cases where emotional validity is a function of the match between the person and the situation and reduce considerations of rational appropriateness to an evaluation of adaptiveness. The adaptive and the appropriate, he says, should not be conflated.

Salmela next distinguishes between two kinds of appropriateness. The first, emotional authenticity is about an immediate fit between a unique individual and a situation. Authentic emotions have an internal, personal warrant. The second, emotional truth requires an individual to reflect on how she or he should respond to the situation while keeping the bigger picture in mind. True emotions have an external, communal warrant. Salmela offers specific criteria for evaluating when an emotional reaction is authentic and when it is true. He concludes by applying his two kinds of rational warrant to the psychotherapy context in order to illustrate the value of considering appropriateness in addition to considering adaptiveness. He suggests that talking about what is rationally appropriate with clients may even offer a better way to address clinicians’ usual concern with adaptiveness.

In Chapter 3, Nancy Potter provides a careful exploration of anger as it occurs in the context of Borderline Personality Disorder (BPD), calling attention to a host of complications that arise in trying to evaluate another person’s anger. Anger here is defined as a moral emotion in which the angry person feels that some wrong has been done. The appropriate response to anger is to try to understand what perceived wrong initiated the anger. Ignoring or criticizing another person’s anger, therefore, minimizes her or his worth as a self-respecting human being.

What is complicated is that although to dismiss a person’s anger risks dismissing her or his moral self-worth, people can also be inappropriately angry. Invoking virtue ethics, Potter notes that we should titrate anger responses so they are proportionate to the situation – both excessive anger and deficient anger are problematic. The DSM-IV, she notes, provides little guidance on this task, treating the BPD symptom of inappropriate, intense anger as a factual matter. Another complication is that even if previous experiences have led a person to develop an angry disposition, not all anger episodes can be interpreted as being about the past; the person may also be responding to the current situation. Potter calls this the doubling effect...
of anger. Rather than seeing anger in persons with BPD as regressive/transferential \textit{a priori}, Potter argues that both past and current precipitants of anger should be acknowledged. She provides some suggestions for distinguishing inappropriate from appropriate anger, and also for deciding when intense anger might be personally warranted (as opposed to socially justified). A final complication is that emotional norms are gendered, and even moderate anger tends to be seen as more \textit{inappropriate} for women in general. This is quite a problem given that most people diagnosed with BPD are women. She closes with some reflections on the therapeutic value of developing a better philosophical understanding of anger.

\textbf{Jennifer Radden}, in Chapter 4, offers a close analysis of the relationship between sensation pain (such as a cut finger) and emotional pain (such as the suffering associated with major depressive disorder). She calls these experiences s-pain and e-pain. Pain is defined as an experience. For example, if I cut my finger but do not feel it, then there is no pain. Sensation pain, therefore, cannot be reduced to tissue damage. Many similarities and differences between s-pain and e-pain are explored. She addresses similarities first. In addition to intensity, a core element of an s-pain experience is unpleasantness, which is clearly an emotional state. In this way s-pain and e-pain are alike. As with other emotional states, the unpleasantness of sensation pain is also mediated by higher cognitive states. Radden next articulates seven important ways that s- and e-pain experiences differ. For example sensation pain involves ‘physical sensation’ and is therefore always localized in the body in a way that emotional pain is not. Also people’s honest first-person reports about whether they are in sensation pain cannot be over-ruled, whereas first-person reports about emotional pain are less privileged.

In her concluding section, Radden raises some questions about the use of pain terms in psychiatry, noting that such use can be ambiguous. For example, terms used in the DSM-IV definition of a mental disorder such as pain, distress and suffering clearly refer to emotional pain, although present distress is also exemplified as “a painful symptom.” She notes that distress in a condition such as a major depressive episode clearly refers to e-pain. In the category of somatoform disorders such as pain disorder, however, painful symptoms-distress also refer to s-pain – such as back pain, headaches or joint pain, but the \textit{clinically significant distress} criterion in pain disorder still suggests e-pain. She notes that such a confusion of pains is deserving of a more careful analysis.

\textbf{Abraham Rudnick} argues in Chapter 5 that psychiatry engages in moral evaluations whenever it is called upon to address issues such as involuntary commitment, the insanity defense, and competence to consent to treatment. The philosophical principles include respecting a person’s choices (autonomy), maximizing benefits and minimizing harm to others (beneficence/non-maleficence), and fair distribution of resources (distributive justice). He points out that the criteria used
to make these moral assessments typically center on cognition. Given that many mental disorders have an important emotional component in addition to a cognitive component, Rudnick proposes to examine the relevance of considering emotional factors in making such moral assessments.

Focusing on the emotional impairment of apathy in major depressive disorder (MDD), he argues that apathy or not caring about one’s self is a primary factor in disrupted competence to approve or refuse treatment. Patients with MDD may therefore have compromised autonomy. Turning to narcissistic personality disorder (NPD), he makes a quite bold argument that the emotional impairment of impoverished caring about others (self-centeredness) may diminish criminal responsibility for people with this diagnosis. He notes the diminished level of responsibility may hold, despite intact practical reasoning on the part of patients with NPD. Discussing schizophrenia, the relevant emotional impairment is the negative symptom sometime called affective flattening, which refers to an emotional emptiness. Rudnick argues that this deficit is related to a diminishment of motivation which interferes with a person’s ability to choose and work toward goals. If so, then psychiatric rehabilitation programs cannot adopt a fully-client centered framework, because some patients with schizophrenia may lack the capacity to choose in general. He concludes with some reflections on how psychiatry will continue to be asked to make these types of moral evaluations, and emphasizes the importance of considering emotional phenomena if such evaluations are to be philosophically sound.

In Chapter 6, Luc Faucher and Christine Tappolet taken on the complicated issue of emotional plasticity, which is the extent to which emotions can be altered by oneself, others, or the community – an important question in moral education particularly if one thinks in terms of virtues. It is a rich chapter that seeks to understand the issue of plasticity in the light of recent research in both affective science and theories of moral development.

Going beyond the usual clichés about nature and nurture both being important, Faucher and Tappolet offer specific models of emotional plasticity ranging from the more hard-wired to the more plastic. They term these models the fully equipped model, the marble model, the avocado pear model, the clay model, the wax model, and the silly putty model. They devote particular attention to the notion that both temperament and basic emotions are hard-wired and present at or shortly after birth, a view that exemplifies the first two of their plasticity models. Consistent with the developmental systems perspective, they suggest that outcomes of emotional development are more variable than is sometimes acknowledged, but not as variable as social constructionists might claim. Both virtue ethics approaches to educating the young and personality change in adulthood, they conclude, are still live options in some cases.
In Chapter 7, Nick Haslam and Stephen Loughnan weave philosophical concepts and empirical data together to explore the various ways in which people attribute emotional experience to others, focusing on how emotional attributions can contribute to dehumanization. Dehumanization involves seeing a target as not fully human – which is associated with perceptions of diminished personhood and diminished moral worth. Haslam and Loughnan point out that there are two typical ways to define human nature. One is to highlight what differentiates humans from other species (usually other mammals) and another is to define the core or essential attributes of humanity. There are, therefore, also two ways to dehumanize others. One is to deny them uniquely human attributes (a sub-human attribution) and the other is to deny them a fully human nature (an inhuman attribution).

They note that social psychological research on *infrahumanization* has been very active in the past few years. For example, research indicates that people have a tendency to simplify the emotions of outgroups. They deny outgroups emotional refinement by attributing them primary but not secondary emotions, such as perceiving lust rather than love or rage rather than moral indignation. With this kind of dehumanization, members of outgroups are conceptualized as being like either animals or children, i.e., less evolved or mature than the ingroup. Historically, the subhuman perception has seemed to flow from colonizers to colonized, masters to slaves, and men to women. In a second type of dehumanization, the outgroup is denied a kind of emotional experience that is considered to be central to humanity such as warmth and empathy. Targets are perceived to be more robotic. For example, women may make these attributions about men; bohemians may make them about corporations and managers.

Haslam and Loughnan’s research has also shown that people may attribute a more authentically human emotional life to themselves than to others. Interestingly, negative emotions related to suffering are often seen to be ‘human,’ and these are also the kinds of negative emotions people are most likely to admit to having. Haslam and Loughnan do not let the self-serving nature of such admissions pass unnoticed. They close with some reflections on the implications that these facts about social perception have for our moral values and our moral behaviors.

Anthony Landreth, in Chapter 8, explores the neuroimaging evidence in support of the sentimentalist theory that moral judgments about ‘wrongdoing’ involve a negative emotional response that is associated with the perception that some norm has been violated. He gives particular focus to the sentimentalism of Jesse Prinz, which advocates a Jamesian thesis that moral judgments co-occur with the perception of bodily changes. Landreth begins by examining the strengths and weaknesses of embodiment theories of emotion (such as Prinz’s) in comparison to appraisal theories of emotion. One challenge faced by embodiment theories is to explain how emotional reactions are intentional – or coordinated with external
events. Appraisal theories do not face this problem because they hold that the evaluation of events with respect to our own interests precede the bodily reactions. He notes that Prinz’s theory does involve a perception of some external event, but it is the physical response that is constitutive of the emotion.

Focusing on neuroimaging methodologies, Landreth next describes what kind of empirical findings should be expected from embodied theories and what kind of findings should be expected from appraisal theories. He reviews the neuroimaging evidence that Prinz claims supports his own theory, but shows that the degree of support is mixed at best. The research results actually indicate that emotional experiences involve frontal and temporal lobe activations of various sorts, and are therefore more consistent with appraisal theories than with embodiment theories. Landreth argues that studies of the orbitofrontal cortex suggest that it tracks our interests in different ways depending on which sub-area is activated. There is some evidence that when moral norms are violated, lateral orbitofrontal activation signals an event to be avoided (punishment-displeasure), providing support to an appraisal-based version of sentimentalism. He closes by articulating a new, pluralistic version of sentimentalism that seeks to accommodate the insights of both embodied and appraisal theories. They key issue, he says, is whether an event is experienced as rewarding (pleasurable) or punishing (displeasureable), not whether the emotional valance is associated with a representation of event connected with our interests or a representation of the state of the body. Better understanding the biological mechanisms underlying reward and punishment, therefore, represents an important research program for the sentimentalist theory of moral judgment.

In Chapter 9, Ralph Ellis focuses his attention on alexithymia, distinguishing two manifestations of the condition. In the first manifestation, people are affect blind. Observers may be able to attribute emotion and motivation to their actions, but those with alexithymia have no conscious experience of emotion. In the second type, persons with alexithymia cannot construe their affect in an emotional framework. They do not experience affect as emotion, or in other terms – they lack psychological mindedness.

He begins his exploration by trying to get a better grasp of the experience of alexithymia by seeking out more normal experiences of having emotion but only being partly aware of them or being mistaken about their meaning. One of the interesting ideas Ellis illuminates is to show that emotion is a complicated phenomenon – and we tend to highlight the rare times when we have emotional clarity, but on a moment to moment basis, we usually lack clarity. He also distinguishes the intensity of feeling from the intentionality of emotion. He claims the bio-cognitive mechanisms for registering intensity differ from the mechanisms for registering intentionality. Intensity involves interoceptive imagery while intentionality involves sensorimotor imagery.
Ellis next explores the possible cognitive and biological bases of the two kinds of alexithymia. The affect blind type of alexithymic is unable to feel. Feeling is related to the intensity mechanism, and is the kind of passive experience of a state of the body discussed by thinkers such as Damasio. Consciousness, however, as Ellis has long claimed, is not merely passive or afferent. It is also active or efferent. Taking his cue from the enactivist theories of Newton and Jeannerod, he notes that to be conscious of an object is also to be aware of how we can interact with the object. Forming object images involves forming images of actions we would perform in relation to those objects (which is a non-conscious quick time-frame process). This kind of emotional consciousness is what may be disturbed in the second kind of alexithymia, says Ellis. With malfunctioning intentional mechanisms, people are aware of feelings in the body, but don’t know what they are about, or what they value, or how those feeling are related to each other.

Returning to the grand theme of the necessity of emotion, Ed Hersch concludes the volume by offering a phenomenological analysis which highlights the ineliminable evaluative aspects of our pre-reflective experience of the world. He emphasizes that this pre-reflective mode of experience initially perceives objects in terms of their immediate value or how they might be good-for-us. In other words, our “cares” organize our experience from the outset. Hersch follows Heidegger in referring to the “care-structure” of experience. According to Hersch, all experience is care-laden, and all caring is emotional. He claims that such ‘primordial’ perceptions precede our secondary apprehension of things as decontextualized objective facts ‘out there’, which are the things (or the level of abstraction) typically studied in the natural sciences.

Hersch argues that if understanding what is really going on is a goal of psychiatrists and psychologists, then the primacy of the emotional dimension in human experience means that the scientific study of human psychological processes shouldn’t be limited to methodologies most appropriate only to the study of uncaring ‘things’ (which lack that dimension). He maintains that attention to the caring-emotional-evaluative dimension to all such processes is crucial if we are to provide an adequate appreciation or understanding of them.

Hersch rejects both what might be called (a) appendage theories of emotion and (b) claims that emotions represent downstream affective events relative to upstream cognitive events. He argues that gaining more accurate and complete answers to a lot of the interesting ‘why questions’ about emotion in cognitive neuroscience and psychiatry will require careful consideration of the care-structure of experience.
References

A moral line in the sand

Alexander Crichton and Philippe Pinel on the psychopathology of the passions

Louis C. Charland

The passions are to be considered, in a medical point of view, as part of our natural constitution. They are mere phenomena, the natural causes of which are to be inquired into … (III. 1. 99)

Alexander Crichton
An Inquiry into the Nature and Origin of Mental Derangement (1798)

1. Introduction

Psychopathology is the science of what mental illnesses are. Affective psychopathology – or, alternately, the ‘psychopathology of affectivity’ – is the branch of psychopathology devoted to the study of mental disorders that implicate mental states associated with moods and emotions and what used to be called ‘passions’ (Berrios 1985). Some segments of the history of affective psychopathology have been skillfully traced (Berrios 1985, 1996). However, there is one episode in that history that has not received the attention it deserves. It concerns medical writers in France, England, and Scotland, during the latter half of the eighteenth century. The issue at stake is whether affective psychopathology should include or exclude questions of morality.

The purpose of the present discussion is to revisit and reformulate this pivotal debate in the history of affective psychopathology. One reason for the lack of attention to the debate may simply be that is has largely been taken to be settled. Thus, it is widely believed that affective psychopathology should steer clear questions of value, especially morality. Another reason may be the fact that the original historical debate is obscured by a partly defunct terminology that does not translate neatly into modern terms. In its original formulation, the debate is stated in
terms of the question whether ‘moral’ considerations should play a part in the psychopathology of affectivity. The problem is that during the period we are concerned with, the term ‘moral’ is used in a number of different ways in both French and English.

2. Problems with ‘Moral’

Among medical writers of the late eighteenth century, the term ‘moral’ is often taken to refer to what is mental as opposed to physical. This intended meaning of the term is usually translated as ‘psychological’ (Porter 1987; Shorter 1993; Wiener 1990). At times, what is ‘moral’ in this general psychological sense is further restricted to what concerns value, especially morality. To complicate things, among these psychological usages of the term, there are also cases where ‘moral’ is specifically used to refer to what concerns the passions or emotions only, though such cases are rare (Grange 1961, Esquirol 1805). Finally, during this period, it is also common to find references to the ‘moral sciences’. This very general sense of the term goes beyond what is considered to be ‘psychological’. It extends to human phenomena and activities generally. In that wide sense, economics and sociology are ‘moral sciences’, though of course they differ from psychology, which is the science of the mental. Note, moreover, that psychology in this scientific sense is usually divorced from normative inquiry into morality. Thus the fact that something is ‘moral’ in the psychological sense should not be taken to imply that it is also ‘moral’ in the ethical sense. That is a distinct and logically independent sense of the term ‘moral’.

The historical meaning of ‘moral’ we are primarily concerned with is the one that concerns psychological matters that fall within the sphere of morality. There appears to a consensus today that ‘moral’ factors in that sense most certainly do not play a substantive role in psychopathology, affective or otherwise. Nor should they. Indeed, it is taken as an article of faith that, if it is to remain scientific, psychopathology must steer clear of questions of fundamental value, especially morality. Psychiatrists and psychotherapists should not pass moral judgment on the goals and activities of their patients or aspire to be like priests. This theoretical stance is so entrenched today that it may be hard to imagine a time when the issue might have been debated. But there was such a time. And although nowadays the issue may seem to have been conclusively settled in favor of a strictly scientific vision of psychopathology, there are cracks in the foundation.

The cracks are most evident in the domain of affectivity. But obviously they are easy to miss, since the issue is seldom seriously debated, even in leading historical discussions of the period. To see where the cracks begin, one must first look in the
right place. This is the effort to construct a purely scientific or ‘medical’ psychopathology of affectivity. It must be conceded that there have been laudable efforts to write the scientific history of modern psychopathology and psychotherapy (Ellenberger 1970; Healy 1990). But they fail to identify the particular cracks we are concerned with, and cross over them without sufficiently exploring the theoretical gulf they create. The cracks start to appear in the early days of ‘moral treatment’, an innovative therapeutic movement associated with medical men like Philippe Pinel (Pinel 1801), Vincenzo Chiarugi (Chiarugi 1793), Johann Christian Reil (Reil 1803), and the lay Quaker William Tuke (Tuke 1813). This coincidentally is the time when the term ‘psychiatry’ (psychiatrie) is first formally introduced by Reil. It is also a time when the passions assumed a prominent place in philosophy, and subsequently medicine, especially psychopathology.

3. Hume and Rousseau on the passions

Let us then turn our gaze to end of the eighteenth century, when the young discipline of psychiatry was just embarking on its fight for professional autonomy and recognition as a legitimate branch of medicine. There are important philosophical developments in the area of the passions that took place in the decades leading up to this period. These undoubtedly influenced the course of related discussions in medicine. Consider, for example, the case of David Hume.

In his new ‘science’ of human nature, Hume attempted to explain the operations and contents of the mind on the basis of observation and experience. Among other things, he tried to explain the operations of the passions in associationist terms inspired by a rather wide conception of ‘natural philosophy’. That effort was bound to stimulate questions about the status of the passions in medicine, most notably psychopathology. Yet in his work on the passions, Hume bypassed the natural sciences entirely in order to concentrate instead on the passions as elements of the mind. In other words, he was not concerned with the passions as a natural scientist; with their anatomy or physiology. Instead he was concerned with their ‘moral’ aspects; not only in the wide, psychological, sense of the term, but also in the narrower sense of morals and morality. Could this ‘moral’ dimension of the passions be accommodated by medicine, which traditionally was more directly aligned with the natural rather than the human sciences? Such questions go to the heart of psychiatry’s ‘hybrid nature’, which is a consequence of ‘the fact that psychiatry unsteadily straddles the natural and human sciences’ (Berrios 2007, 469).

Significantly, during this period the relationship between passion and reason was also being reconsidered. If, as Hume said, reason is a slave to the passions, then when the passions go astray, so probably does reason. Conversely, and more
importantly, before reason can govern mind and behavior in an orderly manner, the passions must first themselves be organized in an orderly fashion. Reason is not irrelevant in this scheme. It functions as the ‘eyes and ears’ of the passions, supplying them with information about the world required for their function. However, ultimately reason cannot move us to action; that primarily is the role of the passions. In Hume’s philosophical system, the organization and ordering of the passions in human life is primarily left to morals and history. It depends on brute facts about human nature, especially ‘sympathy’; the human ability to empathize with the feelings of others.

Hume’s philosophical legacy in the area of the passions raises puzzling questions for any medical attempt to explain their operations in accordance with the principles of physiology and anatomy. What will an explanatory scientific model of this sort appeal to when it inquires into the question how to turn mental disorder into order? What, indeed, is meant by order and disorder? And what is healthy about order as opposed to disorder? Hume’s analysis suggests that we must turn to morals and historical tradition for such answers. However, these explanatory projects clearly fall outside the bounds of medical science, strictly understood. Must we then limit medical inquiry into the passions accordingly? The question is unavoidable.

Leading up to and well into this period, there is also the influence of Rousseau. In *Emile*, he reminds us that without the passions human life and reason essentially have no point (Rousseau 1762). In themselves, the passions are essentially good and healthy. It is society that corrupts. Like Hume, Rousseau also recognizes that the passions can interfere with and disrupt the operations of reason. And like Hume, he insists that the passions play an essential role in the organization and regulation of human life in virtue of their inextricable connection with morals. The passions then can certainly cause problems for reason. But to eliminate them is out of the question. They are essential not only to morals but also for human life and reason generally.

These and other philosophical efforts to rehabilitate the passions undoubtedly exercised a powerful influence on medical thinkers of the day. Specific lines of influence are hard to establish, but circumstantial evidence strongly suggests the influences are there. Especially important is the fact that both Hume and Rousseau stress the close connection between the passions and matters of value and morality in the overall organization of mind and behavior. That raises in an acute way the question whether a medical science of the passions ought to ignore questions associated with value and morality, or alternately must include them. There is also the further question of how the passions figure in mental health: the mental order that mental disorder disrupts. It did not take long before medical writers turned
their attention to the passions and started to confront such questions, which evidently did not appear out of nowhere.

Not surprisingly, towards the end of the eighteenth century, in ‘mental science’ as in philosophy, the passions started to assume a central role. There was a marked shift away from intellectualist conceptions of insanity that focused primarily on erroneous judgments and associations of ideas, to the special role of the passions as embodied visceral forces in the manifestation and etiology of madness. In this emerging new vision of psychopathology, the seat of madness no longer lay simply in the brain, the seat of reason. It also lay in the viscera, the seat of the passions. However, the fact that the passions were so intimately bound up with questions of value and morality created a dilemma for medical scientists of the passions. If the passions are inextricably linked with questions of value and morality, then they seem to fall outside the sphere of medical science, strictly speaking. So to include this aspect of the passions in medical science threatens the scientific credentials of that science. On the other hand, to banish them risks being grossly untrue to the phenomena. Neither option is very inviting; a true dilemma.

This is the point at which our two main protagonists – the scientific fathers of modern affective psychopathology – enter the picture. They are Sir Alexander Crichton (1763–1856), a Scott, and Philippe Pinel (1745–1826), a Frenchman. Both are keenly aware of the above dilemma. Crichton, in fact, tackles it head on, while Pinel tries to work his way through it.

Like many of his contemporaries, Pinel’s thought is imbued with the romantic ideals of Rousseau. On his side, Crichton is evidently inspired by his fellow Scot, Mr. Hume, whom he refers to repeatedly in his major work. Our two medical pioneers are clearly intimately knowledgeable about the dominant philosophical writings of their time. Yet both take radically different routes in response to those philosophical challenges. Pinel’s strategy is to medically embrace Rousseau’s insights on the passions and combine them with the best science and physiology he knows. Crichton’s strategy, instead, is to issue a scientific manifesto designed to undercut any encroachment of morals into affective psychopathology.

4. Crichton and Pinel on the passions

During these formative decades of modern psychiatry, both Alexander Crichton and Philippe Pinel argued that the passions played a central role in psychopathology. They were the first medical writers of their day to do so explicitly; a fact that is immediately evident from the tables of content of their major works. This affective turn in psychopathology challenged the dominant intellectualist paradigm in vogue at the time, according to which mental illness was primarily a cognitive, or
‘intellectual’, matter of disordered ideas. Both Crichton and Pinel stressed the importance of the physiological principle of irritability and the importance of the viscera in accounting for the bodily workings of the passions, while at the same time acknowledging their mentality. However, they strongly disagreed about the place of value and ethics in their plans for a new psychopathology of the passions. Crichton insisted that it was imperative to ignore the ethical and other ‘moral’ dimensions of the passions. He insisted that they be treated as ‘mere’ phenomena only. Pinel disagreed and argued precisely the opposite. Although he agreed with Crichton on the embodied character and the physiological basis of the passions, he also believed that the new psychopathology of the passions had to include elements of value and especially morals along with psychological terms and notions generally. For Pinel, then, ‘moral’ often means what is mental and what, additionally, has to do with morals. But not always, since there are many instances where psychopathology is only concerned with states that are psychological and morals are not involved. What makes the interpretation of Pinel so complicated is the fact that the manner in which these issues and this distinction arise in his work must be addressed on a case by case basis. His is a highly nuanced theory when it comes to this matter of ‘le moral’ and ‘la morale’.

Historians have largely overlooked Pinel’s debt to Crichton on irritability and his disagreement with the Scottish physician on the place of ethics in affective psychopathology. An intriguing consequence of this last point is that Pinel’s ‘moral treatment’ (traitement moral) turns out to have important ethical dimensions. Pinel saw no contradiction between acknowledging the psychological aspects of moral treatment (le moral) while at the same time insisting it had important ethical presuppositions (la morale). Moreover, this did not prevent him from concurring with Crichton in espousing a thoroughgoing materialism that stressed the embodied physiological character of the passions. Whether Pinel succeeded in his attempt to reconcile these two premises in his new ‘philosophical medicine’ (médiciné philosophique) is an important question. His plan for a new medicine of mental illness was based on the rigorous clinical application of Hippocratic observational principles, and a thoroughgoing scientific implementation of the sensationalist epistemology of Locke along the lines articulated by the French philosopher, Condillac. His proposed psychopathology of the passions was thus a beguiling compendium of French sensations and Scottish passions seen through the lens of Greek clinical observation.

So our case study starts towards the end of the eighteenth century, a time when the psychopathology of affectivity underwent remarkable developments. The dominant intellectualist picture of madness inherited from Locke was transformed – one almost wants to say transfigured – by the addition of the passions. Prior to this affective turn, the received view of madness was largely intellectualist in nature
It was ‘essentially delusion, and delusion sprang from intellectual error’ (Porter 1987, 19). This made madness primarily ‘a fault in cognition, rather than in will or passion’ (Porter 2002, 60). Of course, since the ancient Greeks, the passions have often figured among the causes of madness. ‘Raving delirium’ was the hallmark of madness then (Healy 2002, 12). It was also widely acknowledged that excesses in passion could drive one mad (Simon 1975, 90, 185).

The manner in which the passions came to be incorporated into eighteenth and nineteenth century theories of madness marks a radical departure from earlier views. One important factor is the appearance of faculty psychology (Radden 1996). Philosophers like Wolff, Kant, Reid, Stewart, and others, argued that the operations of mind could be subdivided into different faculties or ‘powers’. Pinel, for example, sometimes speaks of lesions of the intellect, or the affective faculty, or the will (Pinel 1809). In tandem with these developments, there emerged a recognition that affectivity might constitute a special psychopathological domain of its own. The passions, in particular, assumed a central place in the new explanatory framework. A good example of the new theoretical orientation can be gleamed from the title of one of the key works of the period. This is Jean-Etienne Esquirol’s *Des Passions considérées comme causes, symptômes et moyens curatifs de l’aliénation mentale* [The passions considered as causes, symptoms and means of cure in cases of insanity] (Esquirol 1805). In this seminal work, Esquirol explicitly acknowledges his debt to Philippe Pinel and Alexander Crichton on the topic of the passions and their relation to mental illness:

> Few authors have studied the relationship of mental alienation to the passions. Crichton offers exact ideas on the origins and development of the passions and their effect on the organism. Professor Pinel agrees with him, regarding the passions as the most frequent cause of upset of our intellectual faculties. (Quoted in Weiner 1990, 385.)

Philippe Pinel (1745–1826) is certainly the more famous of our two protagonists. His application of the Hippocratic method of clinical observation to the study of mental illness, his introduction of statistical methods to the evaluation of treatment outcomes, and his innovative ‘moral treatment’, are regularly cited as landmarks in the birth of modern clinical psychiatry (Healy 1997; Goldstein 2001; Pigeaud 2001; Porter 2002; Shorter 1993; Scull 1993; Weiner 1999). Pinel is also recognized for having made important contributions to the psychopathology of affectivity (Berrios 1985, 1996; Goldstein 2001; Porter 2002; Weiner 1990). Yet little substantial has been written about this specific aspect of his work. Although moral treatment is usually mentioned, its connection to the passions is left largely unexplained. There is therefore a significant lacunae regarding the exact nature of Pinel’s contribution to affective psychopathology. While his focus on the passions
is recognized, even the best exegetical discussions do not delve into the details of which passions are associated with which pathologies, and why (Pigeaud 2001; Weiner 1990).

Our second protagonist is Pinel's Scottish contemporary, Alexander Crichton (1763–1856). He has almost entirely escaped notice in the history of affective psychopathology.

It is openly acknowledged that both Pinel and Esquirol were heavily influenced by Crichton's writings on the passions (Goldstein 2001; Pigeaud; Weiner 1990). However, the details of Crichton's specific contributions to affective psychopathology have yet to be fully explored. His views on the physiological concept of irritability and his theory of how the passions function in the causation of mental illness have not received the attention they deserve. They point to a very different conception of the nature of affective illness and its treatment (Charland forthcoming).

In a telling remark on the state of contemporary scholarship on Crichton, one leading commentator notes that 'his biography … and his major work remain to be analyzed' (Weiner 1990, 338). At the same time, it is recognized that Crichton 'broke new ground' and made important original contributions to psychiatry (Hunter & MacAlpine 1982, 559). It is interesting that the portions of Crichton's work which deal specifically with the passions were simply ignored in early efforts to translate it into French (Weiner 1990, 366). Even more noteworthy is the fact that no official French translation was ever published in Crichton's lifetime (Weiner 1990, 381). Pinel's extensive comments on Crichton's theory of the passions were, in turn, simply omitted in the English translation of his Traité. Thus, 'strange as it may sound, Crichton may not have known of Pinel's extraordinary praise for his book' (Weiner 1990, 383). The end result of these and other historical accidents is that Crichton's discussion of the passions appears to have fallen by the wayside in the history of psychiatry. Perhaps this is why 'until now, no one has explored the relationship of Pinel to Crichton' (Weiner 1990, 366). Again, there are important lacunae in the relevant literature that need to be remedied.

5. **Crichton's manifesto**

The crux of our case study lies in a set of remarks that Pinel makes on Crichton's conception of the passions and their role in mental illness (Pinel 1801, xxj-xlj). These are contained in the *Traité médico-philosophique sur l’aliénation mentale, ou la manie* (Pinel 1801). They also reappear, but in much reduced form, in the second and much enlarged edition of that work, the *Traité médico-philosophique sur l’aliénation mentale* (Pinel 1809). Crichton's contributions to affective psychopathology are contained in his *Inquiry into the Nature and Origin of Mental Derange-
ment (Crichton 1798). While Pinel is admittedly generous in his praise for that work, there are also important points of disagreement which tend to be overlooked. Therein lies the principal motivation for this discussion.

There are two especially important features of Crichton and Pinel’s work on the place of the passions in affective psychopathology that call for special scrutiny. The first has to do with the physiologically embodied character of the passions and how this is reconciled with their mentality. Both thinkers share a profound interest in the physiology of the passions and stress its connections to physical and mental health. Their combined views on the physiological character of the passions probably represents the most important scientific advance in the area since Descartes’ pioneering treatise, Les Passions de L’Âme (Descartes 1650). The physiological concept of ‘irritability’, in particular, is especially important to both authors (Weiner 1990, 375, nt. 136). Yet it is invariably overlooked in discussions of their work and also typically ignored in many histories of physiology. This is unfortunate, since there is much that we can learn from that concept today (Hall 1975; Pagel 1967; Temkin 1964).

Irritability in this context is primarily a notion of physiology and not directly tied to temperament as the modern colloquial meaning of the term suggests. It is a ‘principle of motion’ which is “distinct from nervous energy, and also distinct from the principle of mechanical motion” (Crichton 1798, Vol. 1, 6). Muscular contractions in organisms with lesions of the nervous system are often cited as examples of irritability. One famous experiment involves the elicitation of contractions by applying noxious stimuli to the body parts of frogs where the ‘sensorium has been removed by decollation’ (Ibid., 4). Crichton states that “all muscle fiber is irritable” although he also careful to emphasize that this does not mean that “all irritable bodies and irritable parts of bodies are muscular and fibrous” (Ibid., 10). In some respects, physiological irritability is a precursor of the modern notion of valence (Charland 2005). This is the idea that affective states can be classified as either ‘positive’ or ‘negative’, now a central assumption of modern affective science and psychopathology. A full defense of this thesis, though, must await another occasion. It is mentioned here only to help orient our discussion.

To sum up, both Crichton and Pinel agree that physiological irritability is central to the operation of the passions and that the passions are embodied in virtue of their visceral nature. Both then are materialists regarding the passions in some important sense. However, despite their shared materialist commitments, our two protagonists ultimately take radically different routes on how to construe the relation between the passions and psychopathology. This is the comparison that will especially occupy us. Now it is sometimes observed that Pinel is influenced by Crichton on the topic of the passions, even though their agreement on the importance of irritability and viscerality is not mentioned (Goldstein 2001, 95–96;
Pigeaud 2001). However, more importantly, what is typically missed is that they strongly disagree on the place of the evaluative and ethical character of the passions in affective psychopathology.

Part of the difficulty here lies in ambiguities that surround the historical term ‘moral’. Crichton and Pinel both agree that the passions are ‘moral’ phenomena in the sense that they are partly psychological. This is how the term ‘moral’ is usually understood in these medical discussions. However, the passions can also be considered to be ‘moral’ in a narrower, ethical sense. This is where the disagreement lies. Pinel argues that psychiatry must acknowledge and integrate this ethical ‘moral’ dimension of passions, while Crichton says precisely the opposite. Only one commentator makes this point clearly (Pigeaud 2001, 270–273).

Crichton’s statement about the place of value and ethics in affective psychopathology is stated at the outset of his famous text. It reads like a manifesto for all future affective psychopathology:

The passions are to be considered, in a medical point of view, as part of our natural constitution, which is to be examined with the eye of the natural historian, and the spirit and impartiality of a philosopher. It is of no concern in this work whether the passions be esteemed natural or unnatural, or moral or immoral affections. They are mere phenomena, the natural causes of which are to be inquired into … (III. 1. 99).

In these opening remarks, Crichton also makes it clear that it is fundamentally in virtue of their physiological nature that the passions influence our mental and physical health:

… [T]hey produce constant effects on our corporeal frame, and change the state of our health, sometimes occasioning dreadful distempers, sometimes freeing us from them – these facts are to be carefully observed, examined, and enumerated. They produce beneficial and injurious effects on the faculties of the mind, sometimes exalting them, sometimes occasioning temporary derangement, and permanent ruin …(III. 1. 99).

Thus, in the Inquiry, Crichton’s interest in the passions is ‘principally confined to a physiological and medical point of view’ (III. I. 99). He grants that ‘moralists and metaphysicians have written copiously on the subject’ but counters that their views ‘are of no use whatever to a medical inquirer, except inasmuch as he himself is concerned in the morals of the community he lives in’ (III.1.98). To these observations, he adds the claim that ‘the moral effects of the passions throw no light on the diseases of the mind’ (III.1.133). To understand those diseases, we must turn to physiology.
6. Pinel’s moral treatment

In stark contrast with Crichton’s explicit proposal for an affective psychopathology without value, Philippe Pinel offers a vision of an affective psychopathology that aims to be medical and scientific but at the same time includes value. Yet his stance on the role of value and morality in affective psychopathology is sometimes hard to discern. Like Crichton, he is a medical man, committed to scientific explanation in accordance with explanatory practices the natural sciences. Yet it is this same scientific commitment that forces him to consider and include elements of value and even morality in his overall conception of affective disorder and its treatment.

As a man of science, Pinel is wholeheartedly committed to the art of careful and detailed clinical description. This involves trying to ascertain the causes of mental illness and recording these as accurately as possible. These causes turn out to be overwhelmingly ‘moral’ in nature. Unrequited love; unrestrained lust and greed; grief and disappointment; bankruptcy and economic hardship; fears of political betrayal and condemnation under the revolution; all are ‘moral’ causes of ‘mental alienation’ (Pinel 1809). Of course, it is due to their physically embodied characteristics that these moral factors cause changes in the workings of the mind and viscera, resulting in madness. But according to Pinel, affective psychopathology must employ and resort to the explanatory ‘moral’ vocabulary of the passions in order to correctly and accurately record these causes of alienation. This is one way in which considerations of value and morality form part of his affective psychopathology: the language in which these phenomena are couched – ‘moral’ terms and notions – is essential to proper explanation and description at the level of the causes of mental alienation. Simply speaking in terms of natural causes and physical states of the body is not sufficient.

In Pinel’s affective psychopathology, the language of treatment is equally couched in moral terms and notions that directly or indirectly implicate matters of value and morality. But here the situation is slightly more complicated. One problem is Pinel’s effort to distance his medical conception of moral treatment (traitement moral) from other less ostensibly precedents of the time; notably, the benevolent therapy practiced at William Tuke’s Retreat for Insane persons in York (Tuke 1813). This leads him to selectively emphasize or deemphasize those aspects of moral treatment that revolve around issues of value and morality. This makes it easy to miss the contribution of value and morality to his conception of moral treatment, which is often understood to be simply or merely ‘psychological’ and largely value free.

So while Pinel is careful to try and distance his medical approach to moral treatment from the largely religious therapy practiced at the Retreat, he often obscures the ethical implications and presuppositions of his own moral treatment.
The point is not lost on Michael Foucault, one of Pinel’s most astute critics. According to Foucault, Pinel’s traitement moral is just as ethically laden as Tuke’s moral treatment (Foucault 1976, 576–632). It is in fact an irony of history that while the Retreat is often cited as the birthplace of ‘moral treatment’, the English expression is in fact a translation of Pinel’s French introduced by Samuel Tuke in his later Description of the Retreat (Charland 2007). There is no evidence that the retreat’s founder, William Tuke, ever used the expression ‘moral treatment’, or that he was even remotely aware of Pinel and his work. Writing almost twenty years after the establishment of the Retreat by William, Samuel Tuke, William’s grandson, redescribed its religious philosophy in Pinelian terms to promote its success to the medical world.

Pinel’s decision to include the ethical dimension of the passions in his account of their nature seriously jeopardizes the medical status of his proposed ‘moral treatment’. Many important historical discussions of his moral treatment do not make or sufficiently stress this point (Goldstein 2001; Healy 1997; Porter 2002; Scull 1993; Shorter 1993; Weiner 1990). In fact, the scientific status of Pinel’s moral treatment is quite ambiguous. This is because, on Pinel’s view, ‘moral’ treatment is not simply a matter of intervention at the level of individual and social psychology (le moral). It is also informed by ethics (la morale). It follows that Pinel’s conception of the aims and goals of psychiatry is not as scientifically narrow or clear-cut as his designation as the father of clinical psychiatry suggests. Like Cicero, whom he admired, he was also a moral critic and reformer. Moral treatment then is not purely psychological, as is commonly supposed. In some respects it is also ethical (Charland 2003, 2004, 2007). This emphasis is reflected in Pinel’s admonition that all his readers should read Cicero’s Tusculan Disputations, a key work in the history of Stoic ethics and the passions (Pinel 1801, 80–81; 1809, 12, note 1).

There are other significant examples of Pinel’s commitment to ethics in his plans for a new ‘philosophical medicine’ (médicine philosophique). Surprisingly, even though Pinel is often quite explicit about the relationship between mental health and ethics, these remarks are typically overlooked. One of them comes at the very end of the 2nd edition of the Traité. There, Pinel suggests that lapses in ethics are directly implicated in the rise of mental disorder, and that ethical behavior (une saine morale) is an important component in mental health and illness. Medicine can contribute to ethics, he says, by showing how social and personal lapses in morality can often lead to mental illness:

... on ne peut refuser a la médicine l’avantage de concourir puissamment au retour d’une saine morale, en faisant l’histoire des maux qui résulte de son oubli ...
(Pinel 1809, 492). [... medicine should not be denied the advantage of contributing to the return of a sound morality, which it does by telling the history of the ills that result from its neglect.]
Finally, it is worth emphasizing that psychiatric education today might be very different if we followed Pinel’s injunction that all those who treat mental illness should carefully study and ponder the writings of Cicero. It is easy to forget that his distinct brand of ‘philosophical medicine’ stressed the ethics of Cicero as much as the logic and epistemology of Locke and Condillac.

7. Crichton’s legacy

Since Crichton drew his moral line in the sand and banished morals from the psychopathology of affectivity, psychiatry has largely ceased to debate such matters. His manifesto effectively cleared the road for a thoroughgoing biological psychiatry of affectivity, now perhaps the dominant approach to the topic. Yet Crichton’s legacy in this area is not without its own philosophical problems, which also remain unresolved to this day. The problems are not negligible and raise some doubts about the integrity of his program as whole.

Crichton clearly championed a vision of affective psychopathology that was reductionist in at least one important sense. He drew a moral line in the sand and argued that the psychopathology of affectivity should concern itself with the passions as natural phenomena only, and ignore their ‘moral’ dimensions. By this last term he apparently meant the psychological aspects of the passions, and no doubt anything tied to morals. In the end, according to Crichton, ‘the moral effects of the passions throw no light on the diseases of the mind’ (III.1.133).

Crichton’s focus on physiology and the causal explanation of affective disorder might lead one to suspect he was ultimately a materialist of sorts. Yet he adamantly rejects this philosophical stance and utters harsh criticisms of a major materialist of his time, namely Priestly. Referring to Priestly, he writes:

> Of all species of materialism, that one appears to me to be the most completely absurd which is founded on the supposition that brain and mind are one and the same substance. Yet this doctrine is embraced by one who has added more useful facts to science, and more ornaments to philosophy. than almost any other modern writer (Crichton 2007, 491)

The mind and its mental operations, then, are autonomous and irreducible to ‘matter’. For Crichton, materialism about the mental is out of the question. On the contrary, according to him, the ‘mind’ is crucial and indispensable to psychopathology. Indeed, one of the chief aims of his *Inquiry* is to introduce the ‘philosophy of mind’ of Hume, Locke, and other ‘British psychologists’ to psychopathology. He believed that studying these writings on the operations of the mind would aid the physiologist interested in the medical study of mental derangement by helping
him understand his own mind, something that had to come first (Hunter & MacAlpine 1982, 559). The problem obviously is that this seems totally inconsistent with Crichton’s manifesto that affective psychopathology should dispense with ‘moral’ considerations entirely.

One way to resolve the apparent contradiction is to view Crichton’s commitment to the philosophy of mind as a pledge to the relevance of the mind to ‘descriptive psychopathology’, which is just one aspect of psychopathology; namely, the attempt to identify and catalogue mental symptoms (Berrios 1996). Psychiatric historian German Berrios explains the relevance of descriptive psychopathology to psychopathology this way:

The construction of psychiatry during the nineteenth century occurred under the ægis of medicine. Because of the complexities exhibited by the phenomena of madness, the conceptual structure of psychiatry and its epistemological handles were (rather hastily) made out of fragments taken from contemporary natural and human sciences. Ever since, the former have informed psychiatry of how the target behaviours may be inscribed in the body (currently research is focused on the brain); and the latter have provided the criteria for their very selection and configuration. The natural sciences by themselves cannot create new categories of ‘mental disorder’, nor can any somatic footprint be sufficient to define a mental disorder. This means that the human sciences have epistemological primacy. In other words, a mental disorder cannot be formally ascertained by studying the brain of the patient but by the enactment of the social act of ‘diagnosis’ which is the outcome of an emotional and epistemological partnership between helper and sufferer (Berrios 2007, 470).

Here then is one way to reconcile Crichton’s adamant refusal to admit the ‘moral’ in affective psychopathology, with his insistence that the study of the mind is also sometimes directly relevant to psychopathology. According to this interpretation, ‘moral’ considerations are relevant to affective psychopathology in the domain of descriptive psychopathology, but not in the causal explanation and understanding of mental disorder. None of this settles the question how exactly the mental relates to the physical. We are also left with the question of what exactly is meant by the term ‘mind.’ The former question is one that Crichton seems to ignore. He does consider the latter question but only to abdicate philosophically. Regarding this latter question, he confesses that he has no answer, and also even suggests he simply does not care:

After the physiologist has carried his inquiry as far as possible into the nature and effects of irritability, and nervous sensibility, as well as into those of the mechanical and chemical powers which operate on the human body, he still meets with many phenomena which he cannot account for by the known influence of these agents … He grants that the mind is not an object of external sense; he grants that
it does not occupy space; yet the belief of its existence is forced upon him by the consciousness of what passes within himself; he knows not what to decide, but thinks it is loss of time to dispute about words. He is convinced that the true manner of studying the human mind is by beginning with the study of the human body; and he is persuaded that whoever studies it deeply will be convinced that the mind is totally distinct from that part which is evident to the senses. In other respects it is of very little importance to him whether it be called a matter sui generis, or an immaterial substance. Neither of these terms explain any thing to him (Crichton 2007, 497).

What are we to make of such a philosophical stance? No doubt, contemporary defenders of biological psychiatry and a purely scientific approach to affective psychopathology will probably wish to champion Crichton and add his manifesto to their theoretical arsenal. At the same time, it is probably fair to say that the philosophical problems that bedevil Crichton’s attempt to construct a purely scientific affective psychopathology are still with us today. As we have seen, there are plausible strategies for addressing some of the apparent contradictions in Crichton’s program, but that is far from saying the debate is settled. On the whole, Crichton’s legacy on the status of the ‘moral’ in affective psychopathology is mixed. There are bold and dazzling visionary pronouncements, but they come at a price, leading to mixed results.

8. Pinel’s legacy

The enduring value of Pinel’s legacy on this question of the status of the moral in psychopathology is easier to ascertain. In Pinel’s work, there is an important interplay between theoretical work on nosology and psychopathology and rigorous therapeutic testing of new and existing treatments. On his side, Crichton seems completely unconcerned with issues of treatment and remains at the level of theory. It is especially in the area of treatment that ‘moral’ considerations emerge as paramount for Pinel. What he found is that treatments based on ‘moral’ principles and notions in the widest psychological sense were often extremely successful. The care with which Pinel documented treatment interventions and calculated outcomes is one of his most important contributions to clinical psychiatry as a discipline. What is less recognized is the fact that his moral treatment was also often inherently moral in an ethical sense. Not always. But often enough that it is remarkable that this feature of his moral treatment has attracted so little attention.

It must be emphasized that Pinel’s moral treatment was never intended to lead to or result in religious indoctrination, something that undoubtedly occurred at William Tuke’s Retreat. In fact, Pinel strongly believed that religious excess was
among the most pernicious causes of mental alienation, and patients were regularly discouraged and even sometimes forbidden to indulge in religion while in treatment. However this negativism about religion did not prevent Pinel from insisting on the importance of morality in treatment, through adherence to basic ethical precepts of kindness, respect for self and others, temperance, and deference to authority, as conditions for, and clinical objectives of, moral treatment. He also clearly believed that passion was tied to vice and sensual indulgence were clearly risk factors for mental alienation, and that in mental health the passions needed to be aligned according to sound principles of morality that stressed balance and temperance. In addition to these psychological and ethical ‘moral’ goals, moral treatment was also a holistic intervention that required stabilization of all basic physical parameters of bodily functioning. Thus hygiene in the most general sense was an essential part of ‘moral’ treatment. Special attention was paid to assuring physical comfort as well as proper sleep and nutrition, all under the guise and in the name of ‘medical’ treatment for their condition. For suitable patients, diversions and distractions were also provided, through engagement in various occupations. Play and special festivities were also held to be important in restoring and rehabilitating the wayward minds and passions of patients.

While Pinel’s position may be truer to the phenomena of our psychological life by virtue of the fact that it sometimes includes morals, it also has its problems. One of these is the possibility of falling into excessive moralism. Consider, for example, the following passage with its highly morally condemnatory tone towards vice in the lower classes:

C’est un contraste perpétuel de vices et de vertus qu’offre l’espèce humaine dans l’intérieur de la vie domestique, et si d’un coté on voit des familles prospérer une longue suite d’années, au sein de l’ordre et la concorde, combine d’autres, surtout dans les classes inferieures de la société, affligeant les regards par le tableau repoussant de la débauche, des dissensions et d’une détresse honteuse! C’est la suivant mes notes de chaque jour, la source la plus féconde de l’aliénation qu’on a a traiter dans les hospices … Je m’en m’abstiens de reproduire au grand jour des exemples de cette sorte, dons quelques un honorent l’espèce humaine, mais dont un grand nombre d’autres forment un tableau le plus dégoutant, est semble être pour elle une opprobre (Pinel 1809, 29). [Within domestic situations humanity presents a perpetual contrast between vices and virtues. On one side one sees families which thrive over a course of many years, in the bosom of order and concord, on the other one sees many others, especially in the lower social classes, who offend the eye with the repulsive picture of debauchery, arguments, and shameful distress! Therein lies, according to my daily records, the most prolific source of mental alienation one has to treat in the hospitals … I refrain from publicising to the wide world examples of this kind, of which some are a credit to the human race, but many others make a disgusting picture, and seem a disgrace to humanity.]
Evidently, there is perhaps an excessively moralistic side to Pinel's attitude towards the insane, a fact that tends to get ignored. However, that does not mean that we should automatically discount his insights about the importance of morals in affective psychopathology. It does though underline the fact that both sides of the dilemma we are faced with seem to have genuine problems.

9. The dilemma

In closing, let us review the dilemma we are faced with. It confronts anyone who wishes to study affective psychopathology and at the same time remain within the confines of empirical science, traditionally understood. The dilemma is whether to include or exclude questions that involve direct reference to morals or other normative considerations of value. On the one hand, excluding such questions threatens one's truthfulness to the phenomena under study, since morals and other similar normative questions form an important dimension of how affectivity figures in mind and behavior. On the other hand, including such considerations threatens the integrity of the effort to understand those phenomena scientifically, since such normative ethical and other related evaluative matters seem to fall outside the methodological reach of science, strictly speaking.

Our two protagonists take bold stances in addressing the dilemma. Crichton drew a moral line in the sand and insisted that we keep morals and values and other related psychological matters out of the psychopathology of the affectivity. Pinel was clearly aware of this methodological manifesto of Crichton's, but never openly confronts it in his remarks on the Scottish doctor's work. Nonetheless, on close inspection, Pinel apparently does defy Crichton's edict. In many instances, Pinel crosses over from psychopathology to morals, and values, while attempting to make the case for his own conception of mental illness and his moral treatment.

No one can really be declared the winner in this historical debate. The lesson to be derived from this stalemate is that the dilemma may be more robust and harder to solve than contemporary affective scientists recognize. We should not simply assume that we can distinguish fact and value in affectivity and then proceed to scientifically study the facts alone, as if values can be excluded.
References


How to evaluate the factual basis of emotional appraisals?

Mikko Salmela

1. The problem

William James (1884) memorably wrote that an emotion is the feeling of bodily changes that follow directly the PERCEPTION of the exciting fact (p. 189). I see a bear coming toward me in the forest, start to tremble, and feel afraid, or I hear the news about the death of my aunt, begin to weep, and feel sad. James’s view of emotion has been severely criticized, because it appears that mere perception of facts cannot evoke an emotion without some kind of evaluation of those facts—either automatic, unreflective, and unconscious, or deliberate, reflective, and conscious. Seeing an approaching bear does not evoke fear unless the bear is evaluated as being dangerous; hearing a piece of news does not give rise to sadness unless the meaning of the news is interpreted as implying a significant loss. Indeed, many emotion theorists, including neo-Jamesians like Jesse Prinz (2004), have followed the lead of Richard Lazarus who skips the language of facts and introduces the notion of ‘core relational themes’, such as “a demeaning offence against me or mine” (anger), or “facing an immediate, concrete, and overwhelming physical danger” (fright), to capture those emergent relational meanings whose recognition in the subject-environment relationship gives rise to emotions.

The idea of relational meanings that underlie emotions is an important insight, because it shows that emotions arise neither from environmental conditions, nor from subjective characteristics, but rather from a synthesis of the two. Lazarus (1991) writes, The separate identities of the two subsystems are then lost in favor of an emergent condition, described as one or another relationship with its own relational meaning (p. 90). However, too heavy emphasis on the synthetic nature of relational meanings may cloud James’s perception of the role of facts in emotion elicitation, a role that becomes more obvious when we raise normative questions about the warrant of emotions. For even though relational meanings emerge as
combinations of environmental and organismic attributes, we must analyze these synthetic meanings back to their constituent parts if we want to determine whether the relational meaning—and the associated emotion—is justified in the particular situation. Here the Jamesian idea of “exciting facts” re-emerges in the revised form of facts that warrant the ascription of a particular evaluative meaning, such as danger or offense or loss, to the emotion-eliciting situation.

In some cases, the factual properties of an object provide sufficient warrant for a particular emotion. A maggot-infested piece of meat is disgusting, and an attacking predator is frightening, provided that I am the target. However, this is so because well-being and survival are biologically hard-wired concerns, and those factual properties render the piece of meat disgusting or the predator frightening for virtually every human being. But if I do not share your mental or physical constitution, or your concerns, values, and beliefs, our emotions or rather their appropriate objects differ significantly. A playful dog is harmless to an adult, but it may accidentally hurt a small child; women are often more likely to be harassed in parks at night than men; cocktail parties can be more stressful for temperamentally shy persons than for extroverts; and believers in voodoo may become paralyzed on learning that they have been cursed, even though a similar curse does not constitute a danger for non-believers. In each of these cases, some non-normative facts about the situation appear to warrant the emotion, even though the same emotion would not be warranted for all people. The problem is how to evaluate the factual basis of emotional appraisals, considering that people respond to similar situations in dissimilar ways in light of their individual and cultural differences.

1. There is a persistent philosophical debate about the relationship between facts and values. Even though many philosophers still accept the fact/value distinction in some form, the practical relevance of this principle has been questioned by thinkers who emphasize the entanglement of fact and value in normative practices, such as ethics and aesthetics (see e.g., Putnam, 2002; McDowell, 1998; Wiggins, 1987). They point out that such thick evaluative concepts as ‘cruel’, ‘rude’, or ‘vulgar’ are evaluative, because these concepts ascribe a specific kind of negative attitude to acts characterized accordingly. However, there are also certain non-normative–factual or descriptive–characteristics that acts considered cruel, rude, vulgar, and the like must have in order to instantiate the property in question, since thick concepts cannot be correctly ascribed to just any act or object. Instead, the application of thick concepts is guided by the facts or by the user’s perception of the world, as Bernard Williams (1985, p. 141) points out. For instance, an act is cruel if a person intentionally acts so as to cause severe physical or psychological damage to a sentient being or beings. It is the presence of these factual properties that warrants the thick evaluation of the act as cruel. The same point applies to emotional appraisals of particular objects as disgusting, frightening, offensive, shameful, and so on.
2. The situational warrant of emotion in philosophy and clinical psychology

Both philosophers and clinical psychologists face the problem of how to evaluate the factual basis of emotional appraisals, even if they approach this problem from slightly different perspectives – philosophers from the viewpoint of rationality and psychologists from the perspective of functionality. Thus, philosophers discuss emotional warrant mainly in terms of appropriateness, whereas psychologists prefer the notion of adaptiveness. Differences in terminology manifest divergent standards for the situation-specific warrant of emotions.

Psychologists generally assume that emotions are adaptive, special-purpose mechanisms that help agents to protect or achieve their fundamental needs or goals, such as survival, health, reproduction, well-being, self-esteem, autonomy, and so on. Thus, fear is an adaptive response that enables one to fight or flee dangers; jealousy is an adaptive response to infidelity that enables one to detect and deter sexual rivals. An adaptive emotion can turn maladaptive in two ways: either by changing needs and goals or by changing circumstances. The latter is more usual as psychologists presume that basic human needs are fairly stable and universal.

Philosophers agree that emotions are adaptive, at least in general and most of the time, as the psychologist Nico Frijda (1994) has remarked. Nevertheless, philosophers rarely discuss emotions only in terms of adaptiveness. Rather, they focus on appropriateness. Whereas the notion of adaptiveness focuses on the consequences of experiencing and expressing an emotion, considerations of appropriateness relate to the grounds or reasons for having an emotion. Even if instrumental reasons for an emotion share the consequence-oriented focus of adaptiveness, considerations about appropriateness concern the emotion’s rational fit with the eliciting situation. The question is whether a particular emotion is a warranted response to the situation from an epistemic, moral, prudential, strategic, or overall point of view, as Justin D’Arms and Daniel Jacobson have pointed out.

These brief descriptions of philosophical and psychological approaches to the problem of situational warrant of emotion may suggest that these approaches fall far apart from one another. Indeed, the purportedly value-neutral notions of functionality and adaptiveness exhaust the explicit normative criteria of emotions in clinical psychology, and psychologists rarely discuss problems of emotional
warrant in terms of appropriateness without ultimate concern for adaptiveness. However, the question of appropriateness remains indispensable even if we adopt a functional approach to emotions, because the functionality of emotions requires that they reliably detect those core relational themes in the subject-environment relationship that is their function to detect.

Psychologists tend to assume that there is a correlation between the adaptiveness and appropriateness of an emotion. This correlation is often indeed the case. It is generally both adaptive and appropriate to fight or to flee dangers and to avoid noxious substances. Further still, a well-functioning appraisal process makes it likely that emotions will be appropriate responses to the situations in which they occur, as Ira Roseman and Craig Smith (2001, p. 7) emphasize. However, the difference between adaptiveness and appropriateness can be appreciated by considering the circumstances where maladaptive emotions are most often learned. It is a tragic fact that maladaptive emotions are adaptive in those oppressive, abusive, neglectful, or otherwise deleterious circumstances that are not responsive to the subject’s basic needs. It is adaptive to deny losses and repress offenses that one cannot or is not allowed to mourn or resent in such circumstances.

Nevertheless, this sad fact does not imply that denial and repression were rational responses even in those original situations – quite the contrary! Many clinical psychologists emphasize that it is important to live through denied losses and repressed offenses and to feel sad and angry about these events afterwards, because these situationally appropriate emotions, which would have been appropriate in the original situation, empower the person in the struggle against the maladaptive emotion. The example shows that we can distinguish between short and long term adaptiveness, where the latter, more often than the former, associates with situational appropriateness. Psychotherapists thus implicitly recognize the role of appropriateness as a distinct and complementary standard of emotional warrant along with adaptiveness. But since clinical psychologists do not themselves address the problem of emotional appropriateness themselves, there is a theoretical lacuna in their discussion of the situational warrant of emotions. The question

2. In the psychological discussion on adaptiveness, communal standards of appropriateness function as norms whose internalization or violation bears on the general adaptiveness of emotions. Here social and cultural standards of appropriateness are treated as facts about the relevant community, not as an issue for psychological theorizing. Thus, for instance, James J. Gross and Ross A. Thompson (2007) point out, cultural values are significant in determining what constitutes ‘adaptive response alternatives’ for expressing emotion for persons of any age (p. 15). In a like manner, Batja Mesquita and Dustin Albert (2007) remark that the adaptiveness of emotions depends on their fittingness with cultural models of self and relating.

3. See e.g. Shaver & Mikulincer (2007); Thompson & Mayer (2007); Young, Klosko, and Weishaar, and Greenberg (2003).
then becomes, whether this lacuna can be filled by theoretical insights obtained from philosophical discussions about emotional appropriateness.

3. Philosophical theories of emotional appropriateness

The main problem with the existing philosophical theories of emotional appropriateness is the abstract and impersonal level of discussion. Even though many theorists admit or even emphasize the socially constructed nature of emotions and their intimate connection to personal values, they typically raise and discuss problems of appropriateness either in abstracto, with only roughly sketched examples, or in the context of detailed examples that are tailored to illustrate the theoretical problem at hand.

Justin D’Arms and Daniel Jacobson (2000) provide an example of the former strategy. They suggest that the two proper dimensions of emotional appropriateness as fittingness are the shape and the size of an emotion. My envy, for instance, is unfitting by its shape “if the thing I envy isn’t really possessed by my rival, or if it isn’t really good – indeed is better than mine” (p. 73). On the other hand, my envy is unfitting on grounds of its size, or proportionality, if what I have is almost as good as my rival’s. Thus, if I envy my neighbor’s new car, the neighbor must have purchased a new car, and it must in some relevant sense be better than my car. However, if I or we do not share your values, the objects of our appropriate emotions differ significantly. Handguns may be appropriately envied by the members of the National Rifle Association but hardly by pacifists. Yet if one’s values prove to be unwarranted from a more enlightened perspective, a locally appropriate emotion turns out to be inappropriate after all.

D’Arms and Jacobson’s analysis of emotional appropriateness is unintelligible without reference to some personal or social values but it becomes contentious upon application to particular contexts. This is a typical malady of abstract

---

4. In a more recent article, “Anthropocentric Constraints on Human Value” (2006), D’Arms and Jacobson maintain that envy is fitting if one’s rival possesses positional goods that are relevant to determining one’s social position. They write, Once it is granted that positional goods matter for human flourishing, then it follows that envy is sometimes fitting, because the success of rivals can be bad for an agent in the way envy suggests; it marks a comparative loss (p. 123-124). However, the crucial question is, who is my rival, or rather, who is my worthy rival? If I base my self-esteem on pursuit of excellence in a practice or a role that is either trivial or immoral, it is not obvious that my rival’s possession of positional goods –that qualify as positional goods within the particular practice or role– constitutes a reason for my fitting envy. This indicates that the positional goods whose possession warrants fitting envy must be valuable on some independent grounds, in addition to their being positional goods in social rivalry, and it is this latter, independent value that determines the enviability of the positional good.
theorizing that philosophers sometimes attempt to escape by fabricating detailed examples in support of their theoretical views. However, the snare of this opposite strategy is that meticulous examples threaten to remain as abstract as explicitly general accounts if they are not embedded in a more comprehensive perspective with a particular individual with his or her beliefs, values, personal history, and social context as a reference point. This problem bedevils Patricia Greenspan’s (1988) otherwise insightful analyses of emotional warrant.

In *Emotions and Reasons*, Greenspan argues that contrary emotions about the same situation may be appropriate if both emotions are backed by a substantial subset of the subject’s reasons and total body of evidence. Her example is love of a nonchalant person. This emotion may be warranted by the energizing spur that it gives toward the socially valuable end of bonding, whether or not this end is achievable in the particular case. Yet the subject may appropriately hate the same person for his very nonchalance because that trait hurts her feelings. Therefore, *love and hate may both be appropriate for different reasons, reasons whose significance is assessed in accordance with different conceptions of the general adaptiveness of the emotions based on them*” (p. 131).

Obviously, contrary emotions about the same situation are sometimes warranted. However, the reason that warrants love in Greenspan’s example is quite irrelevant from the subject’s point of view and therefore cannot figure as one of the subject’s reasons for loving the nonchalant person. For even if the subject would recognize and appreciate the social value of bonding, this reason may not warrant her love, since the energizing spur is so evidently misdirected in the situation. Even worse, this impersonal reason is not a plausible reason for love in the first place because it is not a reason for loving this particular person, which is essential to reasons for appropriate romantic love, in contrast, for instance, to sentimental love, where we fall in love with the idea of being in love and use the beloved person as a means to savor and entertain this blissful emotion. Indeed, if an energizing spur toward bonding provided sufficient warrant for love, one could justifiably love just anybody, which is absurd.

Philosophical theorizing about emotional appropriateness thus faces a dual challenge. On the one hand, there is a need for an account of *individual* emotional appropriateness that is more deeply immersed in the cognitive, evaluative, and motivational perspective of the particular person. Such is the problem of *emotional authenticity* that focuses on the criteria of internally justified emotions (Salmela, 2005). How should I feel about this situation, provided that I am this person with this mental and physical constitution, with this personal history, with this cognitive, perceptual, and sensory evidence? This is a question that Greenspan’s account of emotional warrant fails to respond to in a satisfactory manner. On the other hand, a subject’s personally or communally justified emotions may
turn out to be unwarranted from a more global and enlightened point of view. This constitutes the problem of *emotional truth*: how should I feel about this situation, provided that my emotions were based on the most careful, informed, and imaginative reflection of the matter and situation? (Salmela, 2006) The notion of emotional truth responds then to the problems raised by D’Arms and Jacobson’s account of appropriateness.

### 4. Emotional authenticity

The distinction between emotional authenticity and truth opens two perspectives on the situational warrant of emotions. In fact, the distinction between authenticity and truth can be framed in terms of the amount and quality of evidence that the subject possesses for evaluating the basis of an emotion. An emotion is authentic if it is internally justified by the total body of evidence that is either actually or conceivably available to me even though the emotion may not be justified in a more global or external sense that associates with the notion of emotional truth. This notion must be understood in an anti-realist sense, since evidentially unconstrained truth is beyond the reach of emotions anyway. Thus, the notion of external warrant refers to an ideally enlightened communal warrant that is external to both individual persons and actual communities but not to human sensitivities as such.

I have pointed out that non-normative facts about the emotion-eliciting situation either warrant or repudiate a particular emotional appraisal of the situation. This dependence of emotional appraisals on non-normative facts about the situation renders those appraisals vulnerable to rational criticism. We can ask whether the object possesses a factual property or properties that warrant the ascription of a certain evaluative property with respect to an individual person’s sensibility or with respect to a particular community of sensibility, or whether a particular factual property is capable of warranting a distinct emotional appraisal in the first place. The factual warrant of emotional appraisals *qua* authenticity and truth is defeasible in light of enlarged or improved information. For example, our information about what can significantly harm us may change from time to time, consequently affecting our warranted ascriptions of the property of being dangerous. These changes can occur in two ways: either we learn more about the world or we learn more about ourselves. Thus, for instance, I may learn that exposure to sunlight has

5. Ronald de Sousa (2002; 2004) has also proposed an account of emotional truth. However, his account falls prey to problems, similar to D’Arms and Jacobson’s view of emotional appropriateness. See my “True Emotions” (*The Philosophical Quarterly*, 56, 382-405) for a more detailed critique of de Sousa’s proposal.
become harmful to me and other people because of the increased amount of UV-radiation, or I learn from my dermatologist that I have a rare, chronic skin disease that makes exposure to sunlight even more harmful to me than to other people.

I propose that an emotion is authentic if it is situationally appropriate from the individual subject’s point of view. This means that the emotion is a rational response to the situation, given the total body of evidence that is actually or conceivably available to the subject. For instance, fear is authentic if the subject has reasons to fear that the particular object of fear is capable of inflicting significant damage to his or her valuable goal. It is important to realize that the subject need not believe this before or while experiencing the emotion; in fact, emotions are often necessary for acquiring this kind of evaluative information about the subject-environment relationship, as Greenspan (1988), Oakley (1992), and Helm (2001), among others, have pointed out. Emotions both inform us about significant events and features in the environment and motivate us to act in an adaptive manner in a given situation. But since emotions are not justified by themselves but only by reasons, we must relate those reasons to the facts that are cognitively, perceptually, or sensorially available or conceivably available to the subject.

I have proposed an integrity account of emotional authenticity, which states that an emotion is authentic if its evaluative content coheres with the subject’s internally justified mental states (Salmela, 2005). The criteria of internal justification are rationality, autonomy, and processuality. As rational beings, we aim to perceive what actually exists, to believe what is true, and to value things that are worth valuing. However, temporal, historical, cultural, personal, or other unavoidable limitations may distort our perception, understanding, and reflection about the real, the true, and the good, and it does not make sense to suggest that such limitations would pre-empt the possibility of emotional authenticity unless the present limitations are consequences of our past acts or omissions that have shaped our present perceptual, cognitive, and evaluative capacities (Sherman, 1999).

---

6. I remain neutral about the metaphysical status of the real, the true, and the good. These notions are put forward as the formal objects of perceptions, beliefs, and valuations, respectively. Formal objects are objects that tokens of a specific mental state type have in common. Thus, perceptions purport to represent what is real; beliefs, what is true; and valuations, what is good. The fact that some perceptions, beliefs, and valuations are hallucinatory, false, and bad, respectively, does not pre-empt the role of formal objects as categories that render particular perceptions, beliefs, and valuations intelligible as mental states of a certain kind. Moreover, formal objects provide standards of fittingness for mental state types, as de Sousa (2002) points out. Thus, a fitting perception presents what is real, the content of a fitting belief is true, and the object of a fitting valuation is good. This claim is independent from the way in which standards of reality, truth, and goodness are fleshed out.
Yet mere coherence between an emotion and one’s total body of evidence may not be sufficient for authenticity. Let us consider a person who loathes himself for seemingly rational reasons: he has not succeeded in life, he has no proper education, no job, no family, no friends, no wealth, and so on. This is a person who was probably abused, abandoned, oppressed, neglected, rejected, or otherwise mistreated in childhood or adolescence. Such person may find that his present emotion coheres with the available evidence that the subject regards as rational. Yet the main problem is the emotional disposition of self-loathing that has distorted the person’s evaluations of himself, the world, and the future. This maladaptive emotion schema that constitutes a coherent pattern of thematically interrelated memories, emotions, bodily sensations, perceptions, and cognitions continues to plague the person in his present life.

In order to exclude such cases from the scope of authenticity, I propose that coherence between an emotion and its warranting evidence should survive critical acknowledgment of the manner in which the evidence and/or the emotional disposition was adopted. This is the condition of autonomy that highlights critical transparency about one’s own mental constitution (Christman 1991). Observe, however, that this condition is stated in a counterfactual form. Therefore, it does not force us to engage in laborious self-scrutiny, provided we remain open to new evidence that is capable of challenging the warrant of our emotions. However, emotional disorders are typically associated with relatively high immunity to contrary evidence. Accordingly, many psychotherapies emphasize cognitive and emotional return to the paradigm scenarios of maladaptive emotions. This therapeutic recollection helps the person to grasp the violation of his or her vital goals, which, in turn, helps the subject to dissociate from the disordered meanings and appraisals that emerged along with the maladaptive emotion schema.

Finally, processuality is an inescapable aspect of authenticity, because spontaneous emotions constantly provide us with new information that has the capacity to challenge our present understanding of the world and ourselves, including the emotions we conceive of as authentic by virtue of their coherence with our other mental states. We strive to build a congruous whole from the diverse aspects of ourselves, but we do not want to commit ourselves to any particular self, however coherent that self may be. Instead, we want to remain open to new experiences that can trigger change, learning, and growth. In fact, psychological studies of “true feelings” indicate a strong connection between authentic emotions and the resolution of a personal crisis or internal turmoil. This connection indicates that experiences of authentic emotions often involve a transition toward a more coherent and integrated self (Morgan & Averill, 1992; Averill & Nunley 1992). Thus, authentic emotions are both constitutive and expressive of a restored or renewed integrity in one’s self.
5. Emotional truth

By contrast to authenticity, I have suggested that a true emotion is warranted by
reasons that remain undefeated, no matter how much our information about the
relevant facts is or will be enlarged or improved (Salmela, 2006). This account of
emotional truth is based on the observation that emotions meet the syntactic and
disciplinary requirements of minimal truth-aptness, proposed by Crispin Wright
(1992) for moral and aesthetic discourses. First, emotions have an assertoric con-
tent. In the same way as a belief that \( p \) asserts that \( p \) is true, a moral judgment or
an emotion presents its content as warranted. To quote Peter Goldie (2004): *in the
typical case, the emotional response, combined in phenomenology with the percep-
tion of the object as having the emotion-proper property, will involve the experience
of the emotion as being reasonable or justified* (p. 97). Second, emotions have pub-
licly acknowledged standards of warrant. Even when particular objects of emotion
differ significantly between individual persons, the reasons for the emotions must
be intelligible and plausible for other members of one's community. For instance,
the death of a relative, the loss of job, or divorce count as possible and acceptable
reasons for sadness. By contrast, sadness over a lost container of ashes is unintel-
ligible unless the particular heap of ashes has some deeper meaning, such as being
the ashes of a cremated relative. These brief examples suggest that standards of
emotional warrant are at least communal, if not pancultural.

I propose that emotional truth can be modeled on Wright's notion of super-
assertability, which is *the property of being justified by some (in principle accessible)*
state of information and then *remaining justified no matter how that state of infor-
mation might be enlarged upon or improved* (p. 47). Wright applies superassertabil-
ity to comic discourse about what is funny. However, I believe that the notion is
applicable to the emotion of amusement, whose relational meaning is the property
of being funny, and to other emotions as well. For instance, true fears can, in prop-
er conditions, include the fear of radiation or of terrorist attack, but also the fear of
extended unemployment or mental illness, because no further or enlarged inform-
ation alone, without social reform, is capable of eliminating the fact that these
things are actually harmful in contemporary Western societies. The situation is
quite unlike the fear of a voodoo curse, where knowledge about the causal realiza-
tion of the curse through self-suggestion is capable of removing the danger of be-
ing harmed by the curse, even though the removal may not happen overnight.

The main problem with emotional truth concerns the incommensurability of
communal standards of warrant. Fear is a good candidate for a truth-apt emotion
because we may agree that our ascriptions of harmfulness are defeasible in light of
enlarged or improved information. Other emotions, such as amusement, shame,
guilt, or envy, appear to be significantly different, since their appropriateness does not
depend so heavily on factual grounds. Nor need culturally specific reasons for those emotions involve downright false information or distorted reasoning about emotional objects. For instance, Asians may know the same facts about human psychology as Westerners, but they still feel more guilty of their violations against communal norms of behavior because the value of social harmony is more salient to them.

Wright (1996) argues that superassertability is a language-game internal notion, as it were. [It] is a projection of whatever internal discipline informs the discourse (p. 10). This view suggests that we are not allowed to use external criteria to evaluate the warrant of culturally-specific disgust or shame, for instance. The result appears to be some sort of relativism among different communities of sensibility. Each community has its own standards of emotional truth, but the standards of different communities, even within a single society, may differ quite radically, as the example of pacifists vis-à-vis members of the NRA suggests.

Yet superassertability also involves the idea of warrant that survives arbitrarily close scrutiny of its pedigree and arbitrarily extensive increments to or other forms of improvement of our information (Wright, 1992, p. 48). This aspect of superassertability is in conflict with the language-game internalism, narrowly interpreted, because an arbitrarily close scrutiny of pedigree and an arbitrarily extensive enlargement or improvement of information can involve considerations that are capable of challenging or refuting even widely shared communal standards of warrant if these standards are founded on false, insufficient, biased, or otherwise distorted information, or on faulty, inconsistent, prejudicial, ideological, or similar reasoning in the processing of information (see Blackburn, 1998). These criteria may decide the case between pacifists and the NRA as well as other, similar cases where the facts bear on the warrant of the evaluation. Only if communities of sensibility differ on the level of incommensurable fundamental values, such as individual autonomy and social harmony, does it appear to be impossible to adjudicate between them, because both standards remain superassertable in the relevant cultural community.

6. Emotional authenticity and truth in psychotherapeutic practice

In this final section, I will argue that emotional authenticity and truth are ecologically viable and implicitly recognized standards for evaluating the situational warrant of emotions in various psychotherapies, especially in the cognitivist and constructivist traditions. Psychotherapists sometimes refer to authenticity in passing, but the concept does not have a significant theoretical role in their accounts of
mental health, which is conceived in terms of adaptiveness rather than authenticity. Accordingly, therapists do not evaluate their clients’ emotions in epistemic terms as appropriate, true, valid, or rational, but rather in terms of their instrumental usefulness. Thus, emotions are characterized as ‘maladaptive’ or ‘destructive’ or ‘unhealthy’ rather than ‘appropriate’, ‘accurate’, or ‘rational’, whereas the overarching value of all psychotherapies is health or subjective well-being in a comprehensive physical, psychological, and social sense. However, when we take a closer look at theories of clinical psychology, the picture becomes more multifaceted.

Granted that considerations of appropriateness are important in clinical psychology, we can still ask whether emotional authenticity and truth are viable standards of appropriateness in this context. For authenticity, the reply is more clearly affirmative: integrity in the sense of temporal and structural coherence in one’s narrative identity as well as among one’s various mental states is an important hallmark of mental health. The very idea of ‘disorder’ implies lack of order—disorganization, disintegration, and incoherence—in one’s mental life, and its opposite, ‘order’, is associated with organization, integration, and coherence. This association is so evident that therapists rarely pause to explicate the relevance of coherence and integrity as regulative ideals of mental health. Exceptions are those constructivist therapists who emphasize the role of narrative in the construction of the self and reality.

Constructivists analyze psychopathologies as problems in the continuity and coherence of inherently diverse, flexible, and dynamic narratives of the self and the world. They maintain that human beings have an ontological need to make sense of their experiences by giving meaning to them and that this need takes the form of narrative coherence and connectedness in the construction of personal identity and reality. Even though flexibility and openness to alternative meanings are essential to a healthy identity, diversity and coherence do not exclude each other. Instead, coherence is a viable metaphor in the narrative construction of the self. Accordingly, mental disorders are problems in a person’s sense of self-cohesion and narrative identity, and recovery is associated with the restoration of a coherence

7. There are a few exceptions, however. In their schema therapy, Young, Klosko, and Weishaar (2003) propose that humans are happiest and most fulfilled when they are expressing authentic emotions and acting on their natural inclinations (p. 254). With ‘natural inclinations’, Young et al. refer to people’s core emotional needs that they believe to be universal. These needs include secure attachment to others; autonomy, competence, and a sense of identity; freedom to express valid needs and emotions; spontaneity and play; and realistic limits and self-control. Young et al. maintain that many mental problems result from the deprivation of these needs, originally by significant others and later by the person him- or herself through a maladaptive schema that influences the person’s behavior. In contrast, authenticity is associated with adaptive expression and satisfaction of core emotional needs.
that is more flexible and resilient than the previous one (e.g., Arciero & Guidano, 2000; Gonçalves, Korman, and Angus, 2000; Neimeyer, 2000; Mahoney, 1995).

Emotions have a dual role in the creation and maintenance of narrative coherence. On the one hand, emotional structures that descend from previous experiences and early patterns of attachment guide the assimilation of new experiences and events into a meaningful whole that provides a sense of self-permanence and continuity in one's life. Yet on the other hand, disruptive emotions and unexpected events constitute challenges to this self-organizing activity. Pathology results if the discordant situations and emotions cannot be integrated into a cohesive identity, whereas by assimilating these apparently disruptive emotional experiences into the self-narrative, a client can achieve an expanded and more coherent sense of self, as Arciero and Guidano (2000, p. 102) point out. In a like manner, schema therapists propose that a dysfunctional schema is a part of the self that is cut off to some degree from other aspects of the self (Young et al., 2003, p. 40). Greenberg (2003) provides the most philosophical rendering of the role of integrity and coherence in his emotion-focused therapy. He suggests that the adaptiveness of an emotion depends on its coherence with the person's needs and other conscious goals, plans, values, and realistic assessments of the situation (p. 175). This 'integration of head and heart', emotionally-based goals and conscious values and reasons, provides the basis for a reflexive second-order decision on the adaptiveness of emotion. Terminological differences apart, Greenberg's view is congenial to my philosophical account of emotional authenticity.

When we move on from the ideal of authenticity as integrity to its constitutive conditions of rationality, autonomy, and processuality, we can see that processuality is one of the key concepts of narrative therapies. The capacity to assimilate new experiences into one's cohesive identity is a vital hallmark of mental health, according to constructivists. Rationality adds an important qualification to coherence because disordered emotions often cohere with the client's existing beliefs and evaluative judgments. However, this coherence is not internally justified because the relevant beliefs and evaluations are elements of a maladaptive schema that distorts one's beliefs and evaluations about the self, the world, and the future. Cognitive distortions lead the person to misperceive and misinterpret the situation in such a manner that reinforces the maladaptive emotion schema; information that confirms the schema is accentuated, whereas information that contradicts with the schema is minimized or denied (Young et al., 2003). Accordingly, a person's emotional appraisals fail to qualify as authentic because they do not cohere with the person's conceivably available, well-founded beliefs and evaluations.

Considerations of rationality are especially emphasized in cognitive therapies. These therapies maintain that mental problems ensue from systematic departures from reality and logic, including arbitrary inferences, selective abstractions, and
overgeneralizations, as Beck (1979, p. 90) points out. Distorted ideas emerge as automatic thoughts or images that precede and enter into disordered emotions as unrealistic, extravagant, or downright false interpretations of events and their significance to the subject. These egocentric interpretations are difficult to change through reasoning or with contradictory evidence alone. Therefore, cognitive therapists often support rational discussion with other methods and techniques, such as behavioral desensitization, experiential imagery, or pharmacological treatment. Nevertheless, the core of the disorder is still cognitive: incorrect or unrealistic interpretation or evaluation, whose removal is hypothesized to result in an emotional change (see e.g., Beck, 1979; Beck, 1995; Dobson, 2001; Leahy, 2004). Other therapeutic approaches admit that irrational beliefs and evaluations are concomitant with emotional disorders and that their remission brings on cognitive changes, even though these approaches do not regard cognitions and their rationality as the primary target in therapy (e.g., Greenberg, 2003; Mineka & Thomas, 1999).

Finally, autonomy is an important idea for all psychotherapies. The importance of autonomy in the therapeutic context stems from the fact that many disordered and maladaptive emotions are learned in a manner that violated against their personal autonomy. People are abused, abandoned, oppressed, neglected, rejected, or otherwise mistreated in their childhood, adolescence, or later in life in ways that they either disapprove and resist at the time or would disapprove and resist should they be able to understand the full meaning and implications of the treatment. If one’s maladaptive emotion schema and its constitutive beliefs, perceptions, evaluations, and emotions originate from a paradigm scenario that involves this kind of violation of personal autonomy, the resulting schema is not authentic, even though the distinct aspects of the schema form a coherent pattern that is relatively immune to contrary evidence.

The notion of emotional truth may appear less important for clinical psychology. It may not be adaptive to care whether one’s anger, guilt, shame, or other social emotions are warranted by some ideally enlightened point of view as long as they fit the actual standards of one’s community that its other members expect the person to follow. It is true that conformism is often an adaptive strategy. Nevertheless, when clinical psychologists and their clients talk, as they often do, about real dangers or offenses or losses and so on as the objects of adaptive emotions, they have more or less explicit criteria of emotional appropriateness in mind. These criteria presumably coincide with communal feeling rules that both therapist and client share. Yet the justification of those feeling rules does not depend merely on their acceptance by a consensus omnium but on their being grounded in plausible reasons within the particular community of sensibility. For even communal feeling rules are warranted only by reasons that are regarded as undefeated in the particular community at the current level of information and reflection. In many
communities, people have believed or still believe that they have found the absolute and infallible standards of emotional truth, especially in the domain of moral emotions, which are based on the order of things, sacred religious commandments, or other cultural inscriptions. The fact that these standards may lose or have lost their warrant from a more enlightened point of view does not remove the importance of emotional truth as a viable regulative ideal of emotional appropriateness that retains its status even while we recognize the inherent fallibility of human cognitive and reflective capacities.

7. Conclusions

Psychologists emphasize the adaptiveness of emotions, rather than their appropriateness because psychologists typically focus on the behavioral components of emotion which causally adjust the subject-environment relationship to meet the subject's goals in the emotion-eliciting situation. However, the adaptiveness of a particular emotion depends essentially on its appropriateness to the eliciting situation. Therefore, philosophical considerations of emotional appropriateness are relevant for psychotherapeutic practices as well even though such considerations are not explicitly recognized in psychological theorizing. I have proposed that emotional authenticity and truth are important and viable standards for evaluating the situational appropriateness of emotional appraisals. Although these standards have been developed within a philosophical framework, they are plausible and ecologically valid also in clinical psychology.

Emotional authenticity and truth are complementary ways of evaluating the situational warrant of emotional appraisals, internal and external. In both cases, emotional appraisals are warranted by situational facts that justify the emotional appraisal either within an individual person's informed sensibility, or within an ideally enlightened community of sensibility. An authentic emotion is a rational response to a particular situation, given the total body of evidence that is actually or conceivably available to the subject, whereas a true emotion is warranted by reasons that remain undefeated no matter how much our information about the relevant facts is or will be enlarged or improved.

The standards of emotional authenticity and truth resolve the problem of abstractness and impersonality that plagues existing philosophical accounts of emotional appropriateness. Emotional authenticity embeds the question of emotional appropriateness within a comprehensive perspective that takes the particular individual's beliefs, concerns, values, personal history, and social context as its point of reference. However, these mental states and facts are not taken at face value, because they may warrant an emotion only if their coherence with the particular
emotional appraisal passes the criteria of rationality, autonomy, and processuality. Emotional truth, in turn, is a standard of idealized communal warrant that is capable of adjudicating conflicts within and between actual communities of sensibility insofar as these conflicts depend on disagreements about facts or their bearing on valuations. Thus, emotional truth provides a focus imaginarius for a conscientious rational scrutiny of emotional appropriateness, even though a wide cross-communal agreement on true emotions may not be forthcoming in practice.

References


The problem with too much anger

A philosophical approach to understanding anger in borderline personality disordered patients

Nancy Nyquist Potter

1. Vignette

A concerned mother called me to raise the issue of whether some further, perhaps different, treatment might be needed for her daughter. I responded by saying that I would talk with the patient about this concern and would get back to her. After talking with her daughter, I suggested that she call her mother while she was still there and could participate on another line. The patient promptly began a lengthy diatribe at her mother for interfering and poor judgment...After trying unsuccessfully to interject some comments, I put my finger across my lips and went, “Shh,” to my patient. She ragefully slammed down the telephone and left. (Gunderson 1984, p. 110–111)

How are we to view this patient’s anger? Over-the-top? Inappropriate? Did the psychiatrist provoke her anger? Was the anger warranted?

This patient is diagnosed with Borderline Personality Disorder (BPD). The borderline personality is characterized by identity disturbance, feelings of chronic emptiness, impulsive or self-destructive behavior, and unstable intense interpersonal relationships. Loss of a sense of self separate from others, contradictory self-images that are experienced as an inner void, and aggression are key characteristics, according to Goldstein (1995). Distrust, all-or-nothing thinking, extreme sensitivity to perceptions of unfair treatment, and an appearance of normality that quickly unravels under stress are additional features.

Patients with BPD are difficult to interact with in part because of their anger (Gunderson, 2001; Linehan, 1993). Inappropriate or intense anger is one of the criteria for BPD: “…frequent displays of temper, constant anger, recurrent physical fights” (DSM-IV-TR 2000, 710). If we are to correctly identify when this criterion
is present, we need to know the difference between pathological and reasonable anger and what norms govern the distinction. I will argue four things: (1) that anger is a form of communication that requires acknowledgement from the listener; (2) that BPD anger is dispositional (and often dispositionally dysfunctional), but that a patient also can have occurrent anger that is reasonable; (3) that failing to acknowledge BPD anger risks worsening their problems; and (4) that assessments of BPD anger draw upon contextual norms for expression of emotion that may impede giving uptake.

2. What is anger?

Anger is a *moral emotion*, which is to say that moral judgments are normatively paired with particular emotional responses (D’Arms & Jacobson, 1994, p. 748). Marcia Linehan describes emotions as “full systems responses. That is, they are integrated patterns of experiential, cognitive, and expressive, as well as physiological, responses” (Linehan, 1993, p. 68). Grammatically, a person is “angry at” someone or “about” something. As Larry May argues, anger is stimulated by a sense of injury; the object of anger repels (May, 1998, pp. 18–19). May suggests that anger is a kind of reactive assessment, and it may be rudimentary, as when one thinks that it has taken another too long to respond to a question, but still involves some kind of conscious assessment (May, 1998, p. 19).

To elaborate on the concept of anger, I draw upon work by Frye and Austin on language, as well as previous research of my own (Potter, 2000.) “Being angry at someone,” Frye writes, “is somewhat like a speech act in that it has a certain conventional force whereby it sets people up in a certain sort of orientation to each other; and like a speech act, it cannot ‘come off’ if it does not get uptake” (1983). Austin (1975) introduced the concept of uptake to characterize the engagement of the listener in order to secure the meaning of a speech act: when the listener accepts another’s speech act and gives it the conventional understanding, the listener has given the speaker uptake. For example, my promise to you can only be said to be successful when you understand my speech act as one in which I place myself under obligation to you (Austin 1975, p. 571). Your recognition of my speech act as a promise is the uptake.

Uptake, then, occurs when the listening party reorients herself to the communicator and the relation between the two of them “comes off” with an appropriate response. As I discuss elsewhere (Potter, 2000; 2002), oftentimes expressions of anger are *acts of claiming* that call for a response to a person’s claim that she has been wronged; giving uptake to anger requires that the audience acknowledges (1) that a claim is being made, and (2) that that claim is asserting the speaker’s worth.
To get angry is to claim implicitly that one is a certain sort of being, a being which can...stand in a certain relation and position a propos the being one is angry at. One claims that one is in certain ways and dimensions respectable. One makes claims upon respect” (Frye, 1983).

The point is that anger is relational: it has a logic that involves perceived wrongdoer and wronged. As such, a respectful way to respond to anger would be to give it uptake even if it turns out that the angry person's perception of a wrong was incorrect. Anger is not always “about” injury, of course; it could be caused by physiological responses to frustration or stress. I am talking about a kind of anger.

May's and Frye's work is useful in thinking about BPD anger in three ways. First, May suggests that it should be approached as a reactive assessment, which idea emphasizes the cognitive aspect of getting angry. Although anger can be triggered by unconscious associations, it also involves perception, belief, and interpretation and cannot be reduced to the unconscious.

Second, their work suggests that BPD patients may be trying to make a claim about a moral right they think has been violated. Patients are often characterized as having an exaggerated sense of entitlement. When viewed this way, their attempts to assert moral rights are viewed as irrational. An ethical response to expressions of anger includes an attempt to understand what sort of moral claim the patient is making and not to assume that the patient is exaggerating her entitlements. In fact, giving uptake properly is a virtue of its own and is bound up with trustworthiness (cf. Potter, 2002); hence, appropriate responses to patient anger are part of clinician trustworthiness.

Third, at the heart of expressed anger is an attempt to assert one's respect. An alternative way of thinking about borderline anger is that the patient is asserting her worth as a self-respecting human being in the face of perceived injury or insult. As Jeffrie Murphy says, “....If I count morally as much as anyone else (as surely I do), a failure to resent moral injuries done to me is a failure to care about the moral value incarnate in my own moral personality” (Murphy, 1990, p. 505). When a patient's anger is not given uptake, the patient is treated as if she has not communicated; the listener, by refusing even to consider the patient's attempt to orient their relationship as injurer and injured, demeans the patient's assertion of self-worth. In sum, I am suggesting that the patient's anger should be given uptake, because to do so is to recognize her moral worth as a person who is attempting to uphold her own value in the face of perceived hurt.

Of course, patients diagnosed with BPD and the rest of us can get it wrong about anger in a variety of ways, and it is to this point that I now turn.

1. It will also include an analysis of what constitutes a reasonable and unreasonable sense of entitlement, but this task is beyond the scope of this paper.
3. Anger, extremes, and dispositional anger

From an Aristotelian framework, our aim should be to develop a disposition such that we neither overdo nor underdo our responses (Aristotle, 1999). Anger can be inappropriate not only if it lasts too long, or is too vehement for the situation, but also if the target of one’s anger is not the person who did the wrong-doing. On the other hand, Aristotle (like Murphy) believes that not having enough anger is also blameworthy. “For such a person seems to be insensible and to feel no pain, and since he is not angered, he does not seem to be the sort to defend himself. Such willingness to accept insults to oneself and to overlook insults to one’s family and friends is slavish” (NE1126a5–10). Aristotle names the mean the virtue of “mildness,” which is “being undisturbed, not led by feeling, but irritated wherever reason prescribes, and for the length of time it prescribes” (NE1125b35).

The judgment of an “excess” is located in terms of appropriate responses relative to the context, where the mean for anger is the intermediate condition between excess and deficiency in a particular situation. One reason why it’s so difficult to get clear on what counts as excessive anger in BPD patients is that the DSM models are decontextualized. But judgments of extreme or inappropriate anger cannot be made from an abstract point of view. Diagnostic manuals may treat criteria such as inappropriate anger as simple facts, when they actually require evaluative judgments. Furthermore, extreme or inappropriate anger is not necessarily pathological; it may be just a mistake.

On a standard analysis, anger can be either occurrent or dispositional. An occasional expression of excessive, deficient, or inappropriate anger is not so much a cause of concern; the important question is what we have a tendency to do when faced with perceived insults and injuries. If we have a tendency to vent our anger at safe targets rather than the object of our anger, or to explain away all insults as benign and unimportant, we have a disposition toward anger that is flawed.

Excessive or inappropriate anger does not seem to be the only problematic emotion for BPD patients. As Linehan (1993) says, “they often have difficulty regulating the entire pattern of responses associated with particular emotional states” (p. 68). All their negative emotions appear to be dispositional, too. But because of the tendency to associate current injuries with past wounds, BPD anger appears to be only dispositionally maladaptive where, in fact, three issues exist: dispositional anger that is hard-wired, a disposition to bring past anger into the present, and occurrent anger in the present. Here the standard picture of dispositional or occurrent anger is too simple – both because it is most often a doubling of the present with the past, and because norms underlying assessments of anger are themselves complex.

I have indicated that expressions of anger require uptake – but a number of things prevent clinicians from doing so. The next sections review some of those
obstacles and indicate what is involved in giving uptake to someone whose disposition regarding anger is maladaptive.

4. The doubling effect of injury on anger

One difficulty in giving uptake to BPD anger is that it often isn't clear what the anger is “about.” Many clinicians posit BPD anger as a regression to primitive defenses such as “denial, acting out, and projection which prevent the patient’s recognition of the feeling response and its reason” (Gunderson, 2001, p. 53; 1984, p. 35; Kernberg, 1984) and, by definition, primitive defenses are irrational. If a clinician assumes that a patient’s anger is coming from past wounds, the patient may seem unreasonable when she insists she is angry with the clinician for hurt the clinician inflicted. Of course, many clinicians distinguish between the actual relationship between clinician and patient, and the transference relationship. But others are not so insightful and sensitive to this distinction, as the literature suggests, so it is important to bring these philosophical points to light.

The problem with interpreting all BPD anger as regressive is that it denies the possibility that a person can have both warranted, real-time anger and maladaptive anger in response to the past. Linehan encourages clinicians to see the patient’s anger as reasonable, although perhaps mixed with other emotions. She distinguishes “primary” (authentic) emotions from “secondary” (learned) ones. Secondary emotions are “reactions to primary cognitive appraisals and emotional responses; they are the end products of chains of feelings and thoughts.” Maladaptive responses become interwoven with valid ones (Linehan, 1993, 227).

This is what I am calling the doubling effect of injury on anger. Although a patient may have maladaptive anger dispositionally, she may still be able to have occurrent anger that is warranted. Real-time occurrent anger can be an echo of past wounds. Developmental theories of BPD suggest that patients carry deep early childhood wounds that are associated with their dysfunctional and disruptive behaviors. In fact, Linehan argues that “the amelioration of unendurable emotional pain is always suspected as one of the primary motivational factors in borderline dysfunctional behavior” (1993, 265; see also p. 59). The woundedness underlying anger must be given uptake in order for patients to heal. As Linehan points out, it may be difficult to uncover old wounds, partly because the patient herself is confused about her feelings. My concern here, though, is that in viewing real-time anger as merely symbolic of past wounds, clinicians may miss important communicative moments. When clinicians focus exclusively on the irrationality or distorted views of the patient, therapy is undermined (Linehan, 1993, p. 229). And, in fact, inattention to real-time anger may spark a destructive pattern in that
it provokes more hostility and anger when a perceived real-time injury is not taken seriously. The difficulty for these patients is that real-time anger is amplified by past wounds, and the intensity may make it difficult for the patient to regulate it. But amplification from doubling doesn’t mean that it is not important to respond to real-time occurrent anger independently. Precisely because it is a doubled effect and not only an old wound, both injuries need to be addressed. Clinicians can assist patients in distinguishing these types of anger, thus helping them to gain insight into the various sources of their anger.

In the next sections, I focus on that part of BPD anger that is an assessment in real-time.

5. What impedes giving uptake to real-time anger?

The kind of anger we are considering is a cognitively-mediated response to perceived injury or injustice. I have argued that some of those injuries or insults occur in the present moment and that BPD anger cannot all be reduced to pathological responses to past events. So how do we distinguish between anger that hits the mean, extremes of anger, and pathological anger? May suggests that anger that hits the mean would be “in some sense reasonable” (May, 1998, p. 12; emphasis added). That is, from the perspective of the subject, anger is a reasonable response to a perceived violation of a norm that the subject accepts (May, 1998, p. 15; emphasis added).

I will note three things that interfere with giving uptake when considering the idea of “reasonableness.”

One is that I may be angry and believe the anger is reasonable and appropriate but others do not. People differ considerably in what they count as a violation or an injustice. Norms for behavior are not necessarily shared. Furthermore, perspective affects how a moral wrong is viewed or even whether or not it is acknowledged. For example, many people would defend telling a lie who would nevertheless object to being told one (Bok, 1978). Finally, power imbalances affect which norms dominate.

Anger, I’ve suggested, is an assessment that one’s moral rights have been violated. One form that takes is in terms of norms of relationship, where expectations about those norms have been thwarted. Let us contemplate, for a moment, what norms of interaction might be operational for a patient that would make her anger seem reasonable to her if a norm is violated. Possible norms:

- Norm: People ought not to appear bored or restless when I am talking.
- Norm: People ought to listen attentively when I am talking.
- Norm: People ought not to slough me off on others when I am appealing to them for help.
Norm: People should be careful not to automatically discount or suspect me when I explain things to them or express needs.

Norm: People should take me seriously when I say I am frantic and desperate.

Norm: People ought to go out of their way to reassure me that they are not abandoning me if they are planning a holiday.

What is the status of these norms? With the exception of the last one, I suspect many people (in relatively good mental health) subscribe to them. Of course, these norms are applied in different contexts with different results; I too expect to be listened to when I talk to my close friend, but I don’t feel particularly insulted when my students are bored. Still, that such norms are reasonable, from the point of view of the patient, is an important clue to assessment and response of borderline anger. “All norms...are norms of warrant – they’re about what it makes sense to think, to want, or to feel” (D’Arms and Jacobson, 1994, p. 742). I acknowledge that BPD patients can expect too much and that some BPD anger leads to loss of control and so is unreasonable. It is important, though, not to confuse impassioned anger with out-of-control rage. Anger is a passion, and fear of passions may prompt us to judge them as dangerous when they are merely expressions of strong feeling (Fisher, 2002).

My point is that a failure to give uptake to the patient’s sense that a norm or moral right has been violated is to dismiss her norms and expectations as invalid. Such a move cannot be made a priori to giving uptake; we cannot judge whether a real-time injury actually occurred until we’ve explored that possibility. And genuine engagement in exploring possible real-time injuries requires that we be willing to subject our own norms to scrutiny and to take into consideration that the patient’s norms may be, in some sense, reasonable or understandable.

Secondly, norms alone don’t tell us whether or not an emotion is appropriate for the context. Some situations may call for extreme expressions of emotion even when the governing social norms encourage moderation. The anger of BPD patients is often described as an overly sensitive response to slights and inattention, or a reaction to approaching loss or abandonment. If we disregard the experience of double-woundedness – that what the patient is reacting to in the present is an echo of a wound in the past – then perhaps we might view her as overly sensitive. But that way of thinking about patients’ sensitivity risks being dismissive of her situatedness in the context of her current therapeutic relationship.

I follow a distinction drawn by D’Arms and Jacobson: even if a feeling is undesirable (based on norms), this is not grounds for calling it unwarranted (p. 744). Belief is an integral part of emotion. When we feel anger toward someone or about something, it is because we hold a belief about the person or action or situation and we represent it to ourselves in a way that gives rise to anger. Warrant is determined by the content of the belief component of the emotion; “an emotion is
unwarranted when its belief component is unjustified” (D’Arms & Jacobson, 1994, p. 748). Still, one person’s warrant may be another person’s “unwarrant;” warrant must be understood from the patient’s point of view, but the fact that the patient believes she is warranted in a belief does not make it justified by itself. Warrant is a social accomplishment.

The point is that one’s anger or rage may be extreme and therefore outside the norms for anger but that leaves open the question of whether the anger is warranted. The clinician must also establish what beliefs the patient and clinician each are holding and consider to what extent those beliefs are warranted. A problem in giving uptake occurs when the clinician assumes the dysfunctionality of her BPD patient and so has difficulty seeing the patient as a sometimes-reasonable moral agent. Talking about the beliefs that underlie anger is one way to give uptake to the anger without exacerbating it, and may allow the patient to explore other feelings. It may also allow the clinician to identify beliefs and attitudes that are inadvertently contributing to a patient’s echo experience of injury.

But a further difficulty arises when framing appropriate anger in terms of reasonableness – because that norm is based on a theory of rationality that is biased. Reasonableness is assessed similarly to the “reasonable man” standard used in law (cf. Prosser, 1971). But this ideal itself is gendered, as legal scholars argue (cf. Hubin & Haely, 1999; Raigrodski, 1999). The reasonable man is an idealized abstraction and is alleged to be generic but, as so often happens, the generic person turns out to be male.

The problem is not new: since the time of the Pythagoreans, rationality has been linked with maleness and emotionality with femaleness (cf. Whitbeck, 1973). Moreover, the idea that women are not in control of their emotions is not unique to BPD or other mentally disordered women. If appropriate anger is anger that is reasonable, and women by nature are believed to be irrational, then it follows that women’s anger will not be perceived to reasonable. And BPD patients are 75% female. This is the third problem in evaluating anger as reasonable, inappropriate, or extreme.

6. Anger and gender norms

Rom Harré argues that understanding emotions requires that we attend to local systems of rights and obligations, and value in particular contexts (Harre 1986, 6). “Instead of asking the question, ‘What is anger?’”, Harré writes, “we would do well to begin by asking, ‘How is the word “anger,” and other expressions that cluster around it, actually used in this or that cultural milieu and type of episode?’” (1986, 4–5). While I will still ask the standard philosophical question of what anger is, I think Harré is right.
I pursue these ideas by examining norms for behavior as they concern expressions of anger by women.

As May points out, “[a]nger and rage are so commonly expressed by young adult males that we all come to expect it” (May 1998, 8). Not so with women. Frye argues that women’s anger at moral injustices done to them do not get taken seriously and respectfully; instead, women’s anger gets trivialized, pathologized, mocked, and ignored. “Deprived of uptake, the woman’s anger is left as just a burst of expression of individual feeling. As a social act, an act of communication, it just doesn’t happen” (Frye 1983, 89).

Our conventions of language have different norms for giving uptake to women than to men. In particular, the norms for uptake are gendered regarding expressions of anger. Women are not supposed to get angry, or women get provoked by trivial events, and so on. When it comes to women’s claims of injustices and injuries, women are told to stop “playing the victim,” to stop blaming others for their plight, and to take responsibility for their lives (cf. Lamb 1996). In fact, according to Linehan, “The problem for many borderline patients is not the overexperience and expression of anger, but rather the underexpression of it; that is, they are anger-phobic” (Linehan 1993, 356). What they fear is rejection due to their expression of strong emotion. In fact, fear of strong emotions, especially anger, is a more general problem for women. It is a recognized social phenomenon that girls and women are not taught that it is sometimes appropriate to be angry or what appropriate expressions of it look like. Some are unable even to identify what anger feels like.

As I have suggested, cultural analyses reveal that we accord praise and blame for expressions of anger differentially depending on gender. Linehan reflects that whether a BPD patient’s behavior is interpreted as angry depends on who is doing the interpreting (1993, 70). “The overinterpretation of anger and hostile intent, however, can itself generate hostility and anger” (Linehan 1993, 71).

Claims about gender norms for anger and other problems do not eliminate the need for concern about BPD anger. Excessive or inappropriate anger can be dysfunctional, because problems with anger can cause other life difficulties (Linehan 1993, 71) and can be disturbing or disruptive to the patients themselves. But Linehan, too, recognizes the gender norms at work in assessing anger in BPD patients: with respect to women, “even minor expressions of anger may be interpreted as aggression. For example, behavior that is labeled as ‘assertive’ in men may be labeled ‘aggressive’ in women. Perceived aggression begets retaliatory aggression, and thus the cycle of interpersonal conflict is born” (Linehan 1993, 71).

Women are in a double-bind when it comes to anger. If they are vehement in their expressions of anger, they may be viewed as pathologically angry. If they are already viewed as “too angry,” then their denying that they are angry is proof of regression. What the patient takes to be a real-time provocation can be interpreted
as evidence of acting out or projection and, given that standards for rationality are
themselves gendered, it is often difficult for a female patient to make her case that
a present hurt or injustice was, indeed, done to her. Clinicians run the risk of as-
suming that excessive anger is so much a part of a patient’s character that she is
unable ever to be reasonably angry. A patient’s sense of herself as an adequate
reasoner and moral agent is at stake here, and it is important to give her real-time
anger uptake.

Let me return to the opening vignette now and tell you what really happened – because the clinician did give uptake to the patient’s anger, and the outcome was constructive.

A concerned mother called me to raise the issue of whether some further, perhaps
different, treatment might be needed for her daughter. I responded by saying that
I would talk with the patient about this concern and would get back to her. After
talking with her daughter, I suggested that she call her mother while she was still
there and could participate on another line. The patient promptly began a lengthy
diatribe at her mother for interfering and poor judgment. It was late and I was
tired. After trying unsuccessfully to interject some comments, I put my finger
across my lips and went, “Shh,” to my patient. She ragefully slammed down the
telephone and left. She subsequently told me how she felt demeaned, how she felt
I was taking her mother’s side, and that she considered not coming back. I ac-
knowledged that it had been an unfortunate indication of my impatience and a
better means of getting to the issues I felt needed to be addressed should have
been used. (Gunderson 1984, pp.110–111)

The clinician acknowledged a real-time provocation and injury to which the pa-
tient was responding with anger. After admitting his role in the troubled interac-
tion, he reported that the session went on to explore other times the patient felt
demeaned. This simple acknowledgement emphasizes the primacy of taking real-
time BPD anger seriously and respectfully, even when its expression may in part
hark back to other wrongs, and even when the size of it may seem out of place.
Taking responsibility for even minor current provocations can advance therapy
and build a stronger therapeutic alliance.

7. Conclusion

Anger has a present as well as a past. To understand it and assess its appropriateness,
clinicians need to contextualize patients’ anger in relation not only to the
patient’s past injuries but in relation to possible present ones. I have argued that
clinicians need to give uptake to the anger of BPD patients and to consider it as
possibly warranted in real-time as well as doubled. Giving uptake should include considering the beliefs and injury behind the anger. Clinicians will do best to acknowledge occasions when their attitudes and assumptions may, in fact, be slighting or injuring their BPD patients. Giving uptake to real-time anger as prima facie distinct from past wounds defuses it and opens up a space to talk about past wounds and what about a present-time injury is an echo of the past.

Anger has a present and a past, but it also has a future. Because anger implies hope (Sharpe 2003, 35). Anger says, “I still believe in myself; I am self-respecting; I believe in a future in which the injuries I sustain are not all-encompassing, and the hope I bear is that my anger gets taken seriously by you.”

References


2. “Chris felt sick with fear and anger; no, not anger – anger implied hope; hatred.”


A confusion of pains

The sensory and affective components of pain, suffering, and hurt

Jennifer Radden

1. Introduction

For considering the relation of fact and value in emotion, the affective states comprising, and evoked by, pain provide an unmatched example.

Pain is subjective: whatever its origin and status, it enters our lives as a mental state, experienced directly by its subject, and known to others through phenomenological report. Its negative valence is consistent, strong, and seemingly intrinsic. We may adopt additional attitudes toward it (relief, dread, gratitude, or equanimity, for example), but normal pain is immediately abhorrent and undesirable. It comes in varied forms. Some is experienced as localized sensations, akin in many respects to bodily sensations like itches, throbs, or giddiness. Other pain, suffering and hurt is felt as an emotion, or a mood, rather than a sensation. Always, it is a cultural as well as a private state and a social as much as a biological one. For pain involves emotion: studies have confirmed that affective elements and higher cognitive states of expectation and memory frame and interpret even ordinary pain sensations.

These features of pain experience raise a host of conceptual questions, some quite preliminary. Is it important to distinguish localized painful sensations from the pain, suffering and hurt that come to us in the form of emotional states? Does our pain language admit of more and less literal usage, if so? Is its abhorrence intrinsic to all pain experience? The following paper explores questions such as these in light of contemporary neuroscientific and philosophical research about painful sensations. Although alert to the affective elements tempering painful sensations, that research has almost entirely ignored pain and suffering that is non-localized and non-sensory. An older philosophical literature deals with more obviously
emotional pain and suffering, but it, too, is incomplete. These omissions seem likely to limit our understanding of all pain experience.

The nature of the relation between pain experiences has particular urgency, in addition, because of the central part played by emotional pain within clinical and diagnostic psychiatry. There, as a common symptom of several disorders of mood and affect such as depression, as integral to a standard definition of mental disorder, and as central to diagnostic categories involving psychogenic or somatoform “pain disorders,” such pain requires clarification and analysis.

The psychiatric applications of the analysis undertaken here are discussed in Part 2. In Part 1., focus is on the scientific findings, together with some philosophical analyses, that might guide us toward a better understanding of the relation between the affectively framed painful sensations (s-pain) resulting from real or imagined tissue damage and the more straightforwardly emotional pain (e-pain) associated with disorders such as depression. Strong analogies unite these states, undeniably. But there are also phenomenological and conceptual contrasts distinguishing s-pain and e-pain that might discourage us from conflating the two. The purpose of this discussion is to lay out some of what is at stake in the way we depict pain with the hope of avoiding terminological confusion when, in the context of psychiatric analyses, these central states are described and defined.

2. Conceptual clarifications

2.1 Pain is not the stimulation of pain receptors

A much-quoted definition of pain from the International Association for the Study of Pain is good place to begin. It is a problematic definition, yet worth looking at not only because it has been so influential in the last couple of decades of pain research, but because its ambiguities and inconsistencies are emblematic of the seeming confusions in this area. Pain is: “An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage.” This brief definition is followed by a controversial note insisting that “Pain is always subjective…Activity induced in the nociceptor and nociceptive pathways by a noxious stimulus is not pain, which is always a psychological state, even though we may well appreciate that pain most often has a proximate physical cause” (IASP 1986, p. 250)

Most immediately pertinent for us, the IASP account resists treating pain as reducible to the stimulation of pain receptors. In this respect, the note following the definition confirms conclusions drawn in several philosophical analyses of painful sensations (Edwards, 1979, Grahak, 2001, Kripke, 1980, De Grazia, 1991,
Sullivan, 1995, Radden, 2002, Aydede, & Gizeldere, 2002). Without painful sensation, the activation of physiological pain centers (the nociceptive pathways) would not be considered an instance of pain, for pain itself “is an experience, not a stimulus or a response” (Sullivan, 1995, p. 278).

This fundamental point is confirmed by appeal to our use of the term “pain” in everyday language. And, in a recent paper, Murat Aydede and Guven Gizeldere have also pointed out that the whole conceptual apparatus of the scientific study of pain is built on subjective report, and on the separation of first person accounts of pain sensations from investigation into biological events. If painful sensations were reducible to nociceptive stimulation, this research would have been impossible. Thus, of scientific findings about pain sufferers, these researchers point out, “No objective observations of the causes of their condition, non-verbal behavior, and/or the brain damage involved, would all by themselves be strong enough to force scientists to seek functionally and anatomically separate brain mechanisms… the accumulation of this sort of (mostly) subjectively obtained abnormal data… led to the identification of the neural substrates through brain imaging studies: the phenomenology strongly guided what to look for, and where”(Aydede & Guzeldere 2002, p. 10). The findings of these studies are only conceptually coherent in the context where the separation between the phenomenology and the brain states is acknowledged. (With this established, these authors actually propose a nomenclature that acknowledges the dual contribution of neuroscience and phenomenology in the term “neurophenomenology.”)

Summing up this recognition that pain is more than the stimulation of pain receptors, we might use an old philosophers’ tag and say that in the case of pain, to be is to be experienced or perceived (esse est percipi).

Another aspect of the IASP account is worth attention. The brief definition concerns itself with painful sensations to the exclusion of more emotionally toned pain and suffering, apparently accepting, as many seem to do, that when “pain” is used of suffering not involving actual or potential tissue damage, it is employed in some extended or secondary sense. This position is adjusted – or even reversed – in the note following the definition, where it is acknowledged that reported pain in the absence of tissue damage “or any likely pathological cause,” should nonetheless be “accepted as pain.” The inconsistencies between the initial definition and subsequent note are of course confusing. But though rather begrudgingly (and only granting “pathological causes” originating from within the person rather than from, say, external events), the note does appear to allow the pain of depression the status of real pain, appropriately so described.
2.2 Pain is not merely an episode of simple, localized sensory experience

Focused as it has been on the painful sensations associated with tissue damage (real or imagined), pain research has nevertheless come to recognize that all pain is less simple, more cognitively mediated, and thus more like an emotion than had previously been supposed. The unsurprising bit of this that could hardly escape even pre-conceptual insight is that painful sensations comprise distinguishable phenomenal strands, some more sensory, and others, in particular the feeling of abhorrence or displeasure pain brings, more akin to other affective states. Then the surprising aspect: that every part of pain experience is extensively “cognitively mediated.” Such psychic elements as memory, personal and social attitudes, role expectations, and life experience, as well as mental and emotional health, and bodily traits, affect how pain feels.

A range of sources, including the subjective reports of patients having undergone brain surgery (prefrontal lobotomy and cingulotomy), brain imaging, and anatomical studies, have confirmed the first (unsurprising) distinction between “sensory-discriminative” awareness of what the pain is like in terms of quality and severity, and motivational-affective awareness of how intolerable it is (Melzack, 1961, Melzack & Wall 1983, Fernandez & Turk 1992, Fernandez & Milburn 1994, Treede et.al., 1999, Price, 2002). Interestingly, the term “painful” retains the ambiguity between these two traits in everyday (pre-conceptual) usage. Research subjects must be guided to distinguish and report a pain’s intensity separately from its unpleasantness.

Now, to the surprises. Pain experiences are mediated by elements that are complex and unpredictably idiosyncratic, revealing the extensive influence of higher order cognitive states. Efforts have been made to distinguish within motivational-affective awareness. H. L. Fields, for example, separates the “stimulus bound (primary) unpleasantness” from the “secondary unpleasantness” he identifies as a “higher level process” with a “highly variable relationship to stimulus intensity… largely determined by memories and contextual features”(Fields, 1999, S61). Yet primary forms of unpleasantness are also mediated by motivational-affective factors, so not even primary unpleasantness is consistently coupled with stimulus intensity. Voluntarily accepting a painful experience versus being forced to undergo it alters – and lessens – even the degree of (“primary”) unpleasantness of the experience, it has been shown (quoted by Hall, 1989, p. 654). And other studies have demonstrated that the perceived intensity of pain, as well as its (“secondary”) unpleasantness, are both mediated by expectations, beliefs and other cognitive states (Montgomery & Kirsch 1996). At best then, Fields can maintain that the affective influence on secondary unpleasantness will be relatively greater.
Common sense confirms the general point, now regularly noted by researchers, that pain is cognitively mediated this way. The pain of (natural) childbirth for the mother will likely be as stimulus intense and (“primary”) unpleasant as any she has endured, yet relatively bearable, on the dimension of “secondary unpleasantness,” because of what it portends, folk wisdom insists. It goes without saying that pain may be worth enduring for some greater good. And the fact that pain judged natural, healthy, ennobling, or otherwise instrumentally valuable, will be not only less abhorrent but less intense, is also widely acknowledged, along with lore about human suggestibility. The parent’s kiss and touch, children and adults believe, lessen the pain of children’s minor injuries – an assumption confirmed in recent research, where imaging has shown that even placebo “analgyics” activate the brain’s natural opiate-producers that serve to reduce the stimulus intensity and unpleasantness of nociceptive stimulation (Zubieta et al., 2005.)

The basic neuroscience of these observations is also quite well understood. Rather than a simple sensation, pain is modulated by influences from several parts of the brain including the prefrontal cortex that exercises executive control over all other cortical centers. More specifically, there are:

serial interactions between pain sensation intensity, pain unpleasantness, and secondary affect associated with reflection and future implications (i.e., suffering). These pain dimensions and their interactions relate to ascending spinal pathways and a central network of brain structures that process nociceptive information both in series and in parallel. Spinal pathways to amygdala, hypothalamus, reticular formation, medial thalamic nuclei, and limbic cortical structures provide direct inputs to brain areas involved in arousal, bodily regulation, and hence affect. Another major input to these same structures is from spinal pathways to somatosensory thalamic (VPL, VPM) and cortical areas (S-1, S-2, posterior parietal cortex) and from these areas to cortical limbic structures (insula cortex, anterior cingulate cortex). This cortico-limbic pathway integrates nociceptive input with information about overall status of the body and self to provide cognitive mediation of pain affect. Both direct and cortico-limbic pathways converge on the same anterior cingulate cortical and subcortical structures whose function may be to establish emotional valence and response priorities. This entire brain network is under dynamic top-down modulation by brain mechanisms that are associated with anticipation, expectation, and other cognitive factors. (Price 2002, p. 392)

As well as aches and pains, sensations (or what are sometimes known as “bodily sensations”) include such things as itches, tickles, muscle spasms, throbbings, giddiness, and dizziness. They have often been depicted as episodes of localized

---

1. Whether all mothers remember it that way – they do not – is another matter.
2. The mortification of the flesh in medieval asceticism highlights and exemplifies cultural differences in relation to such assessments. (See Kroll & Bachrach 2005.)
sensory experience that are simple, immediate responses to stimuli. In light of the above findings about the cognitive complexity of pain, we can conclude that even painful sensations are not merely sensations thus understood. (And it seems likely that many other sensations are equally complex.)

As the above quoted passage shows, pain is a composite whose separate sensory and affective elements have been identified. Moreover, in rare instances, these elements even detach from one another phenomenally. Cases of "reactive dissociation," for example, are those in which subjects report feeling the pain sensation without any accompanying affective component of unpleasantness. But these are abnormalities. In the normal case, the sensory and affective aspects of painful sensations are inextricably linked. Phenomenologically, they present themselves as a single, unified experience.

2.3 Painful sensations, like ‘merely psychological’ pain and suffering, may be psychogenically caused

Some of the strongest support for the conclusion that pain is not merely an unmediated sensory response comes from the evidence that pain sensations often occur in the absence of any identifiable tissue damage. They are "psychosomatic," as it is sometimes put. Apparent instances of painful sensations, caused not by tissue damage but by psychic states, are frequently documented in the psychiatric literature where, until the term was expunged for its political connotations, they were known as "hysterical" symptoms. Apart from the more controversial diagnoses such as fibromyalgia, whose psychogenic status remains contested, there is the pain-related diagnostic category of “pain disorder,” which exhibits presumed psychogenic underpinnings. Pain disorder is defined as the occurrence of pain in one or more anatomical sites when “psychological factors are judged to have the major role in the onset, severity, exacerbation, or maintenance of the pain.” (APA 1994, p. 461–2)

It is a given that psychogenic factors account for painful psychic states that are not sensations. What is at issue here, then, is not whether such psychogenically caused states occur but how they should be described. As it pertains to terminology, the question becomes whether the pain and suffering brought about, at least in part, by psychological states, is actually – and literally – “pain.”


4. There is "phenomenal unity" as Michael Tye defines it: a matter of simultaneously experiencing perceptual qualities entering into the same phenomenal content (Tye, 2003, p. 36)
2.4 Emotional pain is not a metaphor

The anguish often wrought by depression is as unpleasant as any painful sensation its sufferers experience. But is it pain? Its almost exclusive attention to painful sensations has allowed most recent pain research to avoid this initial matter of terminology. And, psychiatric writing often appears to regard the depression sufferer’s pain as “pain” in an extended or metaphorical sense, as was noted above, and or to restrict the term “pain” to painful sensations, and employ “suffering” to cover experiences like the depressive’s anguish.

Among philosophers, the terms “mental pain” and “physical pain” have been assigned to these two kinds of experience. (See, for instance, Trigg, 1970.) Others have spoken of “psychological” and “bodily” pain (Scarry, 1985). But this usage is misleading and problematic on at least two counts. To begin with, the sensation of pain is as quintessentially phenomenal (and hence, in the sense intended, “mental” or “psychological”), as are painful and distressing emotions. No less or more real than the pain from a blow on the head, and long allied to such pain, depressive anguish is as literally “painful” as, and no more “mental” or “psychological” than, pain from that blow.

Interesting here are recent remarks made by Dr Helen Mayberg, professor of Psychiatry and Neurology at Emory University, who has pioneered surgical depression treatment focused on Area 25 in the brain. She begins pointing out that although a common understanding of depression equates it with a form of deficit, her work suggests the reverse. Talk to a depressed person, she remarks “and you have this bizarre combination of numbness and what William James called ‘an active anguish.’ ‘A sort of psychical neuralgia,’ he said, ‘wholly unknown to healthy life.’ You’re numb but you hurt. You can’t think, but you are in pain. Now, how does your psyche hurt? What a weird choice of words. But it is not an arbitrary choice. It’s there. These people are feeling a particular, indescribable kind of pain” (quoted by Dobbs, 2006, p. 55, my emphasis). Mayberg is a specialist in the pain of depression and she goes to considerable lengths to emphasize that she speaks advisedly and literally when she describes the pain of depression this way. And she follows a long tradition. The suggestion that of two uses one is primary and the other secondary or metaphorical seems unsustainable.

Jamie Mayerfield proposes to use the term “suffering” to refer to pain such as that experienced in depression. (See also De Grazia, 1998.) Despite the considerable disanalogies, however, there are also strong analogies between different forms of pain. And neither “pain” nor “suffering” better fits one than another. (Moreover, we speak of hurting, and being in discomfort and anguish with reference to each kind of experience.) Thus preserving the term “suffering” for states that are not sensations seems arbitrary and confusing. It may be true that, as Mayerfield points
out, the Greek *lupé* was rather distortedly translated as “pain” by the Utilitarians, when the then broader “suffering” would have been more accurate (Mayerfield, 1999). But the word “pain” stuck, and it applies today in the broad way that “suffering” also does. (As Rem Edwards has remarked “Nonlocalized discomforts have been called “pain” time and time again in the discourse of both philosophers and plain men” (Edwards, 1979, p. 36).

Moreover it is instructive, if *lupé* spans both states, suggesting that the Greeks, too, were alive to the analogies encouraging us to use one word for both kinds of suffering.

Edwards’s solution to this terminological matter is to employ the terms “pain 1” and “pain 2” for distinguishing painful emotions and sensations. But the terms emotion pain (e-pain) and sensation pain (s-pain), that have been used to mark the same contrast, enjoy some descriptive advantage, so “e-pain” and “s-pain” will be employed to allow us to keep track of these two kinds of experience in what follows.

2.5 Although analogous, s-pain and e-pain differ

As we saw above, the affective, phenomenal element of its unpleasantness is characteristic of each kind of pain, and that element is moreover mediated by higher order cognitive states such as memory and expectation. What must be emphasized, then, is that s-pain is so named not because it is without any phenomenal affective elements, but because it alone involves sensations. In the normal case, s-pain comprises sensation and affective phenomenal attributes, and in this respect, it appears to differ from e-pain, which is experienced as an emotional state only. S-pain is named for the sensation it includes.

Indicative of the analogies between s-pain and e-pain is the fact that not just the term “pain,” but a broad range of others – such as “hurt,” “suffering,” “discomfort,” “anguish,” and “distress” – span these two kinds of unpleasant state. And important forms of behavioral expression are common to them both: gasps, cries, moans, grimaces and tears, for example. Indeed, a characterization of s-pain provided by Mayerfield could be read as applicable to both kinds of experience (although it was not so intended). Speaking of painful sensations, he remarks “..pain is a particularly useful model,”

Everyone at one time or another has experienced it…We recognize it instantly, and name it unerringly, when it strikes us. There are certain things known to cause it in virtually all people; and even when it has invisible or unlikely causes, we can recognize its occurrence in other people by characteristic cries, grimaces, and recoiling movements. In our own case we can specify with considerable precision when it comes and goes, when it grows more intense, and we are unlikely to con-
To the commonalities such as these that pertain to the way pain experience affects its sufferers, must be added another: pain is a cultural phenomenon. Framed, understood, and interpreted by social and cultural values and meanings, painful sensations as much as painful emotions reflect their sufferer’s place in the social world they find themselves in.

Confronted with these many similarities, it is tempting to presume that the category of “pain” is a unitary one. Yet despite these commonalities, s-pain must not be too hastily equated with e-pain, nor e-pain reduced to s-pain. Several interrelated phenomenological features seem to require us to acknowledge that there are systematic differences between these two sorts of painful experience. Each of these must be acknowledged and explored before we reach any final conclusion as to the best way to understand and portray the relationship between these two groups of experiences.

Eight conceptual and phenomenological differences will be dealt with in turn. Summarizing, we can say that in contrast to e-pains, s-pains are (i) spatially localized; (ii) more temporally localized; (iii) closer to being felt states by their nature; (iv) contain (in the normal case) elements of stimulus intensity and unpleasantness, not mere unpleasantness; (v) enjoy a stronger reportorial authority (while not immune from error); (vi) subject to a particular set of metaphorical descriptions; (vii) always intentional though not as fully intentional; and finally, (viii) not subject to appraisal in light of social norms.

Some of these differences are matters of degree rather than kind, and not all of (i)-(viii) are uncontroversial. Taken together though, (i)-(viii) constitute a set of differences sufficiently formidable to regard s-pain and e-pain as distinguishable. These may encourage us to speak of the term “pain” as having two senses, as some philosophers have wanted to do. Or, “pain” may be a looser (family resemblance) type of category of which s-pain and e-pain represent recognizable variants.

i. Most obviously, sensations are spatially localized while emotions are not. Several of the differences described in (ii)-(vii) are little more than implications of (i) and indeed, its localization has sometimes been treated as alone sufficient to mark the distinction between s-pains and e-pains. (See, for example Edwards, 1979, p. 44). We may be inaccurate in our belief as to where we feel the pain, but with the awareness of a painful sensation comes awareness of where it feels to be. (This feature of pain has given rise to an account of pain as “an emotion at a place”

---

5. It has been questioned whether this trait holds for all sensations. (See Armstrong, 1968.) But no-one challenges the claim at least as it is made of pain.
And this remains true even when, as after a fall, or suffering influenza, we might insist that the painful feeling is located everywhere in our body. “Nonlocalized feeling” is not the same thing as “universally localized feeling,” as it has been put, and: “The former has no definite bodily locus at all, whereas the latter seems to be present “all over”” (Edwards, 1979, 39–40).

Emotional pain is not similarly tied to location. It is true that emotions have sometimes been assigned bodily parts in our Western cultural traditions, as the term “heart ache” vividly illustrates. But modern anatomical knowledge has relegated such associations to the status of lingering metaphors. And when accompanied by somatic sensations – the lump in the throat that comes with a poignant sight, the sinking feeling in the stomach with apprehension – these sensations are regarded as distinct, and are only occasional accompaniments of emotions.

ii. Sensations are also more closely localized in time: they are particular episodes or occurrences with identifiable beginnings and endings. Emotions, in contrast, although their onset can often be timed to a particular occasion, are then frequently less like episodes or occurrences and more like dispositions to act or feel. We readily ascribe emotions to those who are evidently not experiencing them at that time. “She is sleeping, but I know she is still pained by her father’s rejection” is an unexceptional and fairly commonplace way of speaking. “She is sleeping and she is in great pain in her lower back,” seems more problematic. If back pain cannot as readily be attributed to the sleeper as can emotional pain, one way to explain this is to say that back pain is localized temporally as well as spatially.

Another, and more common way to explain the difference between “She is sleeping, but I know she is still pained by her father’s rejection” and “She is sleeping and she is in great pain in her lower back,” is to insist that there cannot be unfelt back pains.

So we get to (iii) s-pains bear a closer relation than do e-pains to being felt. This view, encapsulated in Gilbert Ryle’s remark (about s-pain) that “unnoticed pain is an absurd expression” (Ryle, 1949, p. 203), is widely accepted still. (See for example, Turski, 1996, p. 26).

The impossibility of unfelt (s-) pain has been recently challenged in the philosophical literature. The case of an s-pain that arouses a sleeper seems to contradict the view that there cannot be unfelt (s-)pains and has been raised as a philosophical puzzle or problem for the customary analysis of (s-) pain as a state that must be felt. Attempting to account for such cases of seemingly unfelt pains, Terry Dartnall separates what is from what is known, concluding: “the feeling of the pain didn’t spring into existence when you woke up, but got gradually worse until it woke you up…” What sprang into existence was “your awareness of it” (Dartnall, 2001, p. 99). Others have questioned this analysis. Decrying what he takes to be Dartnall’s misguided “problematizing” of (s-) pain, J. L. Garfield speaks of a mistaken ac-
count of introspective knowledge according to which “introspection gives us inner episodes veridically and in their totality” (Garfield 2001, 1, my emphasis). Necessary but not sufficient for the sensation of pain, stimulation of the nociceptive pathways would at best have been the prompt for the pain that occurred on awakening. In that respect, Garfield’s use of “totality” implies, it was a part, but not the whole, of the pain involved.

Without engaging with every aspect of the exchange over unfelt or unconscious s-pain (and other sensations), I would point out that the apparent puzzles around the case of s-pain that wakes the sleeper are not comparably worrisome when e-pain interferes with sleep. We toss and turn, apparently fretting while half, or fully, asleep; we reach consciousness with unaccountable (“objectless”) feelings of apprehension or gloom; our e-pains enter our dreams and nightmares, sometimes serving to wake us. And these states seem to be remembered on waking, not experienced de novo. Which if any of these experiences rank as unfelt e-pains, or even parallel the case of allegedly unfelt s-pains, may be debated. But at the least we can conclude that such unfelt, or partially felt e-pains, are commonplace and do not seem to merit the status of philosophical puzzles or problems.

So while the possibility of unfelt s-pains is debatable, and debated, that of unfelt e-pains is not, or is less so. As we saw in the case of the sleeping woman pained by her father’s rejection, e-pains are often ascribed (by others) in the absence of conscious awareness of e-pain by their subject. In addition, a kind of “unfelt e-pain” is presupposed in the important concept of masked depression.

Masked depression is frequently attributed within psychiatric theorizing and lore. It is introduced in several ways: where there is an attitude of manic insouciance or one of apathy rather than the depressed mood deemed appropriate or predictable; where certain behaviors “substitute” for the feeling (as when substance abuse is said to be an expression of underlying depression, for example); where such a substitution or conversion transfers the feeling into somatic symptoms; and where mechanisms such as denial, repression or dissociation are said to conceal the underlying e-pain. We do not need to accept every account of masked depression (many are dependent on dubious or arguable theoretical posits) to recognize that no comparable attribution seems to be made in respect of sensation pain. Arguably, we may be the recipients of stimulation to our pain centers that is prevented from entering consciousness. But as was emphasized earlier, the stimulation

---

6. Following the lead of Arthur Kleinman, cross cultural studies have been taken to suggest that depression in non-Western cultures is almost always “somatized,” i.e., experienced in the form of bodily ills rather than the conscious states of (e-) pain and distress that are its central characteristics in our culture (Kleinman 1988, Gaw 1993, Kirmayer and Young 1998 Moerman 2002). Rather than a marginal case, it is suggested, masked depression may be the paradigm, and misleadingly named.
of pain centers is not the same as pain. The term “pain” is reserved, so that only when it enters conscious awareness as an experience does it become actual “pain.”

iv. The elements of stimulus intensity and unpleasantness, respectively, are each present and phenomenologically identifiable in s-pain. This follows, of course, from the fact that s-pain is a felt sensation and stimulus intensity is a sensory measure. Because e-pain is not a sensation, and at most has somatic accompaniments, stimulus intensity and unpleasantness are indistinguishable in the experience of e-pain. (Even in William James’s theory of emotion which allows that emotions are responses to a felt somatic states, the relationship between these separate elements is a causal one: the affect and its sensory cause are only contingently connected (James, 1884).)

v. Also related to the localized nature of s-pains is a further apparent difference, if not of kind then at least of degree, concerning the reportorial authority accorded the subject. Reports on one’s own sensations generally go unchallenged. They are not deemed immune from error – to the contrary. But, sufferers are treated as authorities as to their own s-pains, as the case of phantom limb pain attests. We have no hesitation in speaking of pain here, even while acknowledging that reports of phantom limb pain must be, and must be known to their subject to be, inaccurate as to the location assigned to the pain. Our first person reports – as to whether s-pains are present, their degree of intensity, their location in space and time, and their degree of unpleasantness, and their other sensory attributes – are immune from correction, yet not from error. This feature of our everyday concept of (s-) pain, and pain reports, as Aydede has pointed out, distinguishes (s-) pain reports from perceptual reports, for example.

Despite superficial grammatical similarities, Aydede argues, different truth conditions govern visual reports such as “I see a dark discoloration on the back of my hand”(1) and pain statements such as “I feel a jabbing pain in the back of my hand”(2). By uttering the latter (2) he notes, “I am saying something like “I am undergoing an experience which tells me that some sort of physical disturbance is occurring in the back of my hand.” If so, that there is no physical disturbance occurring in my hand…doesn’t make (2) false. The fact that I can still correctly point to where it really hurts in my hand after hearing from my doctor that nothing is wrong with my hand is explained by reinterpreting what I say and do with that gesture: I am still undergoing an experience which represents my hand as having something physically wrong with it.”(Aydede 2006, p. 5)

As references to masked depression seem to indicate, we do not generally enjoy the same reportorial authority when it comes to e-pain. (See, for instance, the remark that “introspection of emotional states is so much less reliable than that of other states of consciousness” (Seager, 2002, p. 666).) The truth conditions for “I am pained by my father’s rejection” are not the same as those for the visual claim
in (1) above (“I see a dark discoloration on the back of my hand”). But nor are they as immune from revision and correction as (2) (“I feel a jabbing pain in the back of my hand”). E-pain statements apparently fall somewhere between s-pain statements and ordinary perceptual claims with respect to this epistemic dimension of reportorial authority. Two factors likely account for this status. E-pain, as we have seen, is often best understood not as episodic suffering but as longer term states and dispositions; moreover, it is more complex than s-pain because it usually comes embedded in a more extensive network of consciously held cognitive states. These differences can explain why we are treated as more prone to error, to exaggeration, and to distortion, in our reports of e-pain.

vi. **S-pain resists literal description.** In Elaine Scarry’s words, it “has no voice” and “shatters language” (Scarry, 1985, pp. 3, 5). Recognizing that the conventional medical scale of mild to severe captured only one aspect of the phenomenology of s-pain experience, researchers have sorted and classified the metaphors employed by patients into three groups, those referring to “temporal” aspects of the experience (“quivering,” “throbbing,” and “pulsing,” for example); those the “thermal” (such as “burning,” “scalding,” “searing”), and those the “constrictive” (“pinching,” “crushing,” and “cramping”) (Melzack & Wall 1983). In their appeal to metaphor, and their reference to sensory attributes, these descriptions are distinctive to our efforts to convey the experience of s-pains, and are without parallel in descriptions of e-pains.

vii. A related aspect of s-pain apparently distinguishes it from e-pain: as sensations s-pains are not about or of anything beyond themselves. This is sometimes what is meant when they are described as “intransitive.” In contrast, emotions and so e-pains are intentional, that is, they are usually about, over, or directed toward, intentional objects that are, or may be, beyond themselves (Gordon, 1987). These questions of intentionality are complex and contested. In what follows, it will be shown that while s-pains are always intentional, they differ from e-pains in being only minimally so. They are over or about themselves, i.e., they are intransitive. E-pains, in contrast, may be transitive or intransitive. When they are intentional, they are fully and richly intentional. Their objects include other inner states, things, states of affairs, propositions – or themselves.

---

7. Although beyond the scope of this discussion, the political and power implications of this feature of pain’s inexpressibility, are explored and developed in Scarry’s important work (Scarry, 1985).

8. Armstrong uses the term “intransitive” in a rather different way. He distinguishes (transitive) sensations such as touch, and inner (intransitive) sensations such as pains. But because of his perceptual theory of sensation, Armstrong goes on to attribute a “concealed transitivity” to pains as well (Armstrong, 1968, p. 309).
As this contrast suggests, the notion of an intentional object, i.e., that toward which the e-pain is said to be directed, over, or about, requires refinement. It may be either a situation or state of affairs captured propositionally (such as the proposition that my friend has died) or a concrete existent in the world, such as a (living) person, a bodily state, or indeed, a psychic state such as, and including, itself.9 (When I anguished over the effects of my debilitating depression, for example, my own pain and suffering may be the object of my anguish. My depression depresses me.)

The differences just noted may seem attributable to the fact that an adverbial rather than an act-object analysis fits the experience of pain. Understanding (s-) pain on an act-object model, it has been claimed, involves a misapprehension. If all pain is better understood adverbially, then any disparities between e-pain’s and s-pain’s respective transitivity and intransitivity will be beside the point. On this view, pain – and, it is often asserted, pleasure – are properly understood in adverbial terms: they are ways of experiencing something, rather than something we experience. ‘(S-) pain’ now becomes an adverb describing the way we feel, not the thing we feel. In the reasoning of a recent such account, “When we describe a pain, we are … qualifying a verb rather than a noun… “I feel a sharp pain” is an answer to the question “How do you feel?” not to the question” What do you feel?”” (Douglas, 1998, p.129).

Precisely because, as we saw earlier, e-pains are not intransitive and take a range of objects that typically go beyond themselves, the act-object account seems the more obviously applicable. But even in the case of s-pains an adverbial analysis encounters difficulties. At least on its surface grammar, “I feel a sharp (s-)pain” does exhibit an act-object structure – my feeling is over or about the felt (s-)pain. (And indeed, since if s-pain were a kind of perception, it would also invite an act-object analysis, some have found a way to accommodate s-pain within a representational theory of consciousness by proposing that s-pain is an object of experience (Langsam, 1995, Bryne, 2001). It may sometimes also be an answer to “How do you feel?” But depending on the context, “I feel a sharp pain” sometimes seems to be an answer to “What do you feel?;” just as, colloquially, “despair” may be an answer to both “How do you feel?” and “What do you feel?” Applying an adverbial analysis to e-pains would seem to involve similar, context-dependent ambiguities.

The adverbial account is incomplete, then. And concerns over it such as these encourage us to explore where we might stop short of adopting such a position, in order to maintain, as I wish to do, that “I feel a sharp pain” might as equally be the answer to the question “What do you feel?” as “How do you feel?” Thus Trigg, looking at disanalogies between e-pain and s-pain (in his terminology “mental

9. For a careful discussion of this point, see DeLancey, 2002, Chapter 5.
pain” and “physical pain”) finds another feature, not noted thus far. The main difference between the experiences Trigg names mental and physical pain, he proposes, may be “the type of “object” which each has. ‘Mental pain’ could be distress at situations, while ‘physical pain’ would be distress at sensations.” In that case, he goes on, it need not be at all surprising that “…it is logically necessary for someone suffering ‘physical pain’ to feel something. If the sensation were absent, there would be no ‘object’…and hence no pain.” (Trigg, 1970, p. 7 my emphasis).

The difference between e-pains and s-pains, Trigg is asserting here, does not lie in the presence or absence of intentional objects, so long as we remember that these may be either propositions or concrete things in the world. Both kinds of pain have objects, and conform to the basic act-object model. But the sole object of s-pain (and any sensation) is the sensation itself. Sensations may also sometimes be the objects of e-pain, but the considerably richer intentionality of emotions allows that e-pain’s objects include (propositions about) situations and states of affairs, not merely sensations.

In a recent review and sorting of intentionality theses Alex Byrne distinguishes theories according to scope, and arranges them into “unrestricted” and “restricted.” Unrestricted intentionalists hold, while restricted intentionalists deny, he says, that intentionalism also applies to bodily sensations (such as s-pains) (Byrne, 2001, p. 205). Trigg’s account, then, is unrestricted. Unrestricted intentionalism for bodily sensations is the position favored by Byrne himself, on the grounds that bodily sensations possess intentional content related to their phenomenally displayed location in the form of a proposition. Whenever the experience or sensation is endured, as he puts it, “the world seems a certain way, namely, that there is a twinge in the knee” (Byrne, 2001, p. 229). While also maintaining an unrestricted theory, Trigg’s account diverges here: for him, the intentional object is the sensation itself, whereas for Byrne it is a proposition about the sensation. Setting aside these close differences, however, it is possible to recognize that in either variant of the unrestricted intentionality that allows sensations to be intentional objects, the objects of sensations can only be the sensations, or propositions about them. And in this respect, at least, s-pains differ from e-pains, whose intentionality has a broader reach.

Only clear and unequivocal examples of s-pain and e-pain have been described thus far in our discussion. But before we leave this question of intentionality, attention must be drawn to e-pains that are ostensibly without objects, or whose objects are so vaguely defined and pervasive as to be all-encompassing. States such as “uneasiness,” “fearfulness,” “jitteriness,” or “disspiritedness” arguably lie somewhere in between emotions and sensations, and so might fall under some third, hybrid category.
Certainly, aspects of these states can be pointed to that would account for our inclination to regard them as hybrids. First, these are each apparently objectless emotions, detached from the framing intentional structure that makes them about or over something in particular that is beyond themselves. Unlike many emotions, these ones may present themselves as “intransitive,” and closer, in that respect, to sensations. Moreover, each is associated with a well-defined set of sensations that often accompany them. The descriptive terms themselves (“uneasiness” etc.), we can thus suppose, have come to connote their sensory accompaniments. When we think of dispiritedness we are reminded of the felt bodily slump and drooping posture associated with it, and so on. That the terms used to describe these states carry sensory and emotional connotations is undeniable. Yet while the ways we describe them are genuinely ambiguous – “dispiritedness” may allude to either the sagging sensation or to the disheartened frame of mind, for example – the sensations and emotions associated with them remain separable states.

Rather than preventing us from distinguishing e-pain from s-pain by appeal to Trigg’s characterization of the respective nature and scope of their objects, these moods that are sometimes deemed “objectless” reveal the importance of the stress I have placed on the passage from Trigg quoted above. (“...it is logically necessary for someone suffering ‘physical pain’ to feel something. If the sensation were absent, there would be no ‘object’...and hence no pain.”) S-pains must have objects, in the minimal sense involved with intransitive sensations; e-pains need not have them, although when they do, their objects may be sensations or any other inner states, things, situations, or states of affairs – or propositions about any of these. It is not intentionality as such that separates e-pains from s-pains on this analysis, as restricted intentionalism asserts, but rather the type and complexity of that intentionality.

Summing up: the difference between e-pains and s-pains vis à vis intentionality comes to this: S-pains are never “objectless,” as e-pains sometimes appear to be; instead, s-pains are always, but only, minimally intentional, over or about themselves, i.e., they are intransitive. When they are intentional, e-pains are fully and richly intentional, they may be over or about inner states, things, states of affairs, propositions – or themselves.

(viii) S-pain is not judged as appropriate or not. Unlike s-pain, e-pain is subject to appraisal – as appropriate to its circumstances, proper, understandable, even reasonable, or not. This is an implication of its intentionality: such appraisal is based on the aptness of a relation, that between the feeling itself and its object and or occasion. This particular characteristic of e-pain has been proposed as the basis, or part of the basis, for distinguishing pathological states of depression and sadness from more normal and appropriate sadness. Thus, Alan Horwitz and Jerome Wakefield speak of “contextuality” as an inherent, normal aspect of many psychological mechanisms: by this they mean “they are designed to activate in
particular contexts and not to activate in others ... innate mechanisms regulate reactions of sadness, despair, and withdrawal naturally come into play after humans suffer particular kinds of losses” (Horwitz and Wakefield 2007: 15). When they reflect pathology, on this account, loss responses “emerge in situations for which they are not designed, they can be of disproportionate intensity and duration to the situations that evoke them, and ... they can occur spontaneously with no trigger at all” (Horwitz and Wakefield 2007: 15).

2.6 Separate e-pain may accompany s-pain

S-pain has several affective components but not all are phenomenal qualities of the experience. We are aware of the feeling of displeasure, while the affectively toned beliefs, expectations, memories and other states forming these additional ingredients occur beneath—or before—conscious awareness. (They are “sub-doxastic” (Stich) or “sub-personal” states, inaccessible to consciousness (Stich, 1978).)

Independent of the other affective elements framing and shaping our experience of s-pain beneath the surface of conscious awareness, it is also true that separate e-pain sometimes accompanies the experience of s-pain. I may be distressed that the (s-) painful sensation is so unpleasant, for example. (Consider this case: my daughter and I are both to have a medical procedure. I go first. I had persuaded my daughter it would not be painful. It is painful, more than I expected. The s-pain I experience might here be the object over which, because of a sense of having violated my daughter’s trust, I experience (mild) e-pain.) But the unpleasantness of s-pain is a basic phenomenal experience intrinsic to the normal experience of s-pain, and such e-pain is at most a contingent accompaniment.

2.7 There can be no co-morbidity between s-pain and e-pain

Just as s-pain normally involves the affective, phenomenal element of its unpleasantness, conditions such as depression may be similarly accompanied by painful sensations. Empirical data attest to the apparent “co-morbidity” of chronic pain (s-pain) and psychiatric disorder (e-pain); moreover, some studies have also been taken to demonstrate a particular link between chronic pain and depression. Neither alleged link has gone unchallenged (Hardcastle, 1999). But whether or not they are commonly present alongside depression, these pain sensations (s-pain)
do not prevent us from describing the experience of depression as e-pain.\textsuperscript{10} For such sensations are portrayed as accompanying the e-pain of depression, rather than being identified with it. As the term “co-morbidity” implies, the sufferers of chronic pain (s-pain) alongside depression (e-pain) suffer (when they do) two disorders, not one with two distinct pain-related symptoms.

3. Pain in psychiatry

3.1 E-pain is a symptom and indicator of mental disorder

As I have argued elsewhere, it behooves us to understand the concept of e-pain because it is one that enters importantly into clinical psychiatry (Radden 2002). To briefly summarize here, let me point to three examples, all derived from the current diagnostic and statistical manual (APA, 1994). First, the Introduction to DSM-IV employs a definition of “mental disorder” which includes reference to what appears to be e-pain in the phrase “present distress (e.g. a painful symptom)”; second, what appears to be excessive e-pain is noted as a defining symptom of several different mood disorders, including depression; and finally, the diagnostic category of “pain disorder” employs a definition which maintains ambiguity over whether the pain that is the basis for the diagnosis is s-pain or e-pain, when it appears to require both kinds.

The first of these references, at the introduction to the volume, notes that mental disorder is “conceptualized as a clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that is associated with present distress (e.g., a painful symptom) or disability (i.e., impairment in one or more important areas of functioning) with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom”(APA 1994, p. xxi, my emphasis). A general concern with this passage (and especially with the use of the parenthetical example) is its ambiguity: we are left unclear whether something like the distinction between e-pain and s-pain is recognized, and the “present distress” and “painful” symptom referred to describes e-pain. That passage seems to be an offspring of one in the revised previous edition where each mental disorder is said to be “associated with present distress (a painful symptom)” (DSMIII-R 1987, xxiii). The parenthetical paraphrase here apparently serves to indicate that the distress is equated with, rather than merely exemplified in, a painful symptom, but

\textsuperscript{10} Interestingly some of the newer classes of antidepressants and second-generation antiepileptic drugs have proven effective in the treatment of chronic pain. This finding results from the analgesic effect of these drugs, however, not their antidepressant effect, and is believed to modulate pain transmission by interacting with specific neurotransmitters and ion channels.
A confusion of pains

this formulation, too, remains ambiguous. In so central a definition, the relation between present “distress” and the pain of a painful symptom require more thorough and careful analysis.11

The second place we find reliance on the category of what at least appears to be e-pain, is in definitions of affective or mood disorder. References are to “depressed mood,” as a symptom of a range of conditions; and the “symptoms” of Major Depressive Episode are said to include “feelings of worthlessness or guilt,” “recurrent thoughts of death,” “[reports of feeling] depressed, sad, hopeless, discouraged, or “down in the dumps”” and “clinically significant distress” (APA 1994, p. 320). Other “symptoms,” some psychological and some not, are also cited as diagnostic for these sorts of disorder, but I have selected the group of states and reports quite clearly bespeaking e-painful, rather than s-painful phenomenology. Because there is no effort to otherwise explain and define the central category of “depressed mood,” it is arguable that the states listed here are more privileged than some others. Be that as it may, these states appear to be instances of e-pain rather than s-pain, and this status requires some additional acknowledgement.

The definition of “pain disorder” was introduced already. In DSM-IV this condition is said to be “characterized by pain as the predominant focus of clinical attention.” (APA 1994, p. 445) Left unclear is what kind of pain. Pain disorder comes under the broader category of Somatoform Disorders (including Somatization Disorder, Undifferentiated Somatoform Disorder, Conversion Disorder, Hypochondriasis, Body Dysmorphic Disorder and Somatoform Disorder Not Otherwise Specified), said to all share “the presence of physical symptoms that suggest a general medical condition” and that “cause clinically significant distress” (APA 1994, 445, my emphasis). Physical symptoms of the kind identified here need not be painful sensations. (Rashes, for example, may be prickly without being painful.) So while they are believed to have psychogenic causes, the symptoms of pain disorder (and other Somatoform disorders) must be accompanied by distress over or about them – and hence e-pain – in order to warrant this diagnosis. Again, we have a confusion of pains: the “pain” of pain disorder that is the focus of clinical attention seems likely both s-pain and e-pain.

As the architects of future diagnostic and statistical manuals prepare for a revised edition, they should know that with some preparatory acknowledgement of the complexities involved, and the use of consistent language and clear examples, the ambiguities and confusions identified here could be remedied.

11. There are several other critiques of this passage, the most telling of which is by Russell who points out that its employment of the medical term “symptom” involves both a failure to apply and a misapplication of, the medical model (Russell 1994,247)
4. Conclusion

This exploration aimed at conceptual clarification in a complex and important area of psychology, with the ultimate goal of avoiding confusion over pain, suffering and distress as they appear in psychiatric theorizing and lore. Particular points of emphasis were selected based on common ambiguities and misunderstandings over what have sometimes, misleadingly, been termed “physical” (or bodily) and “mental” (or psychological) pain and suffering. It was argued that (i) pain is not merely the stimulation of pain receptors; (ii) pain is not merely a localized episode of sensory experience; (iii) painful sensations may have psychogenic causes; (iv) “emotional pain” is not a metaphor; (vi) although analogous, s-pain and e-pain differ; (vii) separate e-pain may accompany s-pain; and (viii) there can be co-morbidity between s-pain and e-pain.

Because strong analogies also link these two kinds of pain experience, the purpose here was merely to lay out and acknowledge similarities and differences such as these, rather than to insist, as some have, that “pain” possesses two senses.

More generally, the method and epistemological presuppositions employed here have influenced these conclusions in the sense that some of them derive from phenomenological reports only available to conscious awareness. That suggests no new brain science will alone serve to unseat these findings, which are conceptual rather than empirical. While it might encourage us to reconsider the weight we accord the disanalogies outlined above, even the discovery that apparently identical neurons fire, in identical fashion, when e-pain and s-pain occur, would not require us to disregard the distinction between e-pain and s-pain pointed to in this discussion.12

References


---

12. I am grateful to Professors Lawrence Kaye and Murat Aydede, each of whom read and offered suggestions on earlier drafts of this paper. In addition, I have benefited from comments when this material was presented at the James Martin Advanced Research Seminar, at the Philosophy Faculty, Oxford University in November 2006, with a commentary by Guy Kahane, and by incisive help from Frank Keefe.


Ethical implications of emotional impairment

Abraham Rudnick

1. Introduction

Psychiatry systematically describes abnormal mental phenomena in order to understand them and treat individuals afflicted with them. The descriptive enterprise of contemporary psychiatry is documented in contemporary psychiatric classifications such as DSM (American Psychiatric Association, 2000) and ICD (World Health Organization, 1992). Yet psychiatry is not only descriptive but also evaluative, in that it ascribes values to abnormal mental phenomena. These values are of various types, but most important for our purpose are moral values ascribed to abnormal mental phenomena. The goal of this chapter is to examine implications of emotional impairment regarding moral evaluation in psychiatry, i.e., ethical implications of emotional impairment. Thus, fact and value in (abnormal) emotion will be addressed here.

The moral evaluation enterprise of psychiatry is perhaps most conspicuous in the prominent psychiatric practices of involuntary commitment, the insanity defense, and the determination of competence to consent to treatment. In these practices, the individual afflicted with an abnormal mental phenomenon is assessed psychiatrically for (risk of) harm to self or others, for criminal responsibility, and for decisional capacity, respectively. These practices address moral or ethical issues regarding autonomy, beneficence (and non-maleficence), and justice.

Moral evaluation in psychiatry has been based by and large on cognitive criteria, rather than on criteria addressing emotions. This is illustrated in widely agreed-on criteria for involuntary commitment, for the insanity defense, and for competence to consent to treatment. Involuntary commitment, which typically involves forced psychiatric hospitalization (but not necessarily forced treatment), usually requires disrupted reality testing on the part of an individual, i.e., psychosis, or cognitive impairment – typically mental retardation, dementia or delirium – that lead to (risk of) harm to self or others. Psychosis classically consists of delusions and
hallucinations (Rudnick, 1997), of which the former is a thought disorder and hence primarily cognitive, and the latter is a perceptual disorder and hence cognitive too. And cognitive impairment is cognitive by definition. Thus, involuntary commitment is based on cognitive criteria, rather than on criteria addressing emotions.

The insanity defense, which typically involves the judicial removal of criminal responsibility from an offender, traditionally requires a disruption in the ability to know the difference between good and bad. Such an ability and its disruptions are primarily cognitive. Non-cognitive criteria, such as those referring to impulsiveness, have been discussed and have been largely rejected in relation to the insanity defense (Insanity Defense Work Group, 1983).

The determination of competence to consent to treatment, which typically involves assessing decisional capacity of an individual to consent to or refuse treatment, standardly requires a disruption in expression of choice, in understanding, in reasoning or in appreciation (of relevance of information to self) (Grisso & Appelbaum, 1998). Expression of choice is primarily motor rather than cognitive or emotion-related. Understanding and reasoning are primarily cognitive. And there is a debate whether appreciation is primarily cognitive or also emotion-related (Appelbaum, 1998; Charland, 1998).

As demonstrated, mainstream moral evaluation in all these three prominent psychiatric practices is based primarily on cognitive criteria, and not on emotion criteria. I claim that this holds true for most if not all contemporary psychiatric practice. Some recent research in the history of psychiatric ideas by Charland (2007) suggests that a process of disengaging emotion criteria from moral evaluation in psychiatric practice occurred in the 19th century.

Yet emotional impairments and hence emotions are common and central in psychopathology. Many psychiatric disorders manifest emotional impairment, either as core diagnostic symptoms, such as in bipolar disorder (American Psychiatric Association, 2000), or as prevalent comorbid symptoms, such as in eating disorders where depression is common (Woodside & Staab, 2006). Emotional impairments are central in psychopathology as they explain part of the disruption – to quality of life as well as to functioning – characteristic of psychiatric disorders, such as schizophrenia (Bowie et al, 2006; Norholm & Bech, 2006).

As psychopathology is both descriptive and evaluative, and as emotions are common and central in psychopathology, as suggested above, emotional impairments – and hence emotions – should be examined regarding moral evaluation in psychiatry. This has not been done to date, at least not across psychiatric disorders and moral evaluation practices. This chapter will examine implications of emotional impairment regarding moral evaluation in psychiatry, i.e., ethical implications of emotional impairment. The approach used will be that of formulating a conceptual framework for analysis, using it to examine three illustrative psychiatric
disorders in relation to three moral evaluation practices in psychiatry, and then attempting to generalize from these three examinations.

2. Framework

I will use principlism to analyze the ethical implications of emotional impairments. Principlism is the most commonly used approach in health care ethics, and it comprises considerations of autonomy, beneficence (and non-maleficence), justice and context (Beauchamp & Childress, 2001; Hebert, 1996). Autonomy refers to respect for persons and their choices, e.g., respecting a patient’s decision to refuse treatment. Beneficence and non-maleficence refer to maximizing benefit and minimizing harm to persons, e.g., using an effective treatment that has few adverse effects for a patient. Justice refers to fairness to all involved, e.g., allocating health care resources in such a way that patients most in medical need – rather than other patients – receive these resources first. Context refers to the particular circumstances in a given situation, e.g., taking into consideration legal constraints on health care decisions.

Principlism has been criticized as being overly individualistic, westernized, impartial, abstract, and theoretically impoverished. There are also strong rebuttals to such criticism (Gillon, 2003). As principlism is a practical, well-known and relevant approach to prominent ethical issues in psychiatry (Bloch & Green, 2006), it will be used here to examine ethical implications of emotional impairments in major depression, in narcissistic personality disorder, and in schizophrenia. These three psychiatric disorders are sufficiently distinct from each other and demonstrate different pertinent emotional impairments, so that it may be possible to generalize from their examination.

3. Ethical implications of emotional impairment in major depression

Major depression is characterized by one or more major depressive episodes that last at least two weeks (or less if successfully treated) and cause the afflicted person significant suffering and/or disruption to normative functioning. A major depressive episode may consist of neurovegetative symptoms such as disrupted sleep, appetite and sexual drive, of cognitive symptoms such as hopelessness, guilt and memory loss, of behavioral symptoms such as suicidal attempts, neglect of activities of daily living (such as grooming) and truancy (in adolescents), and of emotional symptoms such as sadness, anger and apathy (American Psychiatric Association, 2000).
The most pertinent symptom to our analysis, apathy, may manifest in particular towards the self, where the person does not seem to care much about himself or herself, to the point of not caring whether they live or die. Such impoverished caring about self can manifest more mundanely as not caring about – and hence not preventing – lost income, ill-health and disconnected relationships. This apathy explains part of the prevalent complications of major depression, such as unemployment, general medical morbidity and social isolation.

It is helpful for our purpose to examine the implications of this emotional impairment, i.e., apathy, and more specifically impoverished caring about self, for the determination of competence to consent to treatment. As noted above, the determination of competence to consent to – and refuse – treatment standardly addresses expression of choice, understanding, reasoning and appreciation (of relevance of information to self). In a previous publication, I argued at length that this construct of competence is lacking in that it does not explain the demonstrable disruption of competence due to impoverished caring about self that is sometimes associated with major depression (Rudnick, 2002a). In brief, the argument is that in major depression, expression of choice, understanding, reasoning and appreciation may be intact, yet competence may be disrupted. This is because impoverished caring about self can lead to the afflicted person not caring about his or her treatment vs. non-treatment and their outcomes.

If a person does not care about the outcome of his or her decisions, particularly in relation to matters that have grave personal implications such as morbidity and mortality, as is relevant to health care, I claim that person is not competent to decide on their treatment, i.e., to either consent to or refuse treatment. Thus, impoverished caring about self, which is an emotional impairment that can be caused by major depression, can result in incompetence to consent to treatment.

To examine the ethical implications of impoverished caring about self found in depression, I will make use of the ethical principle of autonomy. The notion of competence – to consent to treatment and in general – is associated with the notion of autonomy. This is so because, as noted above, autonomy refers to respect for persons and their choices, and competence refers to decisional capacity. As decision making assumes choice and results in choices, it follows that autonomy is associated with competence. If so, incompetence involves deficient autonomy of that person. This can be shown for the standard (cognitivist) construct of competence, where an inability to reason coherently restricts fully autonomous choice. But it can also be shown for incompetence associated with the emotional impairment of impoverished caring about self, where choice is restricted and possibly nullified.

An ethical implication of such deficiency of autonomy associated with incompetence is that a person deemed incompetent would not be allowed to decide on his or her own treatment but rather a substitute decision maker would be assigned
to decide on that person’s treatment. Thus, an ethical implication of the emotional impairment of impoverished caring about self found in major depression is that if such an emotional impairment is disruptive enough to lead to incompetence, as it could be, the person afflicted with it may not be allowed to decide on his or her own treatment but rather a substitute decision maker would be assigned to decide on that person’s treatment. This demonstrates that moral evaluation in psychiatry can include emotion criteria, at least regarding ethical considerations related to competence and autonomy.

A caveat is that the determination of incompetence to consent to – or refuse – treatment is perhaps more difficult in relation to emotional impairment such as impoverished caring about self, as compared to cognitive impairment. If a person declines anti-depressant treatment when he or she has a major depressive episode, it may be difficult to determine whether his or her treatment refusal is due to not caring about self, which is directly induced by the depression, or due to non-pathological although unusual beliefs and attitudes, such as a fatalistic belief that whatever happens is pre-determined, unrelated to personal choice. Historical information about that person’s beliefs and attitudes when not depressed may be helpful here, but is not always available or reliable and may not be conclusive (Rudnick 2002a). Still, this may be primarily a clinical problem rather than an ethical problem, and thus may not detract from the conclusion reached above.

4. Ethical implications of emotional impairment in narcissistic personality disorder

Clinically significant narcissism is a personality disorder. The notion of personality disorder has been defined and characterized in various ways. One important characterization of personality disorders is that of Cloninger, who describes personality as the influence of character and temperament, where temperament refers to the innate, genetic, and constitutional influences on personality, and character refers to the learned psychosocial influences. Cloninger hypothesizes that temperament has formed measurable biological substrates – novelty seeking, harm avoidance, reward dependence, and persistence – whereas character has three quantifiable factors: self-directedness, cooperativeness, and self-transcendence. He believes that personality style reflects the individual’s temperament factors plus positive or high scores on the three character factors. Conversely, personality disorders reflect negative low scores on the three character factors (Cloninger et al., 1993; Sperry, 1995, p. 4). Thus, character abnormality is considered a central contributive factor in personality disorders, particularly abnormality in relating to others.
Narcissistic personality disorder is characterized by DSM-IV-TR (American Psychiatric Association, 2000) as consisting of (a) a pervasive pattern of grandiosity (in fantasy or behavior), (b) need for admiration, and (c) lack of empathy, beginning by early adulthood and present in a variety of contexts. Psychoanalytic characterizations of narcissistic personality disorder emphasize fragile self-esteem that requires affirmation from others as the fundamental characteristic of abnormal narcissistic personality organization (McWilliams, 1994).

The particular narcissistic abnormality in relating to others manifests as a sense of entitlement and consequent phenomena such as self-centered use of others. Most pertinent to our analysis, self-centeredness is related to impoverished caring about others. Such impoverished caring about others manifests in behaviors such as temper tantrums, sometimes to the point of uncontrolled rage and aggression. This occurs when the person feels that his or her needs are not met, be it by strangers or by significant others. This self-centeredness explains part of the prevalent complications of narcissistic personality disorder, such as the lack of close social relationships.

It is helpful for our purpose to examine the implications of this emotional impairment, i.e., self-centeredness, and more specifically impoverished caring about others, for the insanity defense (and, more generally, criminal responsibility). As noted above, the insanity defense traditionally addresses the ability to know the difference between good and bad, which is widely understood to be mainly cognitive.

In a previous publication, myself and a co-author argued at length that this (cognitivist) construct of the insanity defense, even when strengthened conceptually, does not allow the removal of criminal responsibility from offenders due to a personality disorder (Rudnick & Levy, 1994). This argument would apply to narcissistic personality disorder as a particular type of personality disorder. In brief, the argument is that, historically, there are two types of reasons for removing criminal responsibility from an offender: ignorance, implying lack of (correct) knowledge, and compulsion, implying lack of control.

It can be demonstrated that rationality is a central constitutive factor of responsibility, hence irrationality is a distinct and fundamental excuse allowing removal of criminal responsibility from an offender, which is the grounds for the insanity defense. Rationality refers here to practical reasoning, which can be defined as the explanation of an action by reasons, formulated as a set of beliefs and desires. For instance, the action of killing may be explained by a belief that the killer was under (otherwise inescapable) mortal threat by the killed person, in conjunction with the desire of the killer to save his/her own life (amounting to self-defense). Irrationality consists of invalid reasoning from a set of beliefs and desires to the action explained or an unacceptable set of beliefs and desires.
Unacceptability refers to incompatibility of the particular set of beliefs and desires with the major body of beliefs and desires held by the person.

If an offender is rational, then I claim that the insanity defense cannot be used to remove criminal responsibility from him or her. It can be shown that people with personality disorders are rational in this sense, as their practical reasoning is valid and acceptable – although perhaps sometimes unpalatable – in the above sense (at least so long as they have no pertinent psychiatric comorbidity).

In particular, the impoverished caring about others associated with narcissistic personality disorder does not disrupt the validity of practical reasoning. It also commonly pervades and is compatible with the major body of beliefs and desires held by the narcissistic person, which is actually an indicator of the narcissistic diagnosis and even more so of its severity. If so, the traditional (cognitivist) construct of the insanity defense cannot be used to remove criminal responsibility from an offender due to a narcissistic personality disorder.

Yet, I propose that if a person’s impoverished caring about others is a result of a psychiatric disorder and hence not a matter of choice of that person, then he or she should not be held criminally responsible if they offended due to their impoverished caring about others. Therefore, I argue that impoverished caring about others, which is an emotional impairment, can remove criminal responsibility from an offender, if it is a result of narcissistic personality disorder.

To examine the ethical implications of the above argument that criminal responsibility can be removed by the emotional impairment of impoverished caring about others found in narcissistic personality disorder, I will make use of the ethical principle of autonomy again. As noted above, autonomy refers to respect for persons and their choices, and criminal responsibility assumes choice of the offender in offending. If so, offending, yet being deemed not criminally responsible, involves deficient autonomy of the offender. This can be shown for the notion of criminal responsibility according to the traditional (cognitivist) construct of the insanity defense, where a (cognitive) inability to learn the difference between good and bad, for instance, restricts fully autonomous choice. But it can also be shown for removal of criminal responsibility associated with impoverished caring about others, where choice is restricted, as argued above.

An ethical implication of such restriction of autonomy associated with removal of criminal responsibility is that an offender deemed not criminally responsible would not usually be sentenced to jail but rather hospitalized or treated alternatively. Thus, an (admittedly controversial) ethical implication of the emotional impairment of impoverished caring about others found in narcissistic personality disorder is that if such an emotional impairment is severe or disruptive enough to lead to removal of criminal responsibility, as it could be, the person afflicted with it would not usually be sentenced to jail but rather hospitalized or treated...
alternatively (unfortunately, narcissistic personality disorder – and particularly its more severe form, termed malignant narcissism – does not respond well to treatment; this could be said by some to justify indefinite involuntary hospitalization of such offenders, as is now being considered in the United Kingdom for psychopaths, who are severely personality disordered and have much in common with severely narcissistic people). This demonstrates that moral evaluation in psychiatry can include emotion criteria, at least regarding ethical considerations related to criminal responsibility and autonomy.

A caveat is that the determination whether choice is (mentally) restricted is perhaps more difficult in relation to emotional impairment such as impoverished caring about others, as compared to cognitive impairment. If a narcissistic person commits a criminal offence, and it is related to his or her impoverished caring about others, it is still unclear whether he or she had no choice and how severe the narcissism or impoverished caring about others should be so that choice is considered sufficiently restricted and criminal responsibility is removed, partly or in full. Determining a threshold – one or more (the latter if the notion of partial criminal responsibility is accepted) – for removing criminal responsibility may be helpful here, but it may not be possible to reach consensus on that, considering that there is no consensus on what amount of narcissism is healthy and adaptive and in which circumstances (McWilliams, 1994). Still, this may be primarily a clinical problem rather than an ethical problem, and thus may not detract from the conclusion reached above.

5. Ethical implications of emotional impairment in schizophrenia

Schizophrenia is defined by so-called positive and negative symptoms that lead to disruption of normative functioning. Schizophrenia is also commonly accompanied by comorbid symptoms such as anxiety and depression, as well as by cognitive impairments, particularly attention, memory and executive function impairments. Positive symptoms consist of psychosis (delusions or hallucinations) or disorganized thought or behavior, and negative symptoms consist of reduction in verbalization, or ideation, in motivation and in emotional expression (American Psychiatric Association, 2000).

Most pertinent to our analysis, it is not only emotional expression but also emotional experience, i.e., feeling, that is commonly reduced in schizophrenia, albeit perhaps less severely or more selectively than emotional expression (Tremblay, 2006). Such impoverished emotional experience involves positive valence or pleasant emotions more than negative valence or unpleasant emotions (Horan et al., 2006). This is manifest in tests such as emotional intensity differentiation tasks
Ethical implications of emotional impairment (Gur et al., 2006). For instance, on average, a person with schizophrenia would feel happiness less frequently and less strongly than sadness, as compared to the general population. This reduced emotional experience explains part of the prevalent complications of schizophrenia, such as lack of motivation of afflicted people to plan and act on their own behalf, and their high rate of suicidality.

It is helpful for our purpose to examine the implications of this emotional impairment, i.e., reduced feeling or impoverished emotional experience, for the controversial matter of coercion in psychiatry, specifically coercion in psychiatric rehabilitation. Psychiatric rehabilitation is the theory and practice of assisting people with severe mental illnesses achieve and maintain goals and roles. The leading approach in psychiatric rehabilitation is client-centered, which means that the goals and roles are those chosen by the severely mentally ill person (Anthony et al., 2002). Coercion can be characterized as the involuntary restriction or modification by one party of another party’s range of choice or actual choice, through various means, such as physical force, emotional extortion, and intellectual manipulation. Although therapeutic coercion is coercion with the goal of helping the coerced party, coercion cannot be part of client-centered psychiatric rehabilitation, as it would consist in someone other than the mentally ill person setting goals and roles for the latter.

In a previous publication, I argued that there are (extraordinary) circumstances where achieving and maintaining certain goals and roles set for the mentally ill person by someone else may be needed (Rudnick, 2002b). In brief, the argument is that there are circumstances where the goals and roles desired by the mentally ill person are induced by his or her mental illness and are considerably detrimental to others or to the person, sometimes even detrimental to his or her mere survival. For instance, the goal and role of being homeless or unsheltered in the freezing winter due to the fear that having an address will lead (imaginary) persecutors to one’s location, is induced by the mental illness and detrimental to that person, including to his or her mere survival. In such cases, setting goals and roles for the mentally ill person by someone else, such as staying in the hospital during the freezing winter, may be needed.

More pertinently, as noted above, lack of motivation is induced by schizophrenia, and leads to difficulty in setting and working towards goals and roles, sometimes to the point of neglecting basic personal hygiene and health to such an extent that physical morbidity and risk of mortality are increased. In such cases, setting goals and roles for the mentally ill person by someone else, such as maintaining good general health and accessing health care for that, may be needed. If so, the client-centered approach cannot fully govern psychiatric rehabilitation.

Thus, I claim that if a person has impoverished emotional experience due to a psychiatric disorder such as schizophrenia, coerced psychiatric rehabilitation may
be needed, albeit only in certain (extraordinary) circumstances. For example, it may be needed where there is considerable risk – such as risk of mortality – to the mentally ill person due to complications of impoverished emotional experience such as lack of motivation.

To examine the ethical implications of the above argument, I will make use of the ethical principle of beneficence. The notion of (therapeutic) coercion is associated with the notion of beneficence, defined as maximizing benefit (and minimizing harm) to a person. As a matter of fact, all paternalistic – hence, coercive – codes of conduct in health care, such as the Hippocratic oath, make beneficence (and non-maleficence) in relation to the patient their main consideration (Lloyd, 1983). An ethical implication of such considerations of beneficence (and non-maleficence) associated with (therapeutic) coercion is that a person deemed unable to plan and act on his or her own goals and roles may be coerced into goals and roles set for him or her by someone else, preferably a formally appointed substitute decision maker who considers that person's best interests (rather than a clinician who may have standard goals and roles in mind). Thus the person afflicted with the emotional impairment of impoverished emotional experience found in schizophrenia may be coerced into psychiatric rehabilitation for his or her best interests. This demonstrates that moral evaluation in psychiatry can include emotion criteria, at least regarding ethical considerations related to coercion and beneficence.

A caveat is that coercion in psychiatric rehabilitation may be more difficult, and perhaps not feasible or even impossible, in psychiatric rehabilitation, at least as compared to psychiatric treatment, where an injection can be given by force relatively easily. If a person declines to plan and act on goals or roles due to negative symptoms such as lack of motivation, even if someone else sets roles and goals for that person, psychiatric rehabilitation may not succeed. Success in rehabilitation requires that the mentally ill person become more motivated and attempt to achieve and maintain roles and goals. The clinical process of psychiatric rehabilitation readiness assessment and development may be helpful in motivating the mentally ill person (Anthony et al., 2002), but not every mentally ill person is willing to engage in it nor is it always successful. Still, this may be primarily a clinical problem rather than an ethical problem, and thus may not detract from the conclusion reached above.

6. Conclusion

First, I want to briefly recapitulate the conclusions of the above sections. In the section on major depression, I reached the conclusion that the emotional impairment of impoverished caring about self can cause incompetence to consent to
Ethical implications of emotional impairment

Treatment. An ethical implication of that is that if this emotional impairment is disruptive enough to lead to incompetence, then that impairment is associated with deficient autonomy. The person afflicted with such an emotional impairment may not be allowed to decide on his or her own treatment.

In the section on narcissistic personality disorder, I reached the conclusion that the emotional impairment of impoverished caring about others can remove criminal responsibility from an offender. An ethical implication of that is that if this emotional impairment is disruptive enough to lead to removal of criminal responsibility, then that impairment is associated with deficient autonomy. The person afflicted with such an emotional impairment may not be sentenced to jail but rather hospitalized or treated alternatively.

In the section on schizophrenia, I reached the conclusion that the emotional impairment of impoverished emotional experience (and hence reduced motivation) can justify (therapeutic) coercion in psychiatric rehabilitation. An ethical implication of that is that if this emotional impairment is disruptive enough to lead to (significant) lack of motivation, then that impairment requires the consideration of beneficence as a primary consideration. The person afflicted with such an emotional impairment may be coerced into psychiatric rehabilitation according to their best interests.

These conclusions, each addressing a distinct psychiatric disorder and a distinct moral evaluation practice in psychiatry, support the more general conclusion that moral evaluation in psychiatry can and does include emotion criteria. That is, moral evaluation in psychiatry can be understood to address emotion, as well as cognition. If so, what could be the consequences of the historically based cognitivist approach to moral evaluation in psychiatry? Two important consequences come to mind, one for clinical practice in psychiatry, the other for moral evaluation in psychiatry.

First, as moral evaluation is inherent in psychiatric practice, as demonstrated above, a cognitivist approach to moral evaluation in psychiatry risks eliminating or minimizing the role of emotions as constructs in psychiatric practice. That would be detrimental to clinical practice in psychiatry, as people with psychiatric disorders manifest a multitude of emotional impairments as well as normal emotions in response to stress.

Second, as moral evaluation in psychiatry addresses criteria involving emotions, as demonstrated above, a cognitivist approach to moral evaluation in psychiatry ignores the role of emotions in moral evaluation in psychiatry, possibly resulting in faulty ethical decision making, such as faulty determination of competence or incompetence, faulty removal or maintenance of criminal responsibility, and faulty coercion or abstinence from coercion. That would be detrimental to moral evaluation in psychiatry, which is based on sound ethical decision-making.
Thus, moral evaluation in psychiatry can and should address emotions and their impairments. This is an area for further research. Such research could address ethical implications of emotional impairments not addressed in this chapter. One such emotional impairment is explosive anger, which is part of various psychiatric disorders, such as narcissistic personality disorder where it manifests as temper tantrums, as noted above; explosive anger is also part of less well known psychiatric disorders, such as intermittent explosive disorder, where it is the central criterion for diagnosis of the disorder. Ethical implications of explosive anger and of other emotional impairments are well worth further study.

References


Facts and values in emotional plasticity

Luc Faucher and Christine Tappolet

1. Introduction

How much can we shape the emotions we experience? This is the question of the plasticity\(^1\) of our emotions. Actually, it is rather misleading to speak of the plasticity question, for the question of how much we can shape our emotions hides a multitude of more particular questions. Depending on how the “we” is defined, the plasticity question splits into several sub-questions. For instance, to what extent is the agent herself able to control, cultivate and modify her emotions? This might well be the most important question from a personal and ethical point of view, but it is not the only one. For one might also ask to what extent surrounding people – parents, teachers, friends, etc. – are able to change the emotions of another person. More generally, to what extent does the socio-cultural environment determine what we feel? Correlatively, what part do our genes, or more generally, the elements of our natural environment that are not socio-cultural, play? Given that emotions involve several aspects or components – the emotional experience, the appraisal, the behavioral disposition, etc. – the plasticity question splits even further. To have a full account of the plasticity of emotions, the plasticity of each component has to be considered.

A further complexity comes from the diversity of the many things that fall or can be considered to fall into the class of the emotions. Is the plasticity of so-called “basic emotions”, such as joy, fear or disgust, different from the plasticity of so-called “higher cognitive emotions”, such as guilt, envy, jealousy? And what about emotions that are considered to be “socially constructed”, like the Japanese emotion *amae*, which is characterized as a pleasurable feeling of dependency (Doi, 1973; Morsbach & Tyler 1986)?\(^2\) From a methodological point of view, it might be

---

1. The term “plasticity” comes from the Greek *plastikos*, which meant *relative to sculpture* and has the same root as *plasma*, which originally meant “thing shaped or sculptured”.

2. The question whether there are basic emotions and also what state would count as such is controversial.
thought that the best policy would be to first consider the question of the plasticity of each putative kind of emotion on its own terms, keeping in mind that our ordinary emotion-categories could well be misleading. Here, however, we plan to discuss more general considerations for and against the plasticity of emotions.

The question of the plasticity of emotions is not only of purely theoretical interest. It is also highly relevant for ethics or more generally for value theory. Whether one is a virtue ethicist or not, one has to acknowledge that moral theory has to make room for character traits, be they virtues or vices. And it is generally acknowledged that Aristotle was right to claim that virtues and vices are concerned with both actions and passions (EN, 1106b). Virtues and vices can be defined as involving character traits, i.e. standing dispositions to act in characteristic ways in given contexts, but also to tone certain percepts, certain thoughts and to feel certain emotions. Thus, the question as to what extent our emotional dispositions depend on us has to be considered if one wants to know to what extent an agent can improve himself. And the same is true of the question whether or not moral education is possible and how one should go about raising children in order to make admirable agents of them. As Aristotle stressed, if virtues involve dispositions to feel emotions, then moral education is as least partly a matter of educating our emotions (EN, 1106b21–23). Moral education would thus presuppose some amount of emotional plasticity.

The question of the plasticity of emotions also has a bearing on how we evaluate agents. For instance, if it turned out that the character traits, or more particularly the emotional traits we have are completely determined by our genes, we would in fact be the victims of chance or “moral luck”, to use Bernard Williams’ expression (1976). We could just hope for a nice genetic makeup or else try to develop some technology to change our genes (or to modify the genes’ expression). In any case, if our emotional dispositions do not depend on us, it is not clear that we would, or should, go on praising the virtuous and blaming the vicious. Indeed, the very responsibility of agents is at stake. Responsibility seems to come with at least minimal control on the kind of persons we are. It seems that we have to agree with Susan Wolf when she claims that moral responsibility for our actions requires moral responsibility for the persons we are. According to Wolf (1987, p.

3. Sophisticated consequentialists such as Peter Railton (1984) and Philip Pettit (1994) claim that virtuous dispositions might be more conducive to the promotion of value than consciously aiming at the promotion of value; and even Kantians now admit that character traits are important (Baron, 1997).

4. See Flanagan 1991, p. 277. The claim that there are such dispositions has been put into question by some, such as Harman (1999) and Doris (1998, 2002). We believe that Flanagan (1991) and Sreenivasan (2001) are right to say that situationist experiments only show that character traits are partly context-sensitive, not that there are no character traits.
60), being responsible for who we are requires two things: a) the ability to evaluate our characters accurately and sensibly, and b) the capacity to transform ourselves in order to correspond to our self-evaluations.

Another idea which is often associated with an Aristotelian account of ethics is that emotions have a role to play in moral discernment. Emotional experiences would key us, to use Karen Jones’ phrase (2003), to certain morally relevant considerations – that an action would be a shameful thing to do, or that it would be cruel, for instance – in a way higher-order cognitive faculties could not. Put in a nutshell, the idea is that emotions are perceptions of values. The question of the plasticity of emotions arises when we wonder whether or not we can refine our grasp of moral reality. But even for those who are uncomfortable with the idea of value perception, it is clear that some moral or more generally evaluative concepts are intimately related to specific emotional responses – shameful and shame, disgusting and disgust, admirable and admiration, to name but a few. To understand these concepts, we need to understand emotions. For instance, the question whether these concepts are relative to a socio-cultural group is closely tied to the question whether the corresponding responses are socio-culturally determined or not.

Finally, and in a way that is quite independent of any theoretical approach in ethics, the plasticity of emotion question bears on the question whether the many rules we have with respect to emotions make sense. We are not supposed to feel hate, envy or jealousy, we are not supposed to express too much pride, we are not supposed to act out of anger, we are not supposed to fear innocuous things, such as spiders or mice; but we are supposed to feel happy, to experience compassion or sympathy (but not pity, an emotion taken to involve feelings of superiority) with those who suffer, etc. These rules are of course cultural: they can be quite different depending on the time and place. And they are often addressed to specific portions of a population: men, women, children, boys and girls, not to speak of socio-economic or ethnic subgroups. The relevance of the plasticity question should be quite obvious. To put it bluntly, if emotions were completely determined, no such rule would make sense. This at least is true if one takes seriously the idea that ought implies can, in the sense that only what is humanly feasible can be morally required from us or, more generally, required from us from a practical point of view.

7. This is often taken to be a conceptual truth, but it might also be taken to be grounded in the thought that requiring what is not feasible from an agent is both cruel and useless. In so far as a non-feasible ideal guides our requirement, we might also ask agents to approximate this ideal. As Flanagan notes (1991, p. 340, fn. 1), what is required might simply be almost feasible, an ideal we can try to approximate.
Thus, our normative practices appear to presuppose at least a certain amount of emotional plasticity.

Most would reject both the radically deterministic conception according to which the emotions we feel are entirely determined by nature and the fundamentally opposed view that claims that our emotions are entirely up to us, whether this is understood in existentialist terms – the individual shapes her own emotions – or in social-constructionist terms – the socio-cultural environment is thought to be the shaping force. As most would agree, emotions are partly under our control, partly not; partly shaped through education and culture, partly not; partly specified by our genetic endowment and our natural environment, partly not. However, though they are true, such statements are far from helpful for those who wish to undertake the moral education of actual human beings.

Our aim here is a very modest one. It is an attempt to draw a framework for future discussions about the plasticity of our emotions. We will consider empirical work only in order to indicate some of the places one might look for data concerning plasticity. We shall start our discussion with a conceptual exercise aimed at different possible models of emotional plasticity. As it is obvious to us that such an exercise is futile without specifying the nature of emotion, we will first propose an analysis of the components of emotions. After these theoretical considerations, we will present some empirical data that might be used to assess the degree of plasticity of emotions as well as the constraints that bear on emotions. In order to do this, we will mainly draw from developmental psychology. In considering the empirical data, it is by no means our intention to provide the last word on the question of plasticity, but we would be happy if we could convince our reader that there are some facts out there that could inform moral practice.

2. Conceptual clarifications

2.1 The components of emotions

In order to address the problem of emotional plasticity, it is useful to have a better idea of what emotions are. Let us start with the common distinction between *occurent emotions*, such as the fear or the anger that a person experiences on a particular occasion, and *emotional dispositions*, such as the disposition to feel fear in certain circumstances. Emotional dispositions can be more or less ingrained. A disposition like irascibility can be short lived – it can disappear as soon as you have

---

8. See Lyons (1980, p. 53-57). As discussed in Faucher & Tappolet (2002), this distinction is actually too crude for there are reasons to distinguish between what we call long-term occurent emotions and short-lived occurent emotions (cf. Goldie 2000).
had your morning coffee, for instance – but it can also be more like a trait. There seems to be a continuum here between passing emotional dispositions and something like a permanent irritability or irascibility. In the latter case, one can talk of *emotional traits*, a term that most naturally refers to the character-defining emotional disposition(s) of a person. We should distinguish between two kinds of emotional traits: temperament and personality trait. A *temperament* involves a set of emotional dispositions that displays variation among individuals and that have a physiological basis thought to be derived from the individual genetic make-up. Here is how Bates (2000) defines it: a temperament is “(...) a set of biologically rooted, early-appearing, and relatively stable individual differences in reactivity to stimuli and self-regulation of that reactivity (...)” (p. 384). Temperaments involve things like the tendency to be fearful in novel situations (what Kagan calls “inhibited”), to be easily distressed, to get angry easily or to be strongly attracted to potential mastery experiences (what Kagan calls ‘uninhibited’). Temperament is thought to be an important part of personality (people do have basic emotional reaction styles), but the second cannot be reduced to the first. However, as Rutter (1994) puts it, “everyday functioning is also influenced by self-concepts, by social cognitions, and by values and attitudes. Humans are thinking beings, and as such their personality will be shaped by the way in which they deal with the temperamental qualities with which they have been endowed, and by the view of themselves and their social worlds that they acquire. Thus, we characterize ourselves and other people in terms of self-esteem, self-confidence, and self-efficacy, as well as by attributes such as suspiciousness, conscientiousness, empathy, and trustworthiness. It seems reasonable to suppose these qualities will be shaped, in part, by temperamental qualities (...).” (p. 28)

Given the distinction between occurrent emotions and emotional dispositions, be they passing or longer-lived dispositions, the plasticity question falls into two sub-questions: a) the plasticity of occurrent emotional states, and b) the plasticity of emotional dispositions.

The first question bears on the direct influence we can have on our occurrent emotions. It is clear that there is some plasticity here. We might control or regulate the expression of an emotion when experiencing it, such as when we make a poker-face when losing an argument. We might get ourselves to experience an

---

9. Kagan (2003) gives a similar definition: “Temperament (...) is a stable psychological profile, characteristic of only a proportion of the population, that has a biological foundation, that emerges during childhood, that is associated with particular affective states and that is not a pathological category” (p. 321).

10. In his discussion, Ben-Ze’ev (2000) mainly concentrates on the impact we can have on emotional experiences.
emotion, for instance by thinking about our worst enemy. Or we might attempt to get rid of emotion, such as when we try to calm down by going for a walk. However, most of the long-term and deep impact we can hope for depends on the plasticity of our emotional dispositions. Thus, it is important to know whether or not we can change our dispositions so our emotional reactions would be triggered by different kinds of things, whether or not we can suppress (some or all) emotional dispositions altogether, whether or not we can extend the range of emotions we are able to experience by acquiring new emotional dispositions. If we could achieve such changes, we might become quite different persons.

Obviously, changes in our emotional dispositions will have an impact on the occurrent emotions we experience. If we get rid of a spider phobia, we will not experience fear when confronted with a small innocuous spider. By contrast, it seems that the impact we can have on our occurrent emotions will make no difference to our emotional dispositions. Someone who learns to control his fits of fear when seeing spiders such as to behave almost normally has not yet been able to get rid of the phobia itself. This is why in general the question of the plasticity of emotional dispositions is seen as the most important one. In fact, as we will see later, things are a bit more complicated. It is possible that one way to get rid of a phobia is to “take one day at a time” and get control over our fear on every occasion, until we are not afraid anymore. In the case of temperament, it is possible that the occurrent emotions we express are disposing people to react a certain way to us, hence making us more likely to express the emotion again and fortifying the disposition we have.

In any case, to understand the nature of emotional dispositions, we have to begin with emotions. Emotions have often been taken to involve different components, some of which, but not all, were thought essential. The debate about the emotion component is of importance for the purpose of our discussion, because different components of a same emotion could be plastic to a different degree. Here is the list of what can be considered the different components of emotions:

1. The cognitive or informational component: If you experience fear of a dog, you need to see (or hear or smell) the dog, or to have specific beliefs about the dog – that the dog is around or running towards you, for instance.\(^1\) Emotions have a cognitive or at least an informational basis: an object needs to be grasped in one way or another for it to become the intentional object of the emotion. It is worth noting that this cognitive or informational state need not involve concepts.

\(^{11}\) It might in fact be sufficient to imagine certain things (that the dog is around, etc.).
2. **Appraisal or evaluative component**: According to some, emotions involve an evaluative or normative judgment. However, if judgments are taken to require conceptual skills, this claim appears to be overstated. As cases of so-called irrational emotions show, the evaluative component is better not conceived of as a judgment in the standard sense of the term. However, it is plausible to claim that emotions involve an evaluative appraisal or an evaluative perception. Fearing something involves assessing that thing negatively, as something that threatens you.

3. **Expression and bodily changes**: Emotions are usually taken to come with characteristic bodily changes. The changes involve a variety of systems controlled by the autonomic nervous system, such as the respiratory, the circulatory, the digestive, the musculoskeletal, the endocrine or the immune system. Thus, an emotion like fear is characterized, among other things, by an increase in the heart rate. Some of these changes, such as when you get red in the face out of anger, are visible to an external observer. They are part of the facial or bodily expressions of emotions. Other aspects of the emotional expressions are not directly related to the activation of the autonomic nervous system.

4. **Emotional Experience**: When you are in a state of fear, you typically undergo an experience which is characterized by a certain feel. Some consider emotional experiences to be related to bodily changes. James, for instance, considered that emotions are perceptions of bodily changes (James, 1884). Damasio (1994, 1999, 2004) recently revived the Jamesian tradition, but added the claim that actual bodily changes are not necessary to experience the emotion: imagining an emotional experience, for instance, activates somatosensory regions and gives rise to emotional experience as well (see also Damasio et al., 2000). Others, like Frijda

---

12. We are conscious that the literature on emotions usually does not distinguish 1 and 2, but we think that there is reason to do so. Perceiving a dog running towards you is not yet to appraise it as a threat.

13. As noted by Griffiths (2003): “Appraisal theorists have come to accept that even (...) apparently conceptually complex dimensions of evaluation as Richard Lazarus’s ‘core relational themes’ can be assessed: 1. Without the information evaluated being available to other cognitive processes, 2. Before perceptual processing of the stimulus has been completed, and 3. Using only simple, sensory concepts to define the property that has to be identified.” (p. 41)


15. Another proposal has been offered by Panksepp (2000, 2003). According to him emotional experiences are the result of the activation of phylogenetically ancient and individually specified subcortical systems (like the rage/anger system or the play/joy system). This is true of what Panksepp calls “blue-ribbon emotions”, that is a group of seven basic emotions. According to him, emotions like guilt and shame “(...) do not really exist as fundamental processes, even in the human brain, but they can easily arise as derivative processes based on social learning that
Luc Faucher and Christine Tappolet (2005), think that it would be a mistake to reduce the emotional experience to the perception of bodily feelings or to a kind of unanalyzable quale. According to Frijda, the structure of experience “contains conscious reflections of the four major nonconscious components of the process of emotions: affect, appraisal, action readiness, and arousal. In addition, it may include the emotion’s felt ‘significance’” (2005, p. 494). Frijda argues further that how emotions will be experienced depends upon where one’s attention is directed. For instance, attention can be turned towards the world or towards oneself. If the attention is turned towards the world, you might experience the situation as depressing or the object as disgusting. If attention is turned towards yourself, you might experience the action tendency characteristic of the emotion or your experience might be more conceptual, such as when you feel that you are a failure. Experience can also be modified by cognition (Frijda calls this sort of experience, “second order experience” and contrasts it with the “core phenomenology” that arises from the amalgam of the unconscious component of emotions plus the direction of attention). He claims that cultural differences in the way situations are appraised and expectations about how one should react to them translate into differences in phenomenology of emotions. While the question of who is right between the advocates of the perception of bodily changes or of a more elaborate view of the emotional experience is far from settled, almost everyone would agree that emotional experience is an essential constituent of emotion.

5. **Action tendencies**: Fear typically comes with either an urge to run away or a tendency to freeze. More generally, emotions are often thought to involve dispositions to act in certain specific ways. Some, like Frijda (1986) even claim emotions are essentially action-tendencies, though his conception of what counts as an action-tendency is very liberal. Others, like Prinz (2004), take the relation to action to be looser. Emotions do not prompt specific behavior, but the bodily changes they involve facilitate action, and the negative and positive valence of emotions dispose us to act in certain general ways, choice being necessary for specific action to ensue.

Weaves such basic feelings as separation distress and social bonding into more complex socio-cultural realities” (2000, p. 138). Noting that decorticated animals still exhibit emotions and that emotional feelings can be triggered by direct electrical stimulation of the subcortical brain systems, Panksepp suggests that emotional experience is independent of cognition: it is a primitive conscious experience, a direct readout from an emotional neural circuit.

16. This position, like Russell’s that we present later, is inspired by the work of Lambie and Marcel (2002).


18. In fact, the behavioral tendencies of fear are more varied. See Tappolet forthcoming.
In addition to these five components, we would like to distinguish two other aspects of emotions. These aspects are not usually taken to be constituents of emotions, but given their importance to the plasticity question it is useful to consider them.

6. Knowledge about emotions: Knowledge about emotions comes in many forms.
   a. Conceptual knowledge about emotions: This includes what types of emotions there are, what the prototype of a particular emotion is, what situation typically elicits this emotion, what kind of behavior typically follows the emotion, etc.
   b. Social norms about emotion: This includes norms concerning when it is appropriate to experience an emotion, when it is appropriate to express (or suppress) the expression of an emotion (i.e. it can be appropriate to feel fear, but not to express it), who can express an emotion, etc.
   c. Personal knowledge about one’s own emotions: With time, it is likely that you come to acquire knowledge about your own emotional dispositions. This knowledge can play a role in keeping you away from certain situations (not drinking with attractive men or women when you are away from home if you want to stay faithful to your partner), or in making you express some emotions more or less often.
   d. Knowledge about others’ emotions: It is crucial for social interaction to be able to perceive and identify others’ emotions. Some of this knowledge (but not all, by any means) might be brought by empathy. Recent work (Decety & Jackson 2006) shows that, at least in some cases, empathy works by using most of the emotional neural structures that are usually active when actually experiencing the target emotion.

7. Coping or regulating skills: It is something to know the social norms about emotions, but it is another to be able to comply with them by either not expressing an emotion or by “influencing which emotions we have, when we have them, and how we experience and express them” (Gross, 2002, p. 282). People might be thought to differ in their ability to cope with or regulate their emotions. Coping or regulating skills are the skills required to comply with social norms, but also with social situations where norms are not spelled out. According to one model we favor on this question, there are at least two ways to cope with an emotion (Gross 2002; Ochsner & Gross 2005): a cognitive and a behavioral one. The first way is to modify your appraisal of the situation (for instance, seeing the fact that you are not getting a job not as a defeat, but as a learning experience) so you don’t experience the emotion that you would normally experience (cognitive regulation). The second way is to decrease or hide the expression of emotion (behavioral regulation). Gross has found that decreasing or hiding an emotion is cognitively more demanding and has an impact on performance on certain cognitive tasks (for instance, it makes
you less responsive to others, maybe because you keep thinking about hiding or expressing your emotions). People might be thought to differ in their degree of control or regulating skills, or by their tendency to use one or the other strategy to regulate their emotions.

Components (6) and (7) can be seen as elements of what is now known as “emotional intelligence” (Pizarro & Salovey, 2002; Salovey et al. 2000; Salovey & Grewal 2005)¹⁹. Researchers have created this concept to make some order in the growing literature on emotions and their relations to cognition and behavior. According to the definition of some of the leading psychologists working on the issue, emotional intelligence comprises four basic abilities (Salovey et al., 2000):

A. the ability to perceive (in us or in others), to appraise (in us or in others) and to express emotion

B. the ability to use feelings in cognitive activities

C. the ability to understand emotion and emotional knowledge

D. the ability to regulate or manage emotions to promote emotional and intellectual growth.

It is clear that the plasticity question also applies to “emotional intelligence”. Indeed, a large amount of the literature on moral development concerns aspects of emotional intelligence, like, for instance, the development of empathy (Eisenberg, 2000; Hoffman, 2000), or the development of the knowledge of the emotional effects of wrongdoing (Arsenio & Kramer 1992). Researchers have already started to look at ways to produce or “increase” emotional intelligence (for a review, see Pizarro & Salovey, 2002)²⁰. But since we are interested mainly in the question of the plasticity of emotion per se, we will leave the exploration of the topic of emotional intelligence for another time.

2.2 Models of plasticity

It is possible to imagine different scenarios about what we come equipped with at birth. For instance, either we come equipped with a fully constituted emotional system, or with a program specifying how to build the emotional system, or the emotional system is the result of contingently putting together different constitu-

---

¹⁹. As Peter Zachar reminded us, these features of emotional knowledge are called “psychological mindedness” in clinical psychology. It is thought to be what people with Autistic Disorder and Asperger’s Disorder lack, and what many successful psychotherapy clients possess.

²⁰. “(...) In any future moral situations, children who were disciplined through the use of inductive methods will continue to act morally even in the absence of external authority or threat, as compared with children who were disciplined merely through an exertion of parental authority.” (Pizarro & Salovey 2002, p. 258)
ents depending upon the physical and/or socio-cultural environment one is in. Here is a list of seven different ways of conceiving emotional systems, going from the degree zero of plasticity to full plasticity. It is easy to imagine many more models, but as we hope will be made clear these seven models cover the main theoretical options. Though it is possible that none of these models apply as such, we believe that they have heuristic value in that they allow one to consider the range of possibilities, as well as to specify the questions that need to be considered.

*The fully equipped model (FEM):* In this model, you come equipped at birth with a fully functional emotional system that is ready to use as soon as you come into the world or have to wait for other non-emotional components to develop. This is the degree zero of plasticity. Nothing changes.

*The marble model (MM):* In this second model, inspired by Leibniz’s view on innateness, we basically come equipped at birth with a system of well-defined emotional dispositions that get shaped, following existing fault lines, by the environment. For instance, we would be disposed to acquire a fear of spiders, be disgusted by meat products, feel anger at cheaters, etc. But these dispositions become actual dispositions of fearing spiders or being disgusted by meat products only if we are exposed to a particular environment.

*The avocado pear model (APM):* According to Peter Goldie (2000, p. 99), emotion theorists such as Ekman (1980, p. 89) and Solomon (1994, p. 267) make the error of embracing what he calls “the avocado pear conception” of our emotional capabilities. This model is a version of the marble model, except that it adds a twist to it. We come equipped with innate emotional systems (this is the hard core), but these systems could be superficially modulated (this is the soft skin). For instance, while some situations activate the emotional expression system, cultural norms or other reasons might make you hide your facial expression. The same idea is captured by Ekman when he claims that emotional kinds, or families, have both themes and variations, for he claims that “the themes may be largely the product of our evolution and given genetically, while the variations reflect learning, both species constant and species variable learning experiences.” (1999, p. 173)

Goldie claims that the avocado pear model should be replaced by a picture of a “single developed capability which has itself been shaped by the culture and environment in which the individual is placed.” (2000, p. 101) Biology and culture interact to shape our emotional systems, but it is not a coherent project to try to disentangle what’s ‘hard’ or neurological and what’s ‘soft’ or cultural. Thus, smiling is “developmentally open or plastic”, but it cannot be decomposed into an “invol-
untary emotional response and a voluntary control thereof” (ibid.).\textsuperscript{21} At least on what we take to be a sensible interpretation of them, this is what happens according to the last three models we propose.

The clay model (ClM): Clay is obviously different from marble in that it is malleable at first, but then, after drying it loses its plasticity. In such a model, the emotional system is constructed (maybe psychologically constructed) in the first few years of life, but it would lose its plasticity after this period. This is analogous to the concept of critical period in animal learning. Peter Gould explains how some birds acquire the songs characteristic of their group (Gould & Marler, 1987). According to him, at a certain critical period of development, a window opens when birds will register the songs of the group of birds they are with. A few days later, using a model of what they heard, they start trying to imitate their fellows. In certain species of birds once this song is learned it crystallizes and stays the same for the rest of their life (other species of birds have to learn songs every season. These species are an example of the next model).

The wax model (WM): According to this model, the emotional system would not only be malleable at the beginning of our life. Just as wax tends to lose its plasticity when it cools down but can be reheated and reshaped forever, the plasticity of the emotional system would last over the years. Thus, though the reshaping of emotions would require some energy, the emotional system would be open to changes even at later stages of one’s life.

The silly-putty model (SPM): Silly-putty is a toy that was a craze among youth in the sixties. Basically, silly-putty can take on any forms you want and keep its malleability for a long time. Not only can silly-putty be used to sculpt your favorite form or used as a ball (it can be bounced as well), it can also, when applied to a printed image, transfer it so you would have a copy of the image. In such a model, the emotional system is considered plastic not only in that it can take different forms, be directly influenced by culture, or assume different functions, but because it also retains its plasticity for the whole life of the subject. This is full plasticity.

These models need not apply only to emotional systems as such. They can apply also to particular emotional dispositions or to emotional traits. Shyness, for instance, could be the result of a disposition that we are born with\textsuperscript{22} and on which the

\textsuperscript{21} Prinz makes a similar, but weaker, claim to the effect that we cannot easily tease apart the contributions of nature and nurture; however, he allows that the disentangling can be done if one looks at a particular emotion’s history (2004, p. 158).

\textsuperscript{22} For instance, as suggested by Thomas and Chess in their landmark study on temperament (1977), maybe those who grow up to be shy are the same who were identified as “slow to warm up” as children.
environment, be it physical or social, has little impact, so that it would be best described by the FEM or the MM. Actually, it might well turn out that different models apply to different kinds of emotional dispositions, or even to different kinds of occurrent emotions. Friends of the distinction between basic and higher-cognitive emotions, for instance, might claim that a model like the MM or the APM applies to our dispositions to experience basic emotions, whereas the social nature of higher-cognitive emotions would make one of the last three models more suitable. Moreover, the models can apply to the components of emotional dispositions. For instance, one might think that the disposition to make given appraisals and the disposition to undergo certain bodily changes and certain conscious experiences, given those appraisals, come together at birth, while the disposition to acquire certain information about our environment has parameters that have to be set by the environment, and the resulting action tendencies are a mix of innate dispositions and social learning.

Now that we have provided a typology of the alternative models of plasticity as well as having offered some precision as to what the different questions of plasticity amount to, the next step is to consider some empirical data. As one might expect, there are many domains of inquiry that produce data that are relevant to questions about the plasticity of emotions. If we were to embark on an exhaustive study we would consider domains like evolutionary theory, animal cognitive ethology, social psychology, neuroeconomy, clinical psychology, anthropology, among others. Such an exhaustive study is desirable and necessary to answer the plasticity question. But for the sake of space, we will concentrate our attention on only one of these domains: developmental psychology.

3. The development of emotions

In this section, we explore the question of plasticity by looking at the development of emotional capacities and dispositions in individuals. As it will become clear, no definitive consensus has been reached as to which model of development is the right one. But the evaluation of the models discussed in the literature might help us to get a clearer idea about our options concerning plasticity. Before presenting these models, we will take a look at the sequence of development of emotions, since it will help narrow our investigation somewhat.

3.1 Order of development of emotions

When do the emotions come ‘on-line’ for the individual? Which emotions are there at birth (or at least very early), which emotions appear later? One reason to
ask ourselves these questions is that we can assume that the later an emotion comes on-line the more plastic it is, or at least the more open to external influences it is. The more time something takes to mature, the more chances there are for external factors to play a role in its making (this seems to be true of the cortex (Quartz & Sejnowski 2002); the amygdala is not fully developed before the first year it is possible to think that it is also subject to the influence of culture, see Joseph 1999 for that kind of view).

The standard account of the sequence of the development of emotions is given by Lewis (2000a)\textsuperscript{23}. According to him, what he calls “the primary emotions” (joy, surprise, sadness, disgust, anger and fear) appear in the first six months of life\textsuperscript{24}. Another group of emotions and emotional capacities (embarrassment, envy and empathy) have to wait for self-consciousness to develop until they can appear. Self-consciousness typically develops around the age of two-and-a-half. Finally, at the age of three, when individuals can evaluate their own actions according to social standards and norms, shame, guilt, pride, embarrassment, and hubris appear. Lewis (2000b) calls the last group \textit{self-conscious evaluative} emotions.\textsuperscript{25} In this account, the last two classes of emotions thus have to wait for the development of non-emotional capacities (the capacity to distinguish self from other and the capacity to understand and recognize norms).

3.2 Models of development

Now how to explain this sequence? There are, as far as we can tell, three ways for explaining the sequence of emotion.

\textsuperscript{23} If there is agreement on the sequence, there is disagreement as to the details of the sequence. For instance, some, like Campos et al. (1996) and Barrett (1998), are interested in emotions we could call “proto-shame” or “proto-pride”, that is, the early predecessors of shame and pride proper (for a similar idea, see also Griffiths and Scarantino, ms). For them, the roots of a sense of shame and pride are to be found early in infancy. Others, like Lewis (2000a, b), do not seem to consider that possibility: for them, shame and pride appear as a whole at a particular age.

\textsuperscript{24} It is not clear that disgust belongs to this group of early emotions. People like Rozin think that disgust appears later in development via the co-optation of distaste facial expressions (Rozin and Fallon 1987). The reason given for this late appearance is that disgust is some kind of “revulsion at the prospect of (oral) incorporation of an offensive substance. That substance has contamination properties if it contacts an otherwise edible substance, it renders it inedible” (Rozin & Fallon 1987, p. 23). So while distaste’s elicitors are confined to the sensory world (bitter taste, weird texture), the elicitors of disgust are ‘ideational’ or ‘highly cognitive’ properties.

\textsuperscript{25} Note that even if Lewis is not concerned about emotions in other cultures, it is possible to accommodate emotions that are not in our repertoire by identifying their cognitive requirements.
1. **The strongly determinist biological view (SDB):** According to this model, which is the most widely known and researched of scientific accounts of emotions, at least one group of emotions (the “primary emotions” or “basic emotions”\(^\text{26}\)) develops according to specification of the genome and according to a predetermined timetable:

   Some features of the emotion system, the overall organization of the discrete emotions as a functional complex, *are primarily a matter of biological development and the unfolding of genetic processes.* (Izard, 1994, p. 361; our emphasis)

   The SDB model also postulates that emotional experiences are distinct and differentiated from the start, so that the experience-expression link is basically the result of phylogeny and is not learned. Overall, SDB suggests a minimal role for the environment in the construction of primary emotions. The explanation of the organization of primary emotion is couched in terms of *biological* maturation, not in terms of *psychological* development (for instance, in terms of learning). Though it usually restricts its claims to emotions that appear early, SDB could also make similar claims for later-developing emotions. For instance, Steinberg (2005) mentions that for romantic love:

   “There is evidence that pubertal development directly influences the development of romantic interest and sexual motivation. There is also evidence that some changes in the frequency and intensity of parent adolescent conflict may be more closely linked to pubertal maturation than age.”

   Thus, in this case too, the explanation of the development of emotions is couched in biological terms, not in psychological ones.

2. **The developmental system view (DS)\(^\text{27}\):** We are borrowing the name of this view from a general theoretical perspective on development, heredity and evolution championed by, among others, Paul Griffiths and Russell Gray (1994). According

---

**Notes:**

26. Izard calls them “independent emotions” by contrast to emotions that need the development of cognitive structure before being able to develop (Dougherty et al, 1996, p. 29).

27. We are conscious that we are grouping together people under a single heading that they might not be familiar with. From our point of view though, they all defend, one way or another, the idea that the diachronical unfolding of emotions is scaffolded by the environment. The same people also sometimes defend the view that the (synchronic) unfolding of a particular emotion is scaffolded by the environment. Griffiths and Scarantino (ms) call the perspective integrating dynamical and synchronical scaffolding of emotions, the “situated perspective” on emotion (another name for that perspective is the “dynamical perspective”, see Lewis 2005). Because we are mainly interested in the developmental aspect of this perspective, we prefer to use the phrase “developmental system model” to group together the studies that challenge the SBD model.
to this perspective, the genes are not the only developmental resources, nor the most important (a thesis called “causal democracy”):

(…) The genes are just one resource that is available to the developmental process. There is a fundamental symmetry between the role of the genes and that of the maternal cytoplasm, or of childhood exposure to language. (Griffiths & Gray 1994, p. 277)

We will not go into a detailed discussion of this perspective and its limits; suffice it to say that the developmental system model underscores two things of interest for us: (A) the source of control of development is not centralized, that is, the adult emotion phenotype is not programmed in the genes, but it depends on the input of many sources including genes, physical factors such as patterns of locomotion, and socio-cultural factors such as parental reactions as well as culturally-dependent concepts of emotion; (B) local constraints channel the development in one direction, a direction not necessarily programmed in the genotype28.

This model of emotional development (advocated by Campos, Kermoian & Witherington 1996; Camras 1994, 2000; Lewis 2005; Russell 2003) typically postulates that we start in life with undifferentiated affective states and that development is a process of differentiation of those states, a process affected by a host of social and cultural factors (cf. de Sousa 1987; Russell 2003). Thus DS assumes a greater plasticity at birth than the SDB view. Advocates of DS also typically assume that there is a greater plasticity to emotions than SDB does (see note 25 and the “Expression” section below).

3. The social constructionist’s view: To be quite frank, constructionists are not big on trying to find an explanation as to why the development of emotions is sequenced the way it is (though see some passing remarks in Averill, 1986). Given that, for them, emotions are “transitory social role[s]” (as Averill puts it), social constructionists are committed to the idea that you cannot experience emotions before you have assimilated some cultural elements, which is pretty late (acculturation can start early, but if having an emotion is a question of mastering certain cultural rules, it seems unlikely that infants can have emotions). Some social constructionists recognize that some of the components of what they call “emotional syndrome” might be innate, but these components still need to be coordinated to

28. “(…) the role of developmental context is not restricted to activating alternative outcomes prefigured in a ‘disjunctive genetic program’ (Griffiths, 1997; Griffiths & Stotz, 2000). Developmental systems are usually competent to produce viable phenotype outside the specific parameter ranges in which they have historically operated. (…) the biological endowment of a healthy human infant determines a norm of reaction which includes a large range of emotional phenotypes, not all of which have been specifically selected for, and not all of which need to have occurred before in human history” (Griffiths & Scarantino, ms).
form an emotion *per se*. Advocates of this view, like Averill et al., (2001), make claims like the following:

> [emotional syndromes] are folk-theoretical constructs, recognized in ordinary language by such abstract names like ‘grief’ in English, *hwa* (‘anger’) in Korean, *amae* (‘dependency’) in Japanese (…) the meaning of emotion qua syndrome depends on a matrix of cultural beliefs (implicit beliefs) about the nature of emotion, (…) emotional syndromes are constituted, in part, by the existential beliefs we hold about them (…) Emotions (…) literally embody the values of society. Stated more formally, emotional syndromes are constituted by social rules as well as by existential beliefs (…) Without the rules of anger, say, there would be no anger, only inarticulate expressions of rage and frustration. (2001, p. 168)

Social constructionists are mainly interested in how emotions are elaborated by beliefs and social rules; they are not very interested in the details of the underlying biology of the components of the emotional syndrome. So basically a social constructionist model of the development of emotions is a model of the development of the capacity of acquisition of beliefs and social rules about emotions, but also of the capacity to recognize the situations that are thought to make emotions appropriate. Their developmental story is about how agents internalize the rules and beliefs about emotions (what they call “emotional schemas”) or how they get more skilled at appraising emotions and using emotions’ categories. Because, from this perspective, the development of emotions rests mainly on the acquisition of non-emotional skills, their developmental story is a non-specific one:

> “From a social constructionist view, emotional development — whether in childhood or as an adult — typically follows a more subtle and non-specific course. Like the development of most other complex forms of behaviour, emotional development tends to be slow, piecemeal and cumulative; indeed, for the most part, emotional development is not even particularly emotional.” (Averill, 1986, p. 112)

Because social constructionists have nothing precise to say about the development of early emotions, we won’t consider them as contenders here. (Another reason for not considering them is that an important part of their claims can be subsumed by the DS view, which also insists on the idea that culture plays an important role in the shaping of the emotions.)

As we see it, the primary battle is between the strongly determinist biological view and the developmental system view. Our strategy in the next sections will be the following. We will first look at some work concerning the construction of components of basic emotions (specifically, the front end and the tail end of emotions,

---

29. As Averill puts it: “From a social constructivist point of view, (…) the most important feature of emotional development involves the acquisition of the social norms and rules that provide the component responses with their meaning and co-ordination.” (1986, p. 105)
that is, the appraisal mechanism and the expression mechanism), but also at the emotional experience. We will leave aside action tendencies for if it can be shown that appraisal mechanisms and expression mechanisms are plastic, then there are strong reasons to believe that action tendencies are also plastic. If both are genetically determined, then there are strong reasons to think that action tendencies are also genetically determined.

Our goal in the next sections is to show that, contrary to the dominant trend in the study of emotions, there is more leeway in the construction of basic (or primary) emotions than acknowledged by advocates of the major theories of basic emotions. If basic emotions are plastic, we can imagine that the same goes for other non-primary emotions. We will then look at work on emotional dispositions. Our goal in this case will be to show that even if emotional dispositions were fixed at birth there are reasons to think that we are not condemned to a particular emotional regimen. Overall, we will argue that there is much more plasticity to our emotions then what SDB typically proposes.

3.3 The plasticity of emotional components

3.3.1 Appraisal mechanisms
Appraisal mechanisms are generally considered to be quite plastic. Most people, from the constructionists to the evolutionary psychologists, agree that what we are afraid of or what we are disgusted by is potentially quite different from culture to culture, from individual to individual, and from one period of the life of an individual to another (see for instance, Rozin, 2003, p. 848). For instance, while banging a spoon on a table may make an infant laugh, it takes much more (we hope!) to make an adult philosopher laugh.

Some of these changes in appraisal are due to changes in cognitive capacities. Hoffman (1982) proposed, for instance, that empathy gets less and less egocentric as children are more capable of differentiating their own internal states from others and their responses become less situated as they get better at imagining “that people can continue to exist over time and contexts” (Eisenberg, 2000, p. 678). Other changes might come from the re-evaluation of goals as one ages (Carstensen & Lockenhoff, 2003).

So, everybody agrees that there are modifications in appraisal, but this agreement masks a deeper disagreement that might have gone unnoticed. For as we have seen in the previous section, biological determinists do not think that the appraisal mechanisms are totally plastic. First, they think that some appraisal mechanisms are set at birth to evaluate a small group of stimuli in a certain way (loud noise are frightful, for instance). Second, the set of things one might be afraid of or disgusted by is seen as somewhat constrained. Certain things could
Facts and values in emotional plasticity

not but be appraised in a certain way (by normal subjects). Miller (1997), for instance, remarks that:

(…) the variation in elicitors of disgust across cultures will hardly look like a random sampling of all things or all actions in the world. Cultures, it seems, have much more leeway in admitting things or actions to the realm of the disgusting than in excluding certain ones from it. Yet even here there are limits. Some things seem almost incapable of eliciting disgust. Animals and animal substances, we can safely assume, will figure more frequently as elicitors of disgust than plants and inanimate objects. And is snow polluting anywhere? Are stones? (…) (p. 16)

Third, not only are there things that are universal elicitors of particular emotions, but there are also mechanisms preparing us to learn to associate certain stimuli with certain emotions. So for instance, we would more easily learn to associate spiders and heights with danger than more recent dangerous stimuli, like guns and electric outlets (Öhman & Mineka, 2001). Not only are those associations easier to establish, but it has been shown that they are also more difficult to undo.

We will leave aside the question of universal elicitors for it would take us away from the literature on development and force us to a fray into the anthropological literature. We will instead concentrate first on learning preparedness.

In his “Preparedness and Phobias” (1995), Davey argues that the phenomena explained by the specific evolved association mechanisms traditionally postulated by psychologists would be better explained by a “human model of classical conditioning”. The human model (described in details in Davey 1992) is different from animal models in that it adds a cognitive dimension to the traditional model of conditioning. According to that model, the determinants of Conditioned Stimulus-UnConditionned Stimulus association (for instance, the association of a spider and a pain sensation or an electric shock) are not only and exclusively dependent on the contiguity of the stimuli (the fact that the subject has actually felt pain or received an electric shock after encountering a spider). It is also dependent on a prior evaluation by the subject of the likelihood of the encounter to be followed by an aversive consequence as well as the evaluation of the degree of pain that such an encounter will cause. Davey calls those determinants “cognitive expectancy biases”.

30. Cosmides and Tooby make a similar claim: “Emotion programs, for example, have a front end that is designed to detect evolutionarily reliable cues that a situation exists”. (2000, p. 93)
31. As shown by Olson et al. (2007), evaluation of the degree of pain that a stimulus will cause can be acquired through verbal means (after all, phobics of microbes fear things that they have never seen, but only been told about) or visual means (like someone else receiving a shock when being presented blue squares).
In other words, your prior fear of an object, as well as your evaluation of the pain that will result from the encounter, are more important factors in establishing the strength of the association between an object and an aversive consequence than what really happens during that encounter. Because the determinants of the expectancy bias are things like estimates about how dangerous something is, “semiotic” similarity between stimuli (angry faces and screaming, for instance) and prior fear about some stimuli, it becomes clear that cultural and ontogenetic factors could play an important role in the explanation of certain characteristics of fear and phobia. For instance, Davey thinks he can explain why fear of spiders is acquired so fast and why it resists extinction by making references to ideas associated with the dangerousness of spiders (and because we fear them we do not approach them, we do not give ourselves the chance to change these beliefs) 32.

Another source of criticism of the SDB explanation comes from the work of Campos and his colleagues (Campos et al., 1996). Campos showed that contrary to what advocates of SDB think, fear of heights is not genetically programmed. According to him, wariness of heights is the result of an experientially constructed process that is linked with the advent of locomotion. It has been shown by putting infants on a glass surface with one shallow side and one deep side that locomotor infants experience a fear of heights while prelocomotor infants do not. If the infant shows signs of cardiac acceleration when put on the deep side, or if he tries to go as quickly as possible to the other side, he is considered to be showing a fear of heights. According to Campos, the origin of this fear is dependent of an expected correlation between visual and vestibular input that is usually characteristic of self-movement. This correspondence is gained by moving voluntarily. Thus, it is only when infants are starting to move by themselves that this correspondence is built and only then that it can be violated. Again, fear seems to depend of an acquired expectation.

3.3.2 Emotional expression
There are a few questions we might ask here about emotional expressions in relation with development (Oster, 2005): (A) are the discrete emotions present at birth or do they develop through a process of differentiation from more global states? (B) Are the expressions characteristic of basic emotions present in their adult-form early in infancy or do they go through transformations? (C) Do facial expressions

32. Isabelle Blanchette (2006) has defended a related idea. Working on visual attention and comparing evolutionary threat relevant stimuli (snakes and spiders) and modern threat stimuli (guns and syringes), she showed that subjects were typically more efficient at locating modern threat than evolutionary-relevant threat. As she noticed, this is in contradiction with predictions from Öhman and her colleagues and more consistent with the DST approach since it implies that the threat detection system is indeed open to ontogenic contingencies.
reliably reflect infants’ affective state or emotional experience at the beginning of life, or does the development allow expression and experience to get disentangled?

One current view on emotional expression development is embedded in Izard’s Differential Emotions Theory (DET) (Izard, 1994; Dougherty et al., 1996; Ackerman et al. 1998). According to DET, emotional expressions, as functional adaptations and the result of phylogeny, are genetically programmed to come in integrated well-organized patterns or modules. Modules do not develop (that is, they are not the result of the process of learning), they mature (leaving some role for a child’s experience):

Emergence of these emotions expressions is primarily a function of the maturation of neural circuits and is independent of cognition. (Ackerman et al. 1998, p. 87)

These patterns or modules are independent of cognitive development or appraisal processes for their activation and they are identical in essential respect to the adult expressions. Moreover, because expressions are, in young infants, directly determined by the emotional experiences it follows that:

The structures of the neural-evaluative component of these emotions must also be functional at birth. (Izard 1994, p. 358; see also Ackerman et al. 1998, p. 9)

According to this view, development allows the individual to have better control over his expression and therefore to sometimes dissociate the experience and the expression or to modulate the expression (blending it, minimizing it, masking it with smiles, etc., (see Keltner 1996, p. 396; for a similar view, see Panksepp and Smith-Pasqualini, 2005). Development consists also in connecting invariant feelings with changing images and thought as well as acquiring better emotional understanding and regulation.

This strong position is opposed by numerous people. For instance, Camras (1994, 2000) proposes that:

(…) the face can assume only a limited number of patterned states due to constraints imposed by lower-order synergistic relationships among muscle actions (i.e. coordinative motor structures). (2000, p. 136)

It is the “coordinative motor structures” – that is, when you move some muscles, you automatically move some others because of the ways in which they are linked – which are responsible for the impression of a well organized system of emotional expression. As Camras shows, these expressions appear also in non-emotional settings. For instance, an infant might show a typical expression of surprise when she raises her head towards a familiar lamp she likes. Moreover, expressions characteristic of sadness, anger and pain/discomfort are seen together in a range of situations from medical inoculations to separation from the mother. This seems
to show that there is no connection between the expression and the emotion in very young children.

More recent work from Camras and her colleagues (Camras et al. 2003), as well as from Oster (2005), has looked at faces of babies in fear and anger situations. What they found is that babies do not produce discrete adult-like facial expressions of these emotions, but instead show varying expressions of distress that do not look like prototypical expressions of anger or fear (though there are differences in visual fixation and body movements). Brow and mouth configurations identified as specific to fear and anger were not differentially produced in the fear or anger paradigms.

Camras, like Russell (2003), and Frijda and Mesquita (1998), as well as some functionalists like Barrett (1998), believes that there is no need to explain the packaging of components in one unified whole, because there is simply no such thing. Using, among other things, ideas about communication put forth by ethologists (Evans & Marler 1994), she proposed that emotions are much less rigid than supposed by the tenants of the DET. On certain occasions it might be useful to smile, while on other occasions it is preferable to keep your smile to yourself. What you will express often depends on who is present. This is obvious in the case of emotions expressed in social relations where hierarchy is involved (but see also Ruiz-Belda et al. 2003, about bowling players that smile after a strike only when they are facing the audience). If such is the case, emotions do not induce facial expression but may only facilitate them.

3.3.3 Emotional experience
As we said earlier, partisans of DET are not only suggesting that facial expressions are genetically determined and present early on in life, but also that there is an innate link between the appraisal mechanisms and the emotional expressions. One consequence is that emotional experience (or what they call “emotional feeling”, granting that feelings can involve cognition; Ackerman et al., 1998) is also present early in life because emotional experience is a product of what is called the “neuro-evaluative” or the “emotion activation process” (Izard also talks of a “non-cognitive motivational condition”). The consequence of this position is that

(...) emotion experience does not change with development. More specifically, the basic motivational/feeling state of an emotion is invariant. (Izard, 1994, p. 359)

or

DET assumes that the core feeling state of any discrete emotion is constant across the life span. (Dougherty et al. 1996, 30)
Or again

The infant experiences anger as a primitive organic sensation. (Magai & McFadden, 1995, p. 154)

This view has been opposed by many, such as Camras (1992), who suggests that development goes from less differentiated forms of motivation (like distress) to more differentiated states (like anger or sadness). Other psychological or social constructionists seem to be committed to this view, because they place cognitively elaborate interpretations (typically not given at birth and subject to learning) at the center of emotional experience. This is the view endorsed by Lewis (2000a), for whom

Emotional experience is the interpretation and evaluation by individuals of their perceived emotional states (i.e. changes in their neurophysiological behavior), as well as the situations in which the changes occur, the behaviors of others, and their own expressions. (Lewis, 2000a, p. 272; our emphasis)

Russell (2003, 2005) develops a similar position. For him, our emotional experience is what he calls a form of “secondary consciousness”. What he means by that is that our experience depends on the culturally and socially informed categorization of some raw feelings. These raw feelings, or “core affects”, as he calls them, vary according to two dimensions: pleasure/displeasure and activation/deactivation. By themselves, the core affects are not ‘emotional’. As Tim Schroeder humoristically puts it: “A woman who tears up because of a blustery wind, while an ill-advised burrito weighs heavily upon her digestive tract, feels an impressive number of the sensations felt by someone who is gut-wrenchingly sad.” (forth.). Imagine moreover that, as you might expect, her state causes her displeasure. You would still not be tempted to say that she is feeling sad. Emotional experiences necessitate the categorizing the core affects. As Russell says, “[one] does not simply introspect and register the reality of a state of fear, (…) rather, (…) much information processing intervenes between the registration of the raw data and the final percept (…). To experience fear is to perceive a strong resemblance between one’s current state as one knows it and the mental script for fear” (2003, p. 164).

This view is also the one favored by the constructionists:

Even responses that are relatively automatic are experienced as emotional only to the extent that they are interpreted within the framework of an emotional syndrome: for example, it is this reflexivity that transforms mere arousal (from climbing the stairs, say) into emotional arousal (e.g., an angry episode). (Averill et al., 2001, p. 170)

It is hard to see how we could settle this debate, as we cannot put ourselves in babies’ heads to know whether or not they feel distinct emotions. One possible strategy would be to use brain imagery to look at the patterns of activation in babies’
brains when they appear to experience emotions. Some work (Damasio et al., 2000; Lawrence & Calder, 2004) have already shown particular patterns of activation for different emotions in adults. If babies’ anger, for instance, shows patterns of activation similar to adult anger, then one might be tempted to conclude that babies have the same emotional experiences as adults. But we take it that psychological and social constructionists would probably not be moved by this. The raw feel that children would have when being frustrated or when being joyful is not what they want to call “emotional experience”. To experience these feelings as emotional, one needs concepts (and scripts). So it seems that here the debate could be semantical, each party having its own understanding of what the expression “emotional experience” means.

3.3.4 Emotional dispositions

Some data from longitudinal studies seems to indicate that the plasticity of emotional dispositions or temperament is somewhat of a vain dream. Some of the best of those studies are the ones conducted by Caspi and his colleagues on a largely intact cohort of 1037 subjects in New Zealand that has been studied every three years from age 3 to 26. The findings are summarized by Caspi (2000) as follows: undercontrolled kids (kids who were impulsive, restless, negativistic, distractible, and labile in their emotional response) at 3 were rated as exhibiting more externalizing problems (fighting, bullying, lying, disobeying) at age 5, 7, 9, and 11, as well as adults. Caspi writes of these children that:

[I]n terms of their personality structure at age 18, they were characterized not only by high levels of impulsivity and thrill seeking but also by aggression and interpersonal alienation. By age 21, undercontrolled children reported more employment difficulties and higher levels of interpersonal conflict at home and in their romantic relationships. They had extensive brushes with the law, and their successful assumption of adult roles was compromised by their abuse of alcohol. People who knew them well corroborated this profile of conflicted interpersonal adjustment in describing undercontrolled children grown up as unreliable and untrustworthy (Caspi, 2000, p. 168).

The stability exhibited by children with a certain temperament, combined with the discovery of the fact that some genes are associated with predisposition to violence or depression (see next few paragraphs), seems to favor an interpretation of the

33. Using a totally different paradigm, Keltner (2003) has looked at women’s college yearbook photos and asked raters to code for the intensity of the smile. They then measured personality, relationships quality and personal well-being over the next 40 years. They have shown that women with the most intense Duchenne’s smiles are, among other things, less likely to remain single and more likely to experience a satisfying marriage.
development of emotional traits focused only on the infant's (innate) biological contribution.\textsuperscript{34}

One should avoid jumping too quickly to this conclusion. Recent work by Caspi and his colleagues (2002, 2003) on depression and violence in adults who have been identified since age 3 as having a certain temperament underlines the role of environment in the shaping of emotional dispositions.

Depression for instance, has been linked with a certain form of polymorphism in the promoter region of the serotonin transporter gene (5-HTT), such that those having one or two copies of the short allele of the 5-HTT exhibited more depressive symptoms than those having two long alleles, but only if having been exposed to stressful life events. As they write in the conclusion of their paper: “(...) no direct association between 5-HTT gene and depression was observed” (p. 389). The simple possession of the gene is not enough because “the gene's effects are expressed only among family members exposed to environmental risks” (p. 389; For a similar conclusion concerning violence, see their 2003, p. 853).

For violence, the same kind of polymorphism exists for the promoter of the Monoamine Oxidase A gene (MAOA). It is known that genetic deficiencies in MAOA are linked with aggression in humans (since it is a X-linked gene, it is more likely to affect males) and in mice (in mice deletion of the gene encoding MAOA increases aggression as well as levels of norepinephrine (NE), serotonin (5-HT) and dopamine (DA), while aggression is restored to a normal level by restoring the expression of MAOA).

Caspi et al. (2002) report that animal studies show that the stress associated with maltreatment early on in life alters the production of neurotransmitters like NE, 5-HT and DA (not to talk about the effect of stress that has also been shown on cerebral structures central to some emotions like the amygdala, see Joseph 1999). Caspi et al. have shown that anti-social behavior is predicted by an interaction between a gene (low activity MAOA genotype) and an environment (maltreatment in early age). By themselves, genes are not enough for one to develop anti-social behavior. As they put it: “[]males with low-MAOA activity genotype who were maltreated in childhood had significantly elevated antisocial scores relative to their low-MAOA counterparts who were not maltreated. In contrast, males with high-MAOA activity did not have elevated antisocial scores, even when they had experienced childhood maltreatment” (2002, p. 853).

So it is not only the genes that are responsible for later behavior and emotional dispositions, but the environment we grow in contributes to a large extent too. That might be a reason to rejoice: despite the gloomy prospects of the longitudinal stud-

\textsuperscript{34} As such, the Caspi studies tell us nothing about how the children comes to have a particular temperament in the first place.
ies of Caspi and Kagan, emotional dispositions are not established once for all. Changes are possible (indeed, childhood prognoses are accurate up to one's 20's, but they do not yield groupings that are valid over one's entire life). This becomes more apparent when we consider that part of the stability of temperament during one's life may be created by the fact that “people create environments that maintain stability”. To give just one example, people who choose spouses similar to them appear to be less likely to go through personality changes than those who choose spouses dissimilar to them (Krueger et al. 1998; Carstensen et al., 2003, p. 728).

The choice of a mate is not the only way by which an individual might contribute, by his actions and choices, to the stability or change in his emotional dispositions. Studies like those by Caspi show that undercontrolled boys are less likely to become involved in crime if they stay in school35; other studies show that investment in social bonds in adulthood can be a turning point in a criminal life; or that a good marriage (not a shotgun one!), regardless of the spouse's own deviant behavior, or a good job, can make one abandon criminal life (Laub, Nagin & Sampson 1998; Sampson & Laub 2004). Finally, experiments done over a period of six months by Keltner (2003) show that romantic partners and roommates are getting closer together emotionally compared to controls, proving the effects that intense and frequent contact with others might have on emotionality.

Changes are not only taking place later in life, they can take place much earlier, as shown by Kagan's data according to which a large number of children who are shy in infancy are no longer shy when they enter preschool. It seems that parents were instrumental in the disappearance of the trait (by, for instance, facilitating encounters in safe environments for inhibited children). Caregivers also promote empathy in helping to notice the effects of transgression on others. The attitudes of the parents towards the emotional demonstrations of their children also have an effect on future emotional regulation. For instance, the reaction of parents to a demonstration of anger by a child has an influence on later use by the child of this strategy (it should be noted that the acceptance of irritability is a function of cultural norms and social contexts36). As Lemerise and Dodge note:

35. True the decision is not always theirs, but this still shows that social factors (like the decision of educators or parents) impact the development of temperament and personality, therefore it is predetermined or inflexible.

36. As Kopp and Neufeld put it: "In Western urban samples, data reveal that infant irritability has long-term adverse relationship consequences. (...) However, it is unclear whether adverse parental responses to infant irritability are a function of the demands of urbanized societies (e.g. job responsibilities, spatial constraints), prevailing cultural norms (e.g. acceptance or rejection of acting out), or a combination of the two (e.g. low tolerance of irritability magnified in families who operate on tight schedules versus high tolerance for toddler irritability where family scheduling is loose). Stated another way, the issue is the extent to which infant and toddler irritabi-
For example, mothers who responded to toddlers’ anger with calm neutrality or cheerful displays had toddlers who showed interest in the environment, positive emotions, and positive responses to strangers in the mothers’ absence. (…) On the other hand, (…), angry maternal responding to toddlers’ difficult behaviors was associated with toddlers’ persisting in angry, noncompliant behavior and being less likely to respond empathically to others. (Lemerise & Dodge 2000, p. 597; see also Kochanska et al. 2005, p. 1937)

Anger during interactions with children produces shyer children, less pro-social behavior towards parents and peers and a poorer understanding of emotion in children. This translates into less social competence and therefore into difficulty in getting others to do what one wants, or more occasions when one gets angry. Indeed, a more general positive disposition of the mother towards the infant contributes to establishing a “form of positive disposition that was more enduring than a mere positive mood (...) More likely, it promoted the child’s deeper and lasting positive disposition toward the mother, which in turn, over time, fostered his or her eagerness to embrace the mother’s rules and values” (Kochanska et al. 2005).

Another factor that has been thought to affect children is inter-parental conflict. It is thought that exposure to inter-adult anger sensitizes children towards anger and makes them more aggressive38. But this is not true of every case:

37. In fact, the appropriateness of mothers’ responses depends on the type of kids they are reacting to: “Importantly, Kochanska (1997) shows that maternal styles interact with toddlers’ temperament in the development of conscience. In fearful children, gentle discipline is ineffective; instead, positive responsiveness from the mother is required for socialization of conscience” (Carstensen et al., 2003, p. 731). Bates makes a similar point concerning the effect of different parenting styles on children with different temperaments: “Resistant children with mothers who responded relatively often to their actual and potential misbehaviors were not as likely to develop externalizing behaviors as resistant children with less controlling mothers; at the same time, nonresistant children with highly controlling mothers showed higher levels of behavior problems than nonresistant children with less controlling mothers.” (2000, p. 391)

38. Harris (1995) maintains that the influence of parents on children is circumscribed to the home. According to her, socialization is context specific, therefore, outside the house, children adapt their behavior and reactions to their peers. As she puts it, “Intra- and intergroup processes, not dyadic relationships, are responsible for the transmission of culture and for environmental modification of children’s personality characteristics” (1995, 458). This seems to contradict the studies we are presenting in this section. But since we are interested in showing that emotional dispositions are open to socio-cultural influence, we would be as happy if Harris was right.
Jenkins and Smith (1990) found that the presence of a grandparent to whom the child was close was associated with a much reduced risk of psychopathology in children experiencing high levels of interparental conflict. Presumably through such relationships children develop a sense that relationships other than those based on dominance can occur. (Jenkins & Oatley 1996, p. 434)

So it seems that all in all emotional dispositions can not only be explained by the infant's biological contribution, but rather by some kind of dynamical process in which the temperament of the infant, the parents, the social environment and the cultural norms all play a part.

3.4 Discussion

So what are we saying about plasticity in development? We pitted two models against one another in this section and we think that the developmental system model has shown that it has to be considered as a serious contender in the explanation of the development of emotions. As such, this might be considered by some as a meager result. But it is not.

First, let us look back at the plasticity models. We saw that biological determinists are tempted by versions of the Marble Model or the Avocado Pear Model both with respect to the entire emotion package and with respect to particular emotional dispositions. Others, the advocates of the developmental system model, are more in favor of something like the Wax Model or the Silly-Putty Model. The results considered (results bearing on children’s development) do not favor the Clay Model as emotional dispositions seem to be open to change during the whole period of childhood and even later. What about the WM (Wax Model) and the SPM (Silly Putty Model)? We see dynamical system advocates as possibly endorsing one of the two. People like Russell make it clear that “each emotional episode (…) is constructed on the occasion of its occurrence. As suggested by dynamic systems theorists (…), no overall pattern is fixed ahead of time” (2003, p. 166–7). This is clearly suggestive of total plasticity all the time, more SPM than WM. On the other hand, given that the categories used to construct the emotions, as well as the core affects, are relatively stable once a person is an adult, WM might be more accurate as a model of adult plasticity.

Now, considering the different components or aspects of emotions, the results are the following:

a. Everybody in the debate acknowledges that the appraisals we make are culturally influenced and that appraisal is pretty plastic. The unsettled issue is how to think about “preparedness”, is it or is it not the result of a complex form of conditioning? If it is, appraisal might be totally plastic (and we are back to the heyday of the behaviorists!). Another issue we mentioned is the one concern-
ing genetically programmed fear. DS advocates claim that this kind of fear appears only in certain conditions, i.e. when children start locomoting. If such is the case, this would be another blow to the SDM model.

b. Concerning the expression of emotions, we have shown that the debate is still on-going concerning whether they are present at birth or whether they are constructed afterwards. One thing for sure is that it seems their development is pretty much canalized, that is to say whatever the cause of our expressions everybody ends up pretty much with the same set of expressions. However, work by Camras et al. (2003) shows that culture might act quite early on emotional expression, as witnessed by their discovery that 11-month old Chinese exhibit emotions less than Japanese and Americans of the same age. This would suggest some level of plasticity very early on.

c. The debate concerning emotional experience seems to revolve around two things. The first one is, do infants have a set of discrete emotional experiences or are they experiencing some kind of blurred experience that gets more and more defined with age? The second one concerns the construction of the experience. Is it done around a very well defined emotional experience (this would be a form of APM, Avocado-Pear Model), or around global hedonic experiences (the core of the avocado would not be emotional).

It might seem that the scientific literature on the plasticity of the appraisal mechanism, of expression and of experience, is not very conclusive. But fixed or plastic, the emotions we undergo are shaped by family and peers’ responses to them. This is good news for us. Put aside the question of the capacity of emotions to be psychologically or socially constructed, it would appear that whatever the temperament or the emotional dispositions you are born with, you can change[^39]! More precisely, the social environment plays an important role in shaping our emotional dispositions. What we need is some social or family engineering.

4. Conclusion

What do our findings entail with respect to evaluative questions? As we have seen, the question of the plasticity of emotion has a bearing on different ethical and evaluative issues. We noted that virtues and vices are usually thought to involve dispositions to feel certain emotions in certain situations. Thus, the plasticity question is relevant to the question whether or not we can get ourselves to become

[^39]: As philosophers, we are interested by the mere theoretical possibility of change. Most people do not choose their family and therefore are determined by this aspect of their environment. So for most people, it is quite possible that this change is not under their control.
more virtuous agents and whether or not we can educate children to become virtuous. As we mentioned, there are reasons to think that the possibility of shaping ourselves emotionally is not only a condition for becoming better agents, but for moral responsibility quite generally. Moreover, given the link between emotions and values, the plasticity of emotions is required if we want to make room for the idea that our grasp of evaluative considerations can improve. Finally, plasticity is presupposed to make sense of the different norms that pertain to emotions.

Now, even though we haven’t tried to give a picture of what the education of emotion involves, the literature we have reviewed allows us to be optimistic about these issues. Firstly, there are reasons to think that appraisal is importantly plastic, so that the kinds of fears or disgust, etc., we experience are not fixed. Secondly, we have seen that emotional dispositions can change. What has to be emphasized is the role of the social environment in shaping emotional dispositions and more generally our temperament. This suggests that often the best an agent can do is to try and immerse herself in the right social environment.

These questions bear on the plasticity of our emotional dispositions. As we have seen, this question has to be distinguished from that of the plasticity of our occurrent emotions, something which involves the possibility of controlling the different components of emotions. We said little about this, but as we mentioned in the section on emotional dispositions, coping skills are something which can be acquired and which depends on the social environment. This question merits further investigation. In particular, the question of how we can get ourselves in and out of occurrent emotions needs to be addressed (Ben-Ze'ev, 2000).

In closing, we would like to make a few remarks about additional limits of our discussion and the directions in which further research should be headed. One problem we have not touched upon much is the plasticity of adult emotions (though we did mention that anti-social behavior can sometimes diminish or disappear, as in the case of an individual who finds himself in a good marriage). Is it possible for adults to change? In ethics, the question is generally not to be understood as “Can we feel the emotions felt in another culture?”, but “Can I react differently to certain situations or people, can I break the mold of my emotional habits?”. We have not really talked about this, but here is a relevant closing observation. Magai and Nusbaum (1996) have been interested in personality changes in adulthood. They propose that some important personality changes appear to be the result of ‘life crises’, that is, some turning point in someone’s life. In order to endure, they note, these crises have to be followed by a period of reflection and emotional work. As they say:
Perturbation in the working system that is personality creates conditions in which established feedback loops are temporarily disrupted, setting into motion novel positive and negative feedback interactions. (Magai and Nussbaum 1996, p. 407)

Those more likely to make changes are those who possessed not only motivation (for instance, because of the “crystalisation of their discontent”) but also those who possessed the cognitive and verbal tools “that fostered active self-reflection”. This is clearly a fruitful research direction. It would imply that one very important tool for change is “emotional intelligence”.

Another issue we left aside (for now) is how we can modify our dispositions to experience an emotion through art (Nussbaum, 1990). Quite obviously, it would also be important to look into the vast empirical literature about different kinds of psychological therapies and their efficiency. Neither did we consider how moods and emotions are produced by drugs or by the manipulation of our genes. We have no principled objections against such strategies. Some theorists might object to that such changes would not involve any personal work, and therefore conclude that they has no moral worth. But this is an open question.40

References


40. We wish to thank Roberto Casati, Jérôme Dokic, Catherine Driscoll, Elisabeth Pacherie, Dan Sperber and Peter Zachar for comments on previous versions of this paper. We would also like to thank audiences at the Institut Jean-Nicod in Paris, the congress of the SSPP in Raleigh, the conference on the modularity of emotions in Montréal, the fifth congress of the ECAP in Lisbon for their comments and questions. We also wish to thank the SSHRC for its financial support.


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


Attributing aberrant emotionality to others

Nick Haslam and Stephen Loughnan

1. Introduction

In this chapter we address the issue of fact and value in emotion from the perspective of social perception. How are emotions ascribed to other people, and what implications follow for how we behave towards them? This question raises factual issues about the emotional repertoire that people typically attribute to other groups and individuals in their social environment. There has been considerable attention to this matter in recent social psychology, as it has important evaluative and normative consequences for social action. In brief, our argument is that people frequently ascribe different and lesser forms of emotionality to other groups and individuals than to their own groups and themselves. This ascription, in turn, may weaken the value that people place on others’ welfare, their emotional engagement with others, and the compunction they feel about harming others. We situate the attribution of emotionality to others in the context of a new theoretical account of dehumanization, in which emotion features prominently.

2. Perceiving emotions in others

From a social psychological standpoint, emotion has normative and evaluative dimensions that can be examined on many levels. Most simply, of course, emotions tend to carry a subjective valence: their experience is desirable or undesirable. More interestingly, many social emotions convey normative judgments about social actors. Pride and shame reflect our perceived standing with respect to our values and standards, just as admiration and disgust reflect the perceived standing of others, and envy and jealousy reveal our judgments of our standing relative to others.

Some writers have taken the further step of examining the moral codes that underpin particular emotions, laying bare the intricate forms of moral cognition.
that generate subtly distinct forms of emotional appraisal. Rozin, Lowery, Imada, and Haidt (1999), for example, show how contempt, anger and disgust are generated from perceived violation of norms of community, autonomy, and divinity. By this sort of account, some emotions, at least, are fundamentally normative, not merely evaluative in feel or judgment, but dependent on elaborate moral reasoning.

Other social psychologists have shown how emotions commonly involve evaluation not only of particular acts or individual actors, but of social groups. Some emotions have an inter-group dimension that involves collective forms of normative judgment. A rapidly growing body of work (e.g., Mackie & Smith, 2003) examines “intergroup emotions” that depend on, and sometimes contribute to, particular forms of intergroup relations. Collective guilt, for example, may be experienced by members of one group in relation to a group that they have harmed in their earlier dealings. Envy may be directed towards groups that succeed relative to one’s own in a competitive struggle, and pity towards those that are perceived to be harmlessly incompetent. In place of a straightforward psychology of group perception and prejudice, where groups are monochromatically liked or disliked, this work proposes a more nuanced analysis in which group evaluations are colored by an emotional palette.

All of these lines of work demonstrate how emotions are intimately involved in the evaluation of acts, individuals, and collectives. However, they all consistently take the perspective of the actor’s emotionality. How, they ask, does the person experience emotion, generate emotion through the appraisal of their own or others’ norm violations, or respond emotionally as a member of one group in relation to other groups? The self is the locus of emotion when these questions are asked. A question that has received much less attention requires a shift in perspective, asking about the emotionality that people ascribe to others. We might ask how people attribute emotions to others, and whether the process of attributing emotions and emotionality to others has normative dimensions in the same way that people’s evaluations of others inform their own emotions.

In this chapter we ask questions of this sort. Our focus is on the ways in which people tend to attribute different types of emotion or different degrees of emotionality to others than they do to themselves, and also on the moral or normative dimensions of these processes. Our position is grounded in our own empirical work on social perception and on the work of an influential group of European social psychologists. In their work and in ours, the attribution of aberrant emotionality to other groups and individuals has come to the fore. We analyze these attributions in terms of the denial of humanness to others (i.e., dehumanization). It need hardly be said that the category “human” has profound normative implications, and that dehumanizing others involves a particularly troublesome form of evaluation. Where our position diverges from that of our European colleagues is that we see
humanness as having two rather distinct senses, which are differently linked to emotion. This distinction warrants a brief analysis before we proceed further.

One intuitively obvious sense of humanness is comparative: humanness is what distinguishes us from animals. As Kagan (2004) writes:

> We can describe an object by listing its features... or by comparing the object with one from a related category... Most answers to the question What is human nature? adopt this second strategy when they nominate the features that are either uniquely human or that are quantitative enhancements on the properties of apes (p. 77)

In the emotional domain, humanness as that which is unique to our species is reflected in the distinction between primary and secondary emotions, where only the latter are uniquely human. For example, anger is primary while guilt is secondary.

However, as Kagan notes, it is also possible to define a category by what is characteristic or typical of its members, rather than by what distinguishes them from a contrastive category. If we understand humanness in this sense, it represents “human nature”: those fundamental, basic and species-typical characteristics that make us who we are, whether or not they distinguish us from other animals. Our research shows that people see emotionality itself – not just the subset of cultivated, secondary emotions – as constitutive of human nature. To deny human nature to a person or group is therefore different from denying them uniquely human characteristics. Each implies the attribution of different forms of aberrant emotionality. In one case, the person or group is perceived as lacking emotional responsiveness, and in the other they are perceived as lacking emotional refinement.

We return to these issues throughout this chapter. We first review research and theory on “infrahumanization”, the process whereby uniquely human secondary emotions are denied to other groups relative to one’s own. According to this research tradition, the emotions that distinguish us from other animals are reserved for the ingroup, so that outgroups are implicitly or explicitly likened to animals, or at least not distinguished from them.

We then turn to our other proposed form of dehumanization, in which others are denied not just a uniquely human type of emotion, but emotionality more broadly. On our view, when groups are seen as lacking human nature, of which emotional responsiveness and warmth or “heart” is a core feature, their members are likened to inert objects or automata rather than animals.

After exploring how the denial of emotionality to others operates in the perception of groups, we then examine recent findings that people attribute greater emotionality – and human nature – to themselves as individuals compared to people in general. If people attribute lesser emotionality to others than to the self, their selfishness, lack of empathy and other forms of inhumanity may be easier to explain.
3. Attributing aberrant emotionality to groups I: Denying uniquely human attributes

As we have argued, there are two primary ways in which humanness can be denied to others. European social psychologists have recently identified a form of bias in group perception, dubbed “infrahumanization”, in which groups are judged to lack humanness in a “uniquely human” sense. Although, in principle, this bias could be investigated by examining the attribution of any uniquely human attribute, research on infrahumanization has rested on the venerable distinction between primary (or “basic”) and secondary emotions (Stein & Oatley, 1992). The former are understood to be shared with our mammalian ancestors whereas the latter are putatively unique to Homo sapiens.

Importantly, infrahumanization researchers have demonstrated that laypeople make a distinction that essentially corresponds to the scientific one, differentiating between emotions shared by both humans and animals (e.g., sadness, joy, fear) and those unique to humans (e.g., nostalgia, exuberance, regret) (Demoulin et al., 2004; Rodriguez et al., 2005). This folk distinction is lexicalized within Romance languages (e.g., sentiments and émotions in French) but implicitly understood in others. Compared to primary emotions, laypeople judge secondary emotions to be experienced differently across cultures, to be less observable, to involve thinking, to emerge later in development, and to implicate moral character. In short, uniquely human emotions are seen as more complex, cultivated, and refined.

Given that all human groups are composed of people, the attribution of emotions to groups that one does or does not belong to (“ingroups” and outgroups”) should be equal. However, infrahumanization research has revealed that people attribute fewer secondary emotions to outgroups (Leyens et al., 2000). Put simply, people deny complex emotionality to outgroup members. This differential attribution of refined emotions occurs for both positively and negatively valenced emotions, and thus cannot be reduced to the well-known phenomena of ingroup favoritism or outgroup derogation (Demoulin et al., 1999).

Interestingly, research into the attribution of primary or raw emotions has yielded conflicting results, with some studies reporting no difference in attributions (Leyens et al., 2000) and others suggesting that outgroups are attributed more primary emotions than the ingroup (Cortes, Demoulin, Rodriguez, Rodriguez, & Leyens, 2005; Leyens et al., 1999). If outgroups are indeed attributed more primary emotions than the ingroup this might suggest a perceived closeness to nature or animality. It may also compound the perception of aberrant emotionality by not only denying outgroups their refined emotions but also emphasising their emotional simplicity.
One interesting aspect of infrahumanization research is its generality. Outgroups that are denied uniquely human emotions need not be only those that are disliked, feared, or in conflict. Leyens and his colleagues have shown that infrahumanization can occur reciprocally between groups that do not have a history of significant friction, status differences, or animosity. It is even possible for a group to be perceived as lacking uniquely human emotions but otherwise to be regarded positively, a combination that may indicate condescending or patronizing attitudes. They are cheerful and happy, but also a little bit simple and childish, their emotional palette containing only primary colors.

Secondary emotions are not only commonly denied to outgroups, but when outgroup members express them their displays are often viewed as at best inaccurate and at worst deliberately misleading. Vaes et al. (2003), for instance, found that people do not trust the judgements of outgroup members reporting secondary emotions, and tend to react negatively to them. Outgroup members’ expressions of secondary emotions may even be systematically misperceived as primary. For example, an outgroup member’s expressed love may be misperceived as lust, their regret as sadness, and their moral indignation as simple rage.

Infrahumanization theorists have been relatively silent on the psychological basis for the effect, but the perception that members of other groups are influenced by simpler emotions might arise from two distinct beliefs. First, one might believe that the simpler emotionality of other groups results from their failure to control their emotions to fit social norms: they express more raw emotionality because they lack the cognitive control to restrain themselves appropriately. Alternatively, one might conclude that members of the other group do not possess or failed to develop the refined emotions. Put simply, the outgroup might be viewed as possessing aberrant emotionality because its members fail to control their basic emotions or because they simply lack refined emotions.

Arguably, perceiving the outgroup as lacking cognitive ability might underpin both views. Previous social psychological research has repeatedly found that people generally view social outgroups as less intelligent than themselves (e.g., Crock- er, Major, & Steele, 1998), and that intelligence, liked secondary emotion, is believed to be a uniquely human attribute. This possibility is nicely illustrated by representations of ethnic groups, specifically perceptions of colonised ‘natives’.

Historical representations of the ‘savage’ or ‘native’ have been predominantly negative. Two recurrent themes in these representations are the ‘savage-as-animal’ and the ‘savage-as-child’ (Jahoda, 1999). For example, Africans were typically portrayed as possessing wild or feral emotions (Jahoda, 1999; Jahoda & Krewer, 1997). Their emotional lives were dominated by lust, fear, hatred, and anger. Of course, even the early explorers and scientists could not deny that Westerners also felt these emotions. Thus, in an attempt to maintain a positive distinction between the
colonised and the coloniser, the native was viewed as unable to develop refined emotions or to control their primitive emotionality on account of their cognitive shortcomings.

The failure to develop refined emotions was supposedly symptomatic of the broader perception that ‘natives’ were unable to develop civilisation or adhere to civilized norms of thought, feeling, and behavior. For many early writers such as Romanes (1880, cited in Jahoda, 1999), the colonised peoples were in a stage of arrested intellectual development somewhere around the age of a 5–7 year old Western child. This arrest was claimed to preclude the development of more sophisticated forms of emotionality, and the ability to exert control over the simpler emotions, leaving Africans in the grip of a ‘savage passion’ (Jahoda, 1999).

A related form of emotion attribution may be involved in relation to gender. Gender stereotypes – especially those held by men when regarding women as an outgroup – often suggest that women possess more emotions than men and that women lack emotional control. Perceptions of women by both genders have tended to focus on the richness of female emotionality (Goldshmidt & Weller, 2000; Hess et al., 2000). Many aspects of this stereotype are socially desirable, involving women’s caring and emotional warmth (Eagly & Mladinic, 1989), but negative stereotypes also portray women as jealous, anxious, and emotionally unstable.

Research investigating attribution of emotions to men and women indicates that women are indeed attributed more emotions than men. Women are more likely than men to be attributed both raw and refined emotions, suggesting that perceptions of females differ from those of ‘natives’ by not denying women complex emotionality (Leyens et al., 1999; Viki & Abrams, 2003). However, ethnic and gender stereotypes potentially converge regarding the possession of sufficient cognitive capacity to control raw emotions. Women have historically been viewed as lacking rationality (Jones, in press; Lloyd, 1984), and ‘traditional women’ are viewed as warm but lacking competence (Fiske, Cuddy, Glick, & Xu, 2002; Fiske, Xu, Cuddy, & Glick, 1999). This failure to meet the normative standards of rationality embodied by men might underlie perceptions of excessive and under-controlled female emotionality. Like natives, females may be seen within sexist discourse simply to lack the cognitive control necessary for emotional restraint.

This discussion barely scratches the surface on the topic of attributions and denials of uniquely human emotions. However, infra-humanisation research demonstrates that aberrant emotionality – in the sense of simpler, less refined, and more animalistic emotion – is ascribed to groups in a surprisingly robust and pervasive manner. More civilized emotionality is the province of the ingroup, even when the outgroup is not disliked or in competition. Although this subtle denial of uniquely human attributes to others is not equivalent to regarding them as animals, and does not constitute dehumanization in a strong sense, it nevertheless provides
a basis for psychological distance. If others are less human than one's own group, then they do not have the same moral standing and norms of humane treatment may not apply to the same degree where they are concerned (Bandura, 2002).

4. Attributing aberrant emotionality to groups II: Denying human nature

The infrahumanization phenomenon represents one way in which certain kinds of emotionality are denied to people on the basis of their group memberships. The dimension on which ingroup and outgroup are distinguished is between secondary and primary emotions, uniquely and non-uniquely human traits more broadly, and implicitly between humans and animals. However, as we have argued, “humanness” goes beyond those attributes and comparisons that define the human/animal boundary. Humanness can also be understood, in the sense of “human nature”, as that which is fundamentally, typically, deeply or universally human.

If there are two distinct senses of humanness, there should be two corresponding types of humanness denial. Moreover, if the two senses of humanness have different relationships to emotion, then their denial will imply different views of others’ emotionality. We argue that emotionality itself is seen as a core feature of human nature, so that when human nature is denied to other people, their emotionality is perceived as lacking in quantity, not merely in quality, as occurs in infrahumanization.

At this point we need to step back and ask whether our two proposed senses of humanness have any psychological reality. It also needs to be shown that emotionality is central to people’s understanding of what it is to be human, and that they believe it to reside deep within the person as a sort of essential core. We have conducted a series of studies (Bain, Haslam, & Kashima, 2006; Haslam, Bain, Douge, Lee & Bastian, 2005; Haslam, Bastian & Bissett, 2004) that offer some encouraging answers.

At the beginning of this line of work (Haslam, Bastian & Bissett, 2004), we had a sample of undergraduates make a series of judgments about the nature or ontological basis of a set of personality traits. These judgments focused on the extent to which particular traits were essentialized: seen as deeply rooted in the person, identity-determining, biologically based, difficult to change, and discrete. We also had our participants rate the extent to which each trait was a component of “human nature” and the degree to which it involved emotion.

Despite the abstractness of these judgments, participants showed high levels of agreement. With respect to the primary concern of the study, participants revealed a coherent essentialist understanding of personality traits. Those traits that were judged to be deep-seated, for example, were also judged to be immutable,
identity-defining, and so on. Importantly, the extent to which traits were essentialized correlated with their perceived emotionality: emotion-related traits are believed to be inhering essences to a greater extent than other traits. Moreover, those traits that were believed to constitute human nature were understood to be both essential and emotional. In short, emotional characteristics are understood to constitute deep and central features of humanness.

This finding that emotion occupies a central place in conceptions of human nature was replicated in some of our later work (Haslam et al., 2005). Once again, ratings of a set of traits indicated that those judged to be emotion-related tended to be seen as basic components of human nature. Traits rated central to human nature were also understood to be cross-culturally universal, prevalent within the population (i.e., descriptively normative), early to emerge in human development, and deep-seated. These traits tended to involve emotionality, openness, warmth, imagination, and vitality. In sharp contrast, traits judged to be uniquely human – those that relate to the sense of humanness adopted by infrahumanization researchers – were judged not to be particularly emotional, and were rated as relatively culture-specific, uncommon in the population, and late to emerge. To our participants, that is, “human nature” represents a somewhat Romantic and universalist sense of humanness, constituting what is shared, inborn, and passionate in human psychology. Their belief that a psychological characteristic is rooted in human nature is no idle philosophical speculation, but also appears to contribute to the importance that they ascribe to it (Bain et al., 2006).

If human nature represents emotionality, openness, warmth and vitality, groups perceived as less than human on this dimension would be seen as inert, rigid, cold, passive, and lacking in vital agency. In previous work (Haslam, 2006) we have argued that some individuals and groups are dehumanized in precisely this fashion. Whereas the denial of uniquely human attributes implicitly or explicitly likens others to animals, the denial of human nature – the animating emotional warmth that is essentially human – likens others to automata or objects.

Many writers on dehumanization are clearly referring to the denial of this sense of humanness to others, where others are seen as robotic, unfeeling, passive, or mechanical. For example, Montague and Matson (1983) present a critique of “technological dehumanization”, a “reduction of humans to machines” (p. 8) that involves the robotic pursuit of efficiency, regularity and conformity, plus apathy, and lack of spontaneity. Hoberman (1992) argues that sport science has dehumanized athletes by attempting to perfect the human engine with artificial regimes of biomechanical and pharmacological improvement. Szasz (1973) takes psychiatric dehumanization to involve a “mechanomorphic” style of thinking that “thingifies” people and treats them as “defective machines” (p.200). Some feminist writers (e.g., MacKinnon, 1987) argue that pornography dehumanizes women in this way,
presenting them as inert and commodified objects lacking agency, autonomy, and feeling, and having only instrumental purposes.

The central role of emotion in this kind of denial of humanness is nicely illustrated by Sobchack’s (1987) analysis of dehumanization in science fiction films. She notes that the key clue for determining who has not (yet) been “taken over” by “cold and passionless alien beings” (p. 120) is the retained capacity to respond affectively to another’s suffering or love. The viewer sees that someone has become an automaton when the person fails to gasp in shock as a dog is about to be run over by a car or fails to return a passionate kiss. People who have been taken over or possessed “look like their human counterparts but are really emotionless simulacra” (p. 122) who respond automatically and lack the free will to do anything but what they are instructed by their wicked masters.

In our own research we have begun to find evidence of subtle denials of human nature in group perception. Most of our work either involves ratings of different groups on personality traits judged to be high or low on human nature, or “implicit” (i.e., automatic) associations between groups and these traits in experiments using social cognition methodologies. Much as infrahumanization researchers find that outgroups are attributed fewer uniquely human nature emotions, we frequently find that certain groups are rated lower than others on human nature traits. For example, Bain and colleagues (e.g., Bain, Haslam, de Souza, & Kashima, 2006) have shown that psychology students attribute more human nature traits to themselves than to medical students, and that Australians rate themselves above some other nations on these same traits. In a similar vein, Loughnan and Haslam (2007) found that experimental participants associated human nature traits less strongly with a group hypothesized to be low in human nature (businesspeople) than with one hypothesized to be high (artists). Moreover, business people were associated more strongly with automata, the non-human other that represents the absence of human nature, whereas artists were more strongly associated with animals, the contrastive category for what is uniquely human.

Conceptualizing humanness as human nature also illuminates the study of gender, which we discussed earlier in relation to what is uniquely human. Bain et al. (2006) demonstrated that women were rated substantially higher than men on human nature traits. Women’s distinctive traits, including their supposed emotionality, are positively framed as fundamentally human. By implication, men were perceived as being alienated from their humanity to a greater extent than women, and are at least in this respect less fully or deeply human. Although, as the previous section of this chapter argued, women may sometimes be seen as lacking in uniquely human qualities and thus implicitly likened to animals, men may sometimes be seen as lacking in human nature traits and thus implicitly likened to
automata. These very associations have been experimentally demonstrated by our research group (Reynolds & Haslam, 2006).

Although research on the differential attribution of human nature to groups is less extensive than work on the attribution of uniquely human characteristics (infrahumanization), the results to date suggest that it is at least equally important in the perception of social groups. The two senses of humanness are plainly distinct in meaning, and have distinct nonhuman contrasts: automata and animals. Importantly, emotion plays a central role in both senses of humanness, and the attribution of emotionality to groups is therefore a crucial domain in which subtle forms of dehumanization can be studied. In both cases the moral implications are almost self-evident. If another group is perceived to lack humanness – whether this is understood as civility and refinement (uniquely human) or as warmth and agency (human nature) – members of that group may be excluded from full moral consideration (Opotow, 1990) and the normal restraints on cruelty and exploitation may be relaxed.

5. Attributing aberrant emotionality to individuals

Up to this point we have discussed the attribution of emotionality in intergroup contexts; how deviant or deficient emotionality is ascribed to people on the basis of their group memberships. Infrahumanization pictures members of outgroups as lacking in refined emotion, and denial of human nature pictures them as affectless automatons. It is also possible that different forms of emotionality are attributed to other people as individuals rather than as members of groups. People might be inclined to perceive themselves as embodying emotion more fully than other individuals, or more fully than average.

There are several plausible reasons why this might be so. For one thing, the subjective feel of emotion is primarily a private matter, and less visible in others than in ourselves. Other people’s emotionality should therefore be less vivid and salient than our own. Second, other people’s emotions may be less salient to us not simply because they are less experientially vivid, but also because they are more likely to be concealed. People actively regulate the public expression of their emotions – perhaps especially those negative emotions that reveal their moral and other weaknesses – and as a result the intensity and range of their emotions are likely to be underestimated by observers. Both of these reasons rely on the relative inaccessibility to perception of other people’s emotions, but there may be another, more conceptual reason for our failure to grasp their full extent. Emotionality may be intimately bound up in our understanding of what it is to be human and if there is any tendency for people to believe that they embody humanness to a greater
extent than others, there may be a corresponding tendency to ascribe greater emotional depth and complexity to the self.

This last suggestion implies that emotionality is widely understood to humanize people. The humanizing power of emotion is beautifully evoked by Paul Auster, in a homage to his ancient typewriter (Auster & Messer, 2002). Referring to an artist's renderings of the machine, he writes:

> Sam has taken possession of my typewriter, and little by little he has turned an inanimate object into a being with a personality and a presence in the world. The typewriter has moods and desires now, it expresses dark angers and exuberant joys, and trapped within its gray, metallic body, you would almost swear that you could hear the beating of a heart. (p.32)

By endowing the object with emotion, that is, it is given personality and heart, a suggestion the object is not all cold and hard surfaces, but also possesses warm depth. Needless to say, people believe that they too have inner depths, and we would argue that one way in which they do so is by believing themselves to possess a more intense or authentic emotionality than other people.

The research we discussed in the previous section shows how people believe emotionality to be a fundamental or core component of human nature, residing within the depths of the person. In a new series of studies, we went on to demonstrate (Haslam et al., 2005) that in addition to attributing humanness to ingroups more than outgroups, people tend to attribute this sense of humanness to themselves more than to others. In a series of questionnaire studies, our participants rated the extent to which they possessed a series of personality traits relative to “the average student” (i.e., more than average, less than average, or no different from average). We examined whether they rated themselves, compared to the average student, as being more likely to possess traits judged to be high in human nature. In doing so, we were at pains to distinguish this effect from the well-established tendency for people to rate themselves as above average on socially desirable traits (“self-enhancement”; Alicke et al., 1985).

In all of our studies, participants ascribed greater humanness (human nature) to themselves than to the average student independent of their tendency to self-enhance. Although the degree to which they self-enhanced was correlated with their self-esteem, their tendency to “self-humanize” was not, implying that the effect is not driven by a desire to regard the self favorably. Interestingly, no corresponding tendency to attribute more uniquely human traits to the self was observed, consistent with previous research on the attribution of emotions (Cortes et al., 2005). It is only one sense of humanness – the emotion-saturated “human nature” – that is linked more to the self than to others.
Further studies have helped to understand the processes that underpin people’s tendency to ascribe greater humanness to themselves than to others. Haslam et al. (2005) found that self-humanizing was explained in part by people’s tendency to attribute to the self traits that are essentialized: they believe they exemplify humanness particularly well because human nature is seen as deep-seated, and they attribute greater depth to themselves than to others. In a related vein, Haslam and Bain (2007) found that people self-humanize less when they compare themselves to a (fictitious) individual than to an undifferentiated “average” person. People even attributed more human nature to their own past selves, selves that have been concretely realized and remembered, than they did to their future selves, which are unrealized and vague.

An implication of these three findings is that people see themselves as more human than others because their representation of others is lacking in depth and concreteness. The self is perceived in three dimensions, and as an individuated unit, whereas the other – and the self in the dimly anticipated future – is commonly perceived shallowly, in terms of traits that are not understood to reflect deep or fundamental dispositions, and abstractly, as an undifferentiated mass, a statistical figment, or an abstract possibility.

Where does this leave us when thinking about the attribution of emotion to self and others? We have argued that greater emotionality is ascribed to the self than to others, and that emotionality is seen as a core, normative component of humanness. When people see themselves as more human, they are seeing themselves as more emotional, in the sense of emotionally open and responsive.

We initially suggested that this perception might simply reflect the greater knowledge people have of their own emotionality: the experiential salience and privacy of their affective states and the fact that other people may actively keep their emotions visible only to themselves. However, this research suggests another factor beyond privileged access or superior knowledge. We may attribute greater emotionality to the self because emotionality, as a fundamental component of human nature, indicates depth and ontological solidity. People ascribe relatively deeply-rooted traits to themselves more than to others because they see themselves as more differentiated and concrete individuals than they see others. Such deeply-rooted traits commonly involve emotionality. In short, we may see ourselves as more emotional than others not just because we know more about ourselves, but because we perceive ourselves – and perhaps are motivated to perceive ourselves – as having a concrete reality and a human depth.

A third basis for the preferential attribution of emotionality to the self may also exist. As we have noted, the traits that people judge to represent human nature tend to involve emotionality, and they are not identical to those that are seen as socially desirable. Many emotions, of course, are unpleasant, and many emotiona...
traits are evaluated negatively. Indeed, when we examined the traits that underpin the tendency to self-humanize (Haslam et al., 2005), we found that the phenomenon appeared to be driven, in part, by participants’ willingness to avow higher levels of negative emotionality than they ascribed to the average student. Although negative emotionality is seen as socially undesirable, it is also seen as evidence of a person’s deep humanity, and therefore more readily attributed to the self than other undesirable traits.

We have repeatedly found that the tendency to self-humanize is stronger for negative traits than for positive ones (Haslam et al., 2005; Haslam & Bain, 2007). Although they see their own strengths as somewhat more human than those of others, people are especially likely to see their own imperfections as more human than other people’s imperfections.

This effect raises the possibility that self-humanizing is not simply a cognitive phenomenon, reflecting differing degrees of knowledge of self and other or differing degrees of perceived concreteness and depth, but that it may have a motivational component. People may be driven to humanize their imperfections so as to minimize their sting. Believing our imperfections to be deep-seated may reduce our sense that we are responsible for them and hence blameworthy, and believing them to be normative (i.e., universal and prevalent) may reduce our sense that we are deviant. By this account, attributing greater negative emotionality to oneself than to others is an acknowledgement of imperfection that is decidedly soft. People may acknowledge that they are unusually or excessively emotional, but find comfort in the belief that they are not truly responsible for this deep-seated affective resonance and that it shows them to be “only human”, a common way of seeking mitigation. This acknowledgement of humanizing flaws may even reflect a kind of vanity, in which people draw a perverse sense of superiority from their emotionality. Yes, they suffer, but their suffering shows that they feel things more deeply and are less alienated from their true emotions than the common run of humanity. This kind of vanity may well play a part in neurotic misery, where some people may identify strongly with the idea of having superior capacity for feeling, and become attached to their suffering as a sign of their specialness and authentic humanity. Such “beautiful souls” are not unknown to clinicians.

Some recent empirical work on the acknowledgement of imperfections indicates that humanness and negative emotionality are indeed implicated in it. Koval and Haslam (2006) had a sample of undergraduates make a series of judgments about a list of undesirable personality traits. Participants rated the extent to which they possessed each trait, chose a subset of traits that best described themselves or the average student, and rated the extent to which the traits were aspects of human nature, undesirable, deep-seated, immutable, prevalent, and immoral. We first replicated the basic self-humanizing effect: people chose traits for themselves that
were adjudged higher in human nature, and higher in emotionality, than the traits they chose for others. Second, we showed that this difference was explained in large part by the extent to which traits were essentialized: people chose imperfections for themselves that were more human than those they chose for others because these imperfections were perceived to be deeper and more fundamental. Emotional traits, in particular, were judged to be particularly deep-seated. Finally, we asked which negative traits were more likely to be acknowledged by our sample. Here the results were quite striking: the strongest predictor of an imperfection's acknowledgement was its perceived humanness. In short, the extent to which people endorse a negative trait is primarily determined by its perceived centrality to human nature, and people tend to attribute human imperfections to themselves more than they do to others. In doing so, they tend to ascribe negative emotionality (e.g., “nervous”, “shy”, “high-strung”) selectively to themselves.

We have argued that interpersonal comparisons – and by implication the perception of self and others – reveals some of the same dynamics that occur in intergroup perception. In both domains, people perceive themselves – whether as individuals or as group members – to be more human than others, and commonly do so by denying others the emotions or emotional traits that they ascribe to themselves. In the case of interpersonal comparisons, the self is seen as possessing more emotionality than others as part of a broader tendency for people to attribute greater human nature to themselves. Although the basis of this phenomenon is only beginning to receive the attention of psychologists, it is likely to reflect some combination of the relatively high salience of the person's emotional experience, the greater perceived concreteness and depth of the self compared to others, and people's desire to acknowledge in themselves only those flaws that reveal their shared humanity.

6. Conclusions

In this chapter we have argued that the attribution of emotionality to others is an interesting and important research question, and one that is intimately bound up with conceptions of humanness. There are two distinct ways in which emotionality can be attributed or denied, linked to two distinct senses of humanness, and these attributions can be examined in the perception of groups, individuals, and the self. With some consistency, it appears that aberrant emotionality tends to be ascribed to others more than to one's group or one's self. Sometimes, the other's emotional repertoire is seen as less refined and complex, differing in quality from our own, and thus locating them closer to animals. Sometimes, the other's emotionality is seen as simply lacking, differing in quantity from our own and thus
locating them closer to inanimate objects or automata. Sometimes we subtly distinguish ourselves from others by the superior depth and vitality of our emotions, perhaps mitigating our failings and limitations by seeing them as rooted in affect, and thus part of our shared humanity. In all cases, the other is seen as explicitly or implicitly as less human than us, and sometimes as less than human.

The normative obligations that govern how we treat members of our species may be weakened when this happens. Inhumanity to others becomes easier when their membership in the human category is seen as marginal or qualified, and when we grant ourselves special emotional depth or discount our misbehaviors by claiming that we are “only human”. This is one clear way in which the facts of perceived emotion can weaken the humane values on which social life depends.

References


1. Introduction

Commonsense leads us to expect that when we make moral judgments, we will often make them passionately. With great indignation I might judge myself to have been treated unfairly. However, I might just as easily have made the same judgment with melancholic resignation. In either case, a negative emotional response attends the judgment that a norm has been violated, and serves as a signal of our commitment to that norm.

At first glance, results from functional neuroimaging appear to support commonsense. When subjects engage in tasks of moral judgment, areas of the brain often associated with emotion show increased activation. However, when interpreting the results of these experiments, it becomes clear that some of the areas activated only qualify as evidence of emotion under substantial assumptions about emotion’s nature. In one study, we find evidence of emotion only if what we mean by emotion is a perception of internal bodily change. In another study, we find evidence only if we mean activity in a mechanism that appraises the status of rewards and punishments.

Recently, Jesse Prinz (2006) has argued that moral judgments are not just accompanied by emotions but depend on emotions, a position called sentimentalism. Simultaneously, Prinz (2004) holds that emotions are composed of perceptions of internal bodily change. Prinz cites a number of neuroimaging studies meant to show that emotions attend moral judgment. But, when we situate these studies constitute part of a larger argument for sentimentalism, based on evidence from social psychology and psychopathology.
studies alongside his theory of emotion, the neuroimaging evidence is less supportive than it would appear.

Prinz’s theory of emotion predicts that the presence of emotion will be indicated by activity in neural mechanisms of bodily perception. Thus, his theory of moral judgment predicts that moral judgments will be accompanied by activity in these same mechanisms. In his own words, Prinz tells us that “emotions co-occur with moral judgments,” and this “…piece of introspective psychology has been confirmed again and again, in every study of what goes on in the brain during moral judgment.” (2006, p. 30) But only two of the studies he cites suggests that perceptions of bodily change co-occur with moral judgment. It may be true that in every brain study of moral judgment, there is evidence of emotion’s presence. But that evidence rarely suggests the presence of what Prinz means by emotion.

In what follows, I will argue that a number of the studies Prinz cites suggest the presence of emotions only when an alternative theory of emotion is adopted. For example, some studies favor the involvement of appraisal processes in accord with an appraisal theory of emotions. Because some of these studies fit a competing account of emotion, they are better regarded as evidence for a competing form of sentimentalism.

To make my case, I will first review Prinz’s sentimentalism in greater detail. This will lead into a discussion of the emotions, where I will review the distinctions between embodied perceptual theories of emotion and appraisal theories. I will then adapt those theories to generate empirical hypotheses in neuroscience. Next, I will review the neuroimaging experiments that Prinz cites in favor of his theory of moral judgment, and explain why Prinz’s interpretation often fails. I will then look at the alternative interpretation of the results and show that in some of the cases, the results favor the involvement of appraisal processes. I will conclude with a discussion of the broader implications for sentimentalism, the theory of emotion and a cognitive neuroscience of moral judgment.

2. Prinz’s sentimentalism

According to the sentimentalist, our capacity for moral judgment depends on our sentiments. On some accounts, a sentiment may be an emotional feeling.3 In the way Prinz uses the term, a sentiment is an emotional disposition, rather than a feeling. When a sentiment is manifested, an emotion occurs in response to the sentiment’s object. For example, if I judge corporal punishment to be wrong, then I will be

---

3. Compare to D’Arms and Jacobson (2000), who regard a sentiment as «any occurrent, object-directed, affect-laden mental state” (p. 723)
disposed to respond with negative emotions toward instances of corporal punishment, and with positive emotions toward the prevention of corporal punishment.

I may or may not get emotional over every case of corporal punishment. Prinz takes this consideration into account, and avoids claiming that a moral judgment must always be emotionally charged. Rather, the idea is that because I am capable of getting emotionally involved over an instance of corporal punishment, I have taken a moral attitude toward that instance. If I am prone to get emotional in a way that is disapproving (e.g. if I’ve gotten indignant), then I have judged an instance of corporal punishment to be wrong. Since judgments are beliefs, Prinz then identifies beliefs concerning the wrongness of acts with sentiments of disapprobation toward those acts.

Prinz’s sentimentalism should not be confused with other prominent accounts. For example, Allen Gibbard (1990) and Simon Blackburn (1998) have separately advocated forms of sentimentalism involving a meta-cognitive component. On these other forms of sentimentalism, when we judge an act to be morally wrong we judge that the act warrants negative emotional responses, such as guilt on behalf of the perpetrator. Thus, a moral judgment is a judgment concerning the appropriateness of emotional responses. Prinz does not make this meta-cognitive ascent.

On Prinz’s view, a moral judgment is not about whether an act warrants a particular kind of emotional response. Rather, certain emotional responses constitute the moral judgment. The act is judged morally wrong, because I suffer guilt for having done it or indignation toward others who have performed the act. I may not feel the emotions on the occasion of the judgment, just as a trucker might not be fully conscious of the highway every moment of an all-night drive (Armstrong, 1981). The absence of emotional feelings does not entail the absence of an emotional response anymore than the absence of the trucker’s awareness of the road entails a lack of visual perception. Both psychological processes, emotion and vision, can occur unconsciously. So, Prinz’s sentimentalism is not one of judgment grounded in feelings. Rather, moral judgment is grounded in a kind of psychological state responsible for emotional feelings. Insofar as emotions are not to be identified with feelings, moral judgments are not to be identified with feelings.

Given some of these observations concerning what emotion is not, one might wonder what Prinz has in mind when he identifies moral beliefs with emotional dispositions, i.e. one might wonder about Prinz’s positive theory of emotion. In previous work Prinz (2004) has argued that emotions are perceptions of bodily state, a view which lends itself nicely to analogies with unconscious visual perception. Forgoing the identification of sentiments with feelings enables Prinz to account for those cases of moral judgment that seem devoid of emotional feeling. The emotion may be present without being felt. One might regard this move as one of the theory’s strengths. But, there are certainly weaknesses.
For example, pre-theoretically it isn’t obvious that there is any connection between somatic perceptions and what one takes moral goodness to consist in. When I conclude that wealthy nations have a moral obligation to meet the needs of burdened societies it isn’t clear why that conclusion should depend on how my body is represented. What perception of bodily state could possibly be relevant to this judgment?

To avoid portraying Prinz’s sentimentalism as a non-starter, it might help to consider the motivation behind Prinz’s theory of emotion. Showing how perceptions of bodily change can be sensibly linked to moral values would be a step in that direction. In the next section I illustrate the kind of dialectic motivating Prinz’s sentimentalism by reviewing embodied perceptual theories and setting them against the appraisal theories of emotion, the leading opposition.

3. Embodiment and appraisal

Different theories of emotion bring with them demands for different kinds of evidence. For example, embodied theories like Prinz’s rely on evidence that emotional experience depends on perceptions of bodily change or somatosensations. In contrast, appraisal theories of the emotions demand evidence that when we respond emotionally we are in some sense attending to how our interests are fairing in the world. In this section, I will review the motivations, strengths, and weaknesses of these approaches to emotion, making plain the kind of dialectic that leads Prinz to his theory.

We will begin with the embodied theories of the emotions, the earliest formulation of which is often called the James-Lange Theory or ‘feeling’ theory of the emotions. To motivate the feeling theory, William James famously wrote:

What kind of an emotion of fear would be left, if the feelings neither of quickened heart-beats nor of shallow breathing, neither of trembling lips nor of weakened limbs, neither of goose-flesh nor of visceral stirrings, were present, it is quite impossible to think. (1884, p. 193)

The passage suggests that we know our emotional states by our feelings, in particular, those feelings that involve the perception of our bodily states. It also suggests that emotions depend on these feelings. Subtract the fearful feelings of your rapid heartbeat, your shallow breathing, your trembling and weakness, and there is no remainder of fear. Fear just is the assembly of such feelings.

4. Alternatively, an embodied theorist might hold that emotions are characteristic physiological changes.
It is now widely held that emotion is more than an assembly of feelings; one can undergo an emotion without feeling it. For example, when we see other people wringing their hands, and when we see their knitted brows, we can tell that they are anxious, and we can ask them why. Yet, they might only come to recognize their anxiety once they are asked. If people can be unaware of their emotional state, and yet appear to be in an emotional state, we might doubt that emotions and feelings are the same.

Observations of this sort shift our focus from emotional feelings to emotions in both their conscious and unconscious manifestations. But acknowledging that emotions can be present without their distinctive feelings does not repudiate the Jamesian insight that emotion is somatosensory. We might allow that bodily perception, like visual perception, can be conscious or unconscious. Then, we might ask whether removing a perception of bodily state removes an emotion.

To further weaken the Jamesian hypothesis, we might not demand that successful perceptual representation be required for emotion. Rather, the mere mental representation of one's body in some select number of states might constitute an emotion, even if the body fails to be in one of those states, or even if bodily representations become disconnected from the peripheral nervous system. This is roughly the notion underlying Antonio Damasio’s view (1999) that emotions can be had in the absence of real perceptual input from the body, and Prinz (2004) follows suit. Perceptions of bodily state might, for example, be hallucinated, and still qualify as emotions. I might represent myself to be trembling, though I do not tremble.

There is much to recommend embodied theories of emotion, but there are also shortcomings. One trouble with pure embodied theories is that they neglect the coordination of our emotional responses with important happenings in the outside world. Perhaps fear is the perception of bodily change underlying one's readiness to flee. But alternatively, one could construe fear as the recognition of a threat. You see the charging bear and you recognize the consequences that might follow from the bear getting to you. For example, you might be killed. Being killed would have dire consequences for your wellbeing, which you care about very much. Fear, we might hold, is the recognition of such threats to our interests.

Appraisal theory does not demand that an emotion be identified with a conscious evaluation of one's situation, though such cases are not ruled out. Pioneers of appraisal theory, such as Magda Arnold and Richard Lazarus, have been explicit on this point (cf. Lazarus (2001) for a review). On Lazarus's view, the appraisals constitutive of emotion concern core relational themes. For example, a judgment that one is in danger is constitutive of fear, while a judgment that something has been lost is constitutive of sadness. Appraisals track the status of our goals, being assessments of whether the world is abiding by our desires or realizing our aversions.
To get a better handle on the nature of a core relational theme, it helps to consult a recent version of the appraisal theory, formulated by Edmund Rolls (2005). On Rolls’ account, we can base our determination of what should go on the list of core relational themes by probing for conditions constitutive of reward and punishment for an organism. A reward is whatever condition an animal will work to obtain, while a punishment is whatever an animal will work to avoid. Thus, rewards and punishments can serve as instrumental reinforcements. Rewards increase the probability of a response on which they are contingent. Punishments reduce the probability of a response on which they are contingent. Some rewards and punishments are primary, being part of our genetic inheritance, while others are secondary, being acquired through their association with primary rewards and punishments. Rolls’ idea is that we might differentiate emotions by reference to both the specific types of rewards and punishments tracked in the brain and the status of those rewards and punishments represented in the brain. Thus, he writes:

Fear is a state that might be produced by a stimulus that has become a secondary reinforcer by virtue of its learned association with a primary negative reinforcer such as pain...Anger is a state that might be produced by the omission of an expected reward, frustrative non-reward, when an active behavioural response is possible. (Rolls, 2005, p. 20)

If the emotion category is more fine-grained, citing the specific type of reward or punishment can add an additional contrast for differentiating the emotion. Thus, Rolls proposes that:

...Guilt may arise when there is a conflict between an available reward and a rule or law of society. Jealousy is an emotion that might be aroused in a male if the faithfulness of his partner seems to be threatened by her liaison (i.e. flirting) with another male...Envy or disappointment might be produced if a prize is obtained by a competitor. (Rolls, 2005, p. 20)

It must be pointed out that Rolls’ analysis of some of these emotions is rather thin. For example, guilt seems to involve more than a conflict between pursuit of a reward and a societal rule. It would seem that one must also regard the rule in question as somehow legitimate, perhaps by internalizing the rule, such that a violation constitutes a source of punishment. Guilt might then be analyzed as the recognition of a form of opportunity cost, pitting the gain from the pursuit of a reward against the punishment of violating an internalized rule.

5. The processes of positive and negative reinforcement in Rolls’ account are representational, so Rolls’ view should not be mistaken for some form of behaviorism.

6. Envy would therefore be an instance of frustrative non-reward, i.e. a case in which a reward was pursued but failed to be obtained.
It should be evident that there are many ways to analyze complex emotions, such as guilt, in terms of rewards and punishments. This might lead one to question the power of Rolls’ approach, but it must be recognized that this is a problem confronting all theories of emotion, and it should be pointed out that insofar as instrumental reinforcement is an empirically well-defined notion, his account makes advances on vaguer versions of the appraisal theory. Not to mention that it is difficult to think of an emotion that does not succumb to this kind of reward and punishment-based analysis.

Insofar as rewards are outcomes that we pursue and punishments are outcomes that we avoid, it is easy to sort emotions into these categories. For example, we pursue the experience of positive emotions, such as joy, and avoid the experience of negative emotions, such as grief. Furthermore, we pursue outcomes that we know are instrumental to these emotions. Knowing that chocolate is instrumental to joy, I pursue chocolate for joy’s sake. If chocolate brought sorrow, I might readily avoid it.7

Despite the benefits of the appraisal approach, one might challenge these theories on the grounds that they ignore the syndrome of sensations to which embodied theories draw our attention. The sensations of muscular tension and sweat, the diastolic discomfort, the shift in body temperature that accompanies many emotional experiences are neglected in most appraisal theories. Of course, the appraisal theorist is not denying that these changes happen. She is merely denying that they are constitutive of the emotions. These changes might be regarded as common effects of emotion. Other theorists argue that the emotions involve both appraisals and representations of bodily state, or in Prinz’s case, that representations of bodily state can be the vehicles of appraisal content (cf. Damasio, 1999; Prinz, 2004).

In any event, it is worth observing how much Prinz’s theory shares with appraisal theories. Similar to Rolls’ theory, Prinz takes the valence of an emotion to consist in the content of the emotion constituting an internal reinforcement. In Prinz’s theory, the occurrence of positive emotions results in positive reinforcement of the actions instrumental in producing those emotions. Perceiving oneself to undergo the syndrome of bodily states associated with joy is itself a positive reinforcer. As mentioned a moment ago, Prinz’s theory also holds that emotions track core relational themes, however they do so not by being judgments, but by

7. One might be reminded of the paradox of tragedy, where audiences appear to be attracted to circumstances instrumental to the experience of negative emotions. However, one must remember that the experience of negative emotions is accompanied by the experience of positive emotions as well. Thus, the negative quality of the sadness experienced while viewing tragedy is a trade-off many are willing to make for the larger reward of the aesthetic experience with which the sorrow is associated.
being representations of bodily state whose activation is locked to the environmental relations that constitute those themes.

When we turn to moral judgments, like the judgment that wealthy nations are morally obligated to assist burdened societies, Prinz can then appeal to the reinforcement capacity of emotions to explain the relevance of body state representations. There are various stories that one could tell concerning the process by which I would come to form the moral judgment that I do. For example, I might empathize with the plight of burdened people. In empathizing, I would undergo the bodily sensations similar to what these people undergo when they attend to their personal misery. These bodily sensations would constitute punishments. Thus, through experiencing them, I would become less likely to engage in activity on which they are contingent, and more likely to engage in activity that would lead to their omission. Among those activities that would lead to their omission is action that improves the living conditions of the burdened. Among those conditions that would lead to their occurrence would be others’ callousness toward the burdened.

If emotions constitute reinforcers, moral judgment could depend on emotional dispositions without every instance of moral judgment involving an emotional experience. Emotions would only be needed to develop positions on moral questions or to alter positions. Thus, in answering moral questions an emotionally uninvolved response would still depend on an agent’s specific emotional reinforcement history.

This concludes our discussion of the content of Prinz’s theory. It should be clear how the bridge is built between perceptions of bodily change, when they constitute reinforcements, and various moral judgments. In the next section, we will consider how Prinz’s theory of emotion might be tested against appraisal theories of emotion in the context of the moral judgment literature in cognitive neuroscience.

4. Testing, testing

To test a theory of emotion against neural data, one must look for neural responses that precede emotional behavior with relative invariance, and that occur when the agent is put in the right circumstances. Ideally, we would find that intervening on specific neural networks would enable us to selectively reproduce or impair emotional reactions (cf. Bickle, Mandik, & Landreth, 2006; Craver, forthcoming). This capacity for precise manipulation would enable us to show that specific neural systems are the mediators between, for example, threat and panicky behavior or between reward and joyful behavior, because these neural systems carry information concerning our bodily states (in the case of embodied theories) or information concerning our interests (in the case of appraisal theories).
So, perceptual embodied theorists should look for mediators (i.e. neural networks) whose activity encodes such bodily states as dilation of the esophagus (as occurs in sadness) or a warmth in one's face (as occurs in embarrassment). It will not do for such a substrate to merely encode the presence of warmth or dilation independently of bodily location. Rather, we should look for conjunctive representations that encode both the occurrence of a somatic property and the bodily location of that property. For example, a substrate that encoded information concerning the painfulness of a stimulus, as reported by the subject, would not carry a body state representation per se unless it encoded where in the body the painful stimulation occurred. If the relevant neural signal does not encode the status of a particular bodily location, it is not a somatosensory signal.

On the other side, appraisal theorists should look for a substrate with cells that encode information concerning states outside of the agent's body. Perhaps this substrate will include somatosensory information as well. Our interests are not confined to what's going on in the outside world, nor are they confined to our bodies. Our interests encompass both domains of experience. A substrate of appraisal should reflect this domain-genericity.

Furthermore, a substrate of appraisal should encode signals that indicate the satisfaction and frustration of one's goals or interests. The neural activity in this substrate should encode both information concerning the thing valued and the value of the thing. That is to say, it should treat rewards differently from punishments.

We now have a rough sketch of somatosensory and appraisal theories of the emotions. And, we have some indication of what to look for in the brain as evidence of their presence. This places us in a position to evaluate Prinz's handling of the functional neuroimaging evidence when he claims it in support of his sentimentality. In the next two sections, I will review the assortment of neuroimaging studies of moral judgment cited in Prinz (2006) and apply our two theories to the results, demonstrating first that the evidence often fails to fit the predictions of Prinz's theory and then later, that the evidence sometimes fits the predictions of an appraisal theory.

5. Tasks of moral judgment

Prinz (2006) cites five studies focused on tasks of moral judgment. These studies show increased activation in brain areas often associated with emotional experi-

---

8. Ideally the network would be relatively localized. But a psychological construct may fail to map onto a neural process in a region that is distinct in histological or morphological terms. What is crucial is that we uncover at least components of the relevant neural mechanisms.
ence. But only two of these studies satisfy the predictions of Prinz’s sentimentalism, given his theory of emotion. In this section, I will review the studies Prinz discusses. However, my review will be limited in certain respects. A full discussion of all of the brain regions activated in these studies would take us too far afield. To limit the number of regions for discussion, I will concentrate on activations the investigators believed to be typical of emotional processing. This will enable us to maintain focus on the argument against Prinz’s handling of the neuroimaging evidence.

Embodied theories of the emotions (of the perceptual variety) predict that there will be activations in parts of the brain involved in somatosensation when subjects undergo emotional responses.9 For example, we should expect activation in the primary somatosensory areas of parietal cortex and the insula, since these parts of the brain carry information concerning bodily states.

Appraisal theories predict that there will be activations in parts of the brain involved in appraisals, i.e. judgments concerning how one’s interests are fairing in the world. Insofar as rewards and punishments constitute objects of value, we should expect reward and punishment tracking parts of the brain to be activated when appraisal-based emotions occur. For example, we might expect orbitofrontal cortex to be activated, since it tracks the status of rewards and punishments. Let’s now consider the studies that Prinz discusses.

We will begin with Berthoz et al. (2002), who investigated the relationship between embarrassment and judgments of intentional and unintentional violations of social norms.10 The task stimuli consisted of sentences telling stories with three different endings and sentences composed of “unrelated words”. Subjects were directed to read the sentences and try to imagine what the protagonist of the story would feel if they were in that situation. The stories were told in either a personal form (where ‘you’ are the protagonist) or impersonal form (e.g., where ‘Joanna’ is the protagonist). Berthoz et al. (2002) provide the following examples of stimulus sentences:

**Personal Stories**

*Beginning.* ‘You are invited for a Japanese dinner at your friend’s house.’

*Endings.* (i) Normal: ‘You have a bite of the first course and like it, and congratulate your friend for her good cooking.’ (ii) Embarrassing: ‘You have a bite

9. This prediction depends on a limited form of localizationism, according to which the sites of activation appearing in a functional image of the brain will be correlated with neural processes responsible for specific psychological processes. Thus, while tactile perception might be carried out in multiple neural regions, we would still expect to find performance on a tactile task to be correlated with activation in some subset of those regions.

10. Many social norms are not regarded as moral norms, of course, which perhaps complicates the interpretation of the results.
of the first course, you choke and spit out the food while you are coughing. (iii) Violation: ‘You have a bite of the first course, but do not like it and spit the food back into your plate.’

**Impersonal Stories**

**Beginning.** ‘Joanna is invited for a Japanese dinner at her friend’s house.’

**Ending.** (i) Normal: ‘She has a bite of the first course and likes it, and congratulates her friend for her good cooking.’ (ii) Embarrassing: ‘She has a bite of the first course, chokes and spits out the food while she is coughing.’ (iii) Violation: ‘She has a bite of the first course, but does not like it and spits the food back on her plate.’

Subjects were asked to rate the stories in terms of how embarrassing they found them. Brain scans taken during readings of normal stories were subtracted from brain scans taken during embarrassing stories, so that the difference in neural activations between task conditions could be assessed. Activations were found in the right medial and superior prefrontal cortex, left orbitofrontal cortex, anterior and middle temporal pole bilaterally, left temporo-parietal junction and occipital cortex. When brain scans taken during the normal readings were subtracted from those taken during readings of the norm violation stories, activations were found in left medial and superior prefrontal cortex, left middle and inferior prefrontal cortex, left orbitofrontal cortex, anterior temporal pole (bilaterally), left temporo-parietal junction, occipital cortex with foci in cuneous and posterior fusiform gyrus, and the brainstem. Out of all of these regions of interest, the authors picked the orbitofrontal cortex and medial prefrontal cortex as the regions associated with emotion, because these regions were found in the norm violation images (post-subtraction) and are commonly activated in emotion studies. Prinz’s theory of emotion must predict that somatosensory regions of the brain would be activated in this study, if he is to regard the study as evidence in favor of his theory. But neither activity in orbitofrontal cortex nor medial prefrontal cortex has been shown to signal the bodily location of somatic states.

In a different study, Greene et al. (2001) investigated processes of cognitive control in moral judgment, where cognitive control refers to a capacity to suppress a prepotent response, and instead offer a response better suited to the details of the current situation. The Stroop task is a classic example of a test of cognitive control. In the Stroop task, subjects are presented with words printed in fonts of different colors. Each of the words presented also names a color. So, for example, the word “red” might be printed in a green font. Subject are challenged to suppress the tendency to read the word as “green” rather than “red”, and some fair better than others.

Green et al. hypothesized that in tasks of moral judgment, cognitive control may play a role when subjects must override an emotional response in order to respond
in accord with a consequentialist moral principle. In their study, Greene et al. posed subjects with the following dilemma: five people stand on one train rail and one person stands on another. A trolley approaching a junction between the two tracks will collide with either group. If you pull a switch, the trolley will be diverted and kill the one person instead of the five. Call this dilemma the “trolley” dilemma. If you do not pull the switch, the five will be killed. What should you do? Most subjects say pull the switch, but when the case is altered slightly, judgments change.

Suppose that instead of pulling a switch, the only way for you to stop the trolley from colliding with the group of five people is to push another person from a footbridge onto the tracks, where the fallen person’s body stops the trolley. Call this dilemma the “footbridge” dilemma. If you do not push the other person onto the tracks, the five will die. What should you do? In response to this latter dilemma, subjects largely deny that one should push the bystander onto the tracks.

As in most functional neuroimaging experiments, a variety of brain regions were activated in task performance. Unlike Berthoz et al.’s study, Greene and his colleagues were unable to acquire reliable data concerning orbitofrontal activation. fMRI is subject to distortion near the sinuses, and that is where orbitofrontal cortex is located. So, while no data was acquired implicating orbitofrontal cortex, the results from the footbridge dilemma found activations in the superior temporal sulcus, the posterior cingulate, and the medial prefrontal cortex. These were taken to be the regions of interest indicating emotional control, none of which are somatosensory parts of the brain. Regions of interest implying cognitive control include the anterior cingulate and dorsolateral prefrontal cortex. These areas were activated in the trolley dilemma.

Aside from these two studies, Prinz (2006) cites an article by Moll et al. (2003) and points to their review of an experiment in which subjects were given auditory presentations of statements with moral and non-moral content. Subjects were then asked to judge whether the sentence was right or wrong. Examples of statements with moral content include the following: “We break the law if necessary”, and “The elderly are useless”. Examples of statements without moral content include: “Stones are made of water”, and “Telephones never ring”. The authors found that the orbitofrontal cortex, frontopolar cortex, anterior temporal and anterior cingu-

11. Special thanks to Adina Roskies for this point.

12. Greene et al. (2001) suggest that the average response to the trolley dilemma (i.e. pull the switch) is motivated by subject’s commitment to a consequentialist rule, while the average response to the footbridge dilemma is motivated by an emotional response. The interpretation of Greene et al.’s results and the design of their experiment are highly controversial issues (cf. Moll, 2005). But, Greene appears to deny that some moral judgments are immediately caused by emotional responses. Of course, this does not affect Prinz’s claim that the development of moral positions depends on emotional responses.
late cortices, the thalamus, midbrain and basal forebrain were activated during judgments of the correctness of moral statements (as opposed to judgments of the correctness of non-moral statements). The authors were not explicit regarding which of these regions they took to be indicative of emotional processing in this particular experiment, but in other portions of their review they suggest that those regions consist of the orbitofrontal cortex and anterior temporal cortex, neither of which are somatosensory parts of the brain.

The three studies we have considered thus far direct us to the following regions of interest: orbitofrontal cortex, posterior cingulate, anterior temporal cortex, and medial prefrontal cortex. All of these areas are regarded as substrates of emotional processing because they are normally activated in tasks that are designed to provoke emotional responses. But, none of these areas of the brain are normally regarded as substrates of somatosensation. Some of these regions may play a role in motor control, and some of these regions may carry information concerning the hedonic aspects of bodily stimulation. But these regions have not been shown to carry information concerning the status of particular body parts, which is what we must demand of a substrate of somatosensation.

Briefly, the orbitofrontal cortex is known to encode information concerning stimuli in visual, auditory, tactile, gustatory, and olfactory modalities (Rolls, 2005). The orbitofrontal cortex is intensely studied by emotion researchers because it is especially responsive to the status of reward and punishment stimuli. Damage to the orbitofrontal cortex can severely impair subject's performance on emotion tasks. So, the orbitofrontal cortex appears necessary for emotion, and particularly important for normal emotional processing.

The posterior cingulate is activated under conditions in which subjects appear to experience joy/amusement and sadness (Britton et al., 2006). But cellular recording studies show that it also encodes information that isn't obviously connected to emotion-related somatosensation, e.g. the spatial orientation of an animal's visual attention (Small et al. 2003). Additional research reveals activation in the posterior cingulate in episodic memory retrieval and pain (Nielsen et al., 2005). Its involvement in pain might seem to suggest that this area underlies somatosensation, but it isn’t clear that the bodily locations of the painful experiences can be read off of posterior cingulate activity. So, the data cannot distinguish between somatosensations of pain and judgments concerning somatosensations processed elsewhere.

Cells in anterior temporal cortex are known to encode information concerning facial expressions. Once again, activation of this region might seem promising for an embodied theory if a region responsive to other people’s facial expressions were also responsive to one’s own facial expressions. This result would be promising, because it might suggest that these cells encode their information in the sub-
ject’s bodily coordinates (e.g. in a sensorimotor map of the subject’s own face). Perhaps anterior temporal cortex does double-duty, tracking changes in one’s own facial musculature, and tracking changes in others’ facial expressions visually. Unfortunately, studies of visual self-recognition have failed to find activation in anterior temporal cortex (Sugiura et al., 2005).

Finally, medial prefrontal cortex is often activated in tasks designed to provoke sympathy, however, the region can be divided into two parts, one of which is implicated in sympathy, the other of which is implicated in social cognition tasks (Saxe, 2006). In the latter case, the medial prefrontal cortex is activated when subjects must keep track of an object of attention held by at least two agents. For example, when subjects are asked to direct their attention to an object, and a human-like image of a face turns its gaze toward the same object, this latter region of the medial prefrontal cortex is activated. However, when subjects are asked to direct attention to an object, and no other agent’s attentional state is involved, this portion of the medial prefrontal cortex is not activated. These results, combined with others, indicate that this region (the dorsal medial prefrontal cortex) is preferentially responsive to signs of mutual interest. Again, there is nothing in these results to suggest the involvement of somatosensory states.

One might wonder which regions of the brain would be activated in these tasks if the embodied emotions were involved. Rather than leaving this question unanswered, I will offer a short list of candidate structures. These are all areas of the brain that are known to be involved in bodily perception.

First, let us consider neural networks that constitute somatotopic maps. Somatotopic maps are regions of the brain that can be parsed such that adjacent areas of the neural region are sensitive to stimuli on adjacent parts of the body. Textbooks in neuroscience often depict somatotopy in the anterior parietal cortex arranged like an oddly proportioned homunculus. In such regions, the functional organization of the brain mirrors the organization of the body. So, for example, measurements from different parts of the parietal cortex can be used to accurately predict which part of the body is being touched by a stimulus. Other parts of the brain, such as the cerebellum, encode somatotopic information concerning the stretch of muscle fibers, which enables the brain to keep track of the limbs’ position in space. These structures are widely separated from each other, suggesting that the substrate of emotional bodily perception may also be widely distributed. Lesion studies suggest that damage to parietal somatosensory cortex impairs subjects in gauging their own and other’s emotional state (Adolphs et al., 2000). In the latter case, patients performed poorly when directed to identify the emotions expressed in pictures of faces.

Somatotopy insures that a region of the brain is involved in sensing bodily changes. But somatotopy may not be a necessary property of a substrate for bodily
perception. Though a region of the brain may not be functionally and spatially organized to mirror bodily organization, single cells distributed haphazardly may individually, or in a group, encode body state information. So, the absence of somatotopy does not entail the absence of somatosensory function.

The insular cortex is an example of a neural region often associated with bodily perception, but sustaining no apparent somatotopy. Much of the evidence associating the insula with somatosensation is indirect, and when placed in the context of emotion studies is controversial. The insula receives input from a variety of sensory sources, including the secondary somatosensory cortex. In humans, the insula is responsive to stimulation of the esophagus, and appears to be involved in the regulation of heart rate (Augustine, 1996; Oppenheimer et al., 1992). These facts are favorable to the idea that the insula might track bodily states constitutive of the feelings we associate with emotions like sadness (in esophageal dilation) or fear (in the case of heart rate). Additional evidence shows that the insula is active when subjects undergo emotions, such as disgust.

These data are all encouraging for the hypothesis that the insula is a component in a mechanism of somatosensory emotion.13

As mentioned earlier, two of the five studies cited by Prinz confirm the prediction that somatosensory emotions co-occur with moral judgment. These studies show activation in the insula. The first of these studies comes from Alan Sanfey’s lab (Sanfey et al., 2003).

In the task, subjects engaged in an economic exchange called the Ultimatum Game. There are two players in this game and two roles that players can assume. One role is the “proposer” and the other the “responder”. The proposer begins with a sum of money that will be split with the responder. The proposer offers a deal for splitting the money, and the responder either accepts or rejects the deal. If the responder rejects the deal, neither proposer nor responder receives any of the money.

Fair offers in the Ultimatum Game consisted of 50/50 splits. Unfair offers involved splits that favored the proposer. When proposers presented responders with unfair offers, responders showed anterior insula activation. Unfair offers leading to a rejection of the deal involved stronger anterior insula activations, which the investigators regarded as an indicator of negative emotions, such as anger.

We might plausibly interpret the results of the Ultimatum Game as the product of fairness judgments, a species of moral judgment. It would therefore seem that Prinz has one piece of evidence that somatosensory states accompany moral

---

13. Wilder Penfield’s results challenge the hypothesis that the insula is a complete neural mechanism of emotion. Penfield found that stimulating the insula in human subjects could lead to taste sensations, oropharyngeal, esophageal, and even gastrointestinal sensations, but subjects who received this stimulation did not undergo emotional change (Saper, 2002).
judgments. Another piece of evidence comes from a study of reactions to the faces of the 2004 presidential candidates. Briefly, Kaplan, Freedman, and Iacoboni (forthcoming) presented subjects with pictures of their preferred candidate and compared that brain scan with the result of presenting subjects with a picture of their non-preferred candidate. For example, Republicans viewing pictures of John Kerry showed increased activation in the insula compared to their viewing of George W. Bush.

In contrast to the Ultimatum Game, the Kaplan et al. study requires fancier footwork to spin as a task of moral judgment. Perhaps judgments of the candidates’ moral character were involved. But subjects in the study were not asked to explain the basis of their preference for one candidate over another. So it isn’t clear that moral concepts guided the judgments, even by the subjects’ lights.

If we exclude the Kaplan et al. study, the results of the Sanfey study are the exception, not the rule. This leads one to wonder whether a sentimentalist explanation of the other three studies is available, or whether sentimentalism finds no support in the neuroimaging literature. In the next section, we will look at an alternative hypothesis according to which moral judgments should co-occur with activation in mechanisms of appraisal.

6. Evidence of appraisal

Suppose that we were to adopt an appraisal theory of the emotions along Rolls’ lines and interpret the data through the lens of that theory. Appraisal theorists hold that emotions are appraisals or judgments of how our interests are faring in the world. In Rolls’ terms, the relevant appraisals concern the status of rewards and punishments. As I’ve already mentioned, the orbitofrontal cortex is regarded as a key component of emotional systems, because of its involvement in reward and punishment processing. Orbitofrontal cortex responds to pleasant touch, pleasant odors, visual predictors of primary reinforcers, monetary rewards, pleasant music, attractive faces, pleasant tastes, and gains in points during game play.

14. Here, I am using the word “interest” in the general sense involving whatever we care about, including our altruistic concerns.
Orbitofrontal regions also respond to a variety of punishments including unpleasant touch, unpleasant odors, visual predictors of primary punishers, monetary losses, musical dissonance, and loss of points in game play. Insofar as rewards and punishments are stimuli we will work to obtain or avoid, and being that the orbitofrontal cortex tracks rewards and punishments, it would appear that the orbitofrontal cortex tracks things that we value. But, it is important to keep in mind that the orbitofrontal cortex does not merely respond to a variety of stimuli. This area of the brain also responds to losses and gains with respect to these stimuli. It carries information concerning the status of things that we care about: whether the music meets our taste, whether the food is palatable, whether we made or lost money.

Furthermore, the cell populations responding to these stimuli appear to cluster according to the valence and abstractness of the stimulus types (Kringelbach & Rolls, 2004). For example, responses to monetary losses and gains cluster in anterior regions, while responses to food rewards and losses cluster in the posterior regions. Lateral areas appear to respond more to punishment than medial areas, which respond more to rewards. So, it would appear that the organization of the orbitofrontal cortex reflects a capacity to discriminate between the statuses of various interests. Monetary interests are treated differently from nutritional interests. Gains are treated differently from losses, and this is reflected in the segregation of orbitofrontal components.

Because of the sheer diversity of reward and punishment types to which the orbitofrontal cortex responds, and because of the syndrome of deficits that attends damage to this area, the data support the localization of appraisal processes to orbitofrontal cortex (cf. Rolls, 2005 for coverage of lesion cases). But the data do not suggest that appraisal processes are confined to the orbitofrontal cortex. Motor and sensorimotor areas of the brain carry reward related information as well. Spikes from cell populations in the posterior parietal cortex predict the amount of food or juice, as well as the relative preference for the food or juice, that monkeys will receive for performing a particular action (cf. Musallam, 2004). But one would expect that these areas of the brain would encode their reward values on the basis of which rewards or punishments were prioritized in prefrontal regions at the time. That job

15. The orbitofrontal cortex is responsive to rewarding and punishing bodily states, such as pleasant and unpleasant touch, which might seem to suggest that this area is a site of somatosensory emotion. But that conclusion would neglect the variety of information encoded here. Bodily state is only a very small sample of the information to which this part of prefrontal cortex is responsive, and functional neuroimaging results do not discriminate between tactile and non-tactile orbitofrontal responses. Furthermore, it is unclear whether any measures from orbitofrontal cortex carry information concerning the bodily locations of stimuli. So we should not regard activation in orbitofrontal cortex as evidence of somatosensory emotions.
may be left to the orbitofrontal cortex. Let’s consider now how to fit the data concerning orbitofrontal cortex with the kinds of tasks reviewed in Section 2.

Given that punishments are states we will work to avoid, and that generally speaking, we would like to prevent violations of the norms we accept, it is plausible to construe violations of our accepted norms as punishments. Given the tendency of punishment activations to cluster laterally, we should predict that lateral areas of the orbitofrontal cortex would be activated when one of an agent’s moral values is violated. In fact, this prediction has been confirmed.

Recall that from among the stimuli that Berthoz et al. (2002) administered, some sentences were identified as intentional violations of social norms and others were identified as unintentional violations of social norms. In the case of an unintentional violation, you spit out your food, because you are choking. In the case of an intentional violation, you spit out your food because you don’t like it. The unintentional violation led to an increase in activation of the lateral orbitofrontal cortex, while the intentional violation led to an even greater increase in activation of the same neural region. These results are encouraging for an appraisal-based interpretation of the data, because we would expect negative emotions (i.e. negative appraisals) to pair with judgments of norm violations, assuming that the norm violated is one we accept.

While the Berthoz study fits an appraisal-based sentimentalism, the Moll study was the only other evidence for orbitofrontal activation. 2 out of 5 studies favor appraisal-based sentimentalism, while 1 favors a somatosensory sentimentalism. It would appear that an appraisal-based sentimentalism is only one up on Prinz’s sentimentalism.

An advocate of appraisal sentimentalism might attempt to explain away the remaining studies. Imaging studies of orbitofrontal cortex are susceptible to false negatives, leaving hope that with the right techniques orbitofrontal activation would be found (Rolls, 2005). But, hope can hardly be summoned as evidence that appraisal processes normally accompany moral judgment. What the advocate of an appraisal sentimentalism needs is evidence of reward and punishment processing elsewhere.

Such evidence can be found in activations of the anterior cingulate, reported in the Greene and Sanfey studies (Rolls, McCabe & Redoute 2007; Schroeder, 2004). The anterior cingulate has been intensely studied by cognitive neuroscientists, and it has been found that a subregion of the anterior cingulate has its activity correlated with rewards and punishments.16 Damage to this portion of the anterior cingulate is also associated with emotional impairment. The upshot is that

---

16. It should be noted that the relevant portion of the anterior cingulate, in this case the pregenual cingulate, is not a somatosensory area of the brain.
all of the studies show activation in areas of the brain associated with reward and punishment processing. So an appraisal theory wins, right?

Perhaps, but a lot depends on the kind of appraisal made. Recall that Prinz advocates an embodied appraisal theory, according to which representations of bodily state encode information regarding core relational themes. While much of the evidence Prinz cites cannot be used to corroborate his sentimentalism, there is evidence that the insula is involved in reward processing. Perhaps sometimes we make moral judgments in the way that Prinz describes and other times we make moral judgments without reference to our bodily states. The emotions composing moral judgments might sometimes be appraisals without somatic content, and on other occasions involve appraisals with somatic content. In the next section, I will offer a sketch of the sort of process that might unify emotion mechanisms, and thereby provide an interpretation of the neuroimaging results that corroborates a more comprehensive sentimentalism—a sentimentalism that accommodates different emotional styles of moral judgment.

7. **Structure of an emotional mechanism**

As I noted in Section 2, Prinz’s view shares important commitments with an appraisal theory like Rolls’. Both theorists regard reinforcement as the key to understanding the valence of an emotional state. Both theorists hold that emotions are attuned to important relationships between their bearer and the environment.

The main difference between a theory like Rolls’ and Prinz’s theory is in the vehicles of emotional content. In Prinz’s view, emotions are always composed of representations of bodily state. In Rolls’ view, emotions are typically composed of representations of states external to the agent. Notably, Rolls’ theory does not rule out there being emotions composed of representations of bodily state. Rather, his theory says that if there are somatosensory emotions, they are not the norm. Rolls’ is a more general theory for allowing emotions to be composed by mental representations in different formats.

Clearly some bodily states feel good and some feel bad, and some states of affairs represented visually also make us feel good and bad. I see a picture of a mutilated face. That was a sight I would work to avoid seeing, because it looks bad. Thus, the sight of the face is a punishment in Rolls’ terms.

---

17. It should also be noted that states of affairs look good or bad, such as the socks I wear with my sandals, which look fine to me and horrible to my sister. Some rewards and punishments may be entirely visually constituted.
My lover cheats. In my jealousy, my body temperature rises and my eyes feel like they’re about to pop out of my head. I’m in an uncomfortable bodily state—one that I would work to avoid. Thus, the bodily state that results from seeing my lover cheat is also a punishment. The bodily state feels bad; the visual stimulus looks bad. Representations in somatosensory and visual formats can both constitute punishments.

For the representations that constitute rewards and punishments to result in reinforcement, an operation on those representations must occur, where a reward is represented as obtaining, or being expected to obtain, or being greater or less than what was expected. The same can be said for punishments. One way of formulating a pluralist position is to identify emotions with these kinds of operations, performed on representations of rewards and punishments regardless of the format.

In some people, or perhaps only under certain conditions, a particular kind of action will be associated with an emotional state, constituted by operations on a somatosensory representation. For example, representations of unfair treatment might trigger bodily processes that when perceptually represented constitute uncomfortable bodily states, and thus punishments. Some representations of this bodily response might form the core of anger, shared by all instances of indignation. A perception of unfair treatment could then come to be judged morally bad simply because it is instrumental in inducing the punishing bodily state we identify with anger. Perhaps this explains Sanfey’s results.

By contrast, the process by which one perceives oneself to have fallen in social stature may involve operations on representations in a visual format, in which case the respective punishment would be constituted visually. Thus, imagining oneself to have done something that could lead to a fall in social stature, such as displaying poor table manners, might be judged morally bad because that action is instrumental to that form of punishment. There are bodily states associated with embarrassment of course, but these bodily states might only be rendered punishing because of their association with the visual representations of the disgust and disappointment we see in the faces of the people we have offended. Perhaps this is the right story to tell about the Berthoz results.

Both of the examples I have given for how we might flesh out a pluralist picture are exceedingly egocentric. But, it is a short step to rendering them less so when we allow the perceived status of other people to constitute rewards and punishments as well. Your pain might serve as a negative reinforcement in my actions, and your pleasure a source of positive reinforcement. The important point is that there is no bar to providing an account of reward and punishment that explains altruistic and empathetic conduct.

18. A multimodal format is probably more likely, but the important point here is that the kinds of representations involved would not be distinctively somatic.
If different neural structures traffic in representations in different formats, some of them unimodal, some multimodal, some maybe even amodal, then we might expect mechanisms of emotion to be multiply instanced in the brain, but unified by the kind of operations performed on their respective representation types. The diverse set of neural activations found in tasks of emotion, including those in the studies reviewed here, might warrant taking this pluralist picture seriously.

Presently, the medium of positive reinforcement is believed to be a phasic or bursting release of dopamine from cells in midbrain structures. One way to start narrowing down the set of neural networks that could constitute mechanisms of emotion might be tracing effective projections into dopaminergic structures. Orbitofrontal cortex offers such projections. Perhaps the insula and anterior cingulate do as well.

8. Conclusion

Given the three theories of emotion I have canvassed, we can formulate three distinct kinds of sentimentalism. The first variety is a “somatosensory sentimentalism” to fit Prinz’s theory. The second variety we might call “appraisal sentimentalism”. This is the view that appraisals of external circumstances determine moral judgments. The third can be called “instrumental sentimentalism” after the instrumental learning effect produced by rewards and punishments. According to instrumental sentimentalism, capacities to represent the status of rewarding and punishing circumstances determine moral judgment. And, rewards and punishments may be constituted by representations of either bodily conditions or external conditions.

More than functional neuroimaging evidence will be required to test these theories (Coltheart, 2005; Landreth & Richardson, 2004). As we all know, correlation is not causation and the sentimentalist is committed to a causal claim about moral judgment. Neuroimaging reveals only correlations. But research efforts, employing other techniques (including the study of pathological cases) are underway. A causal analysis may not be beyond our reach. It will be interesting to know what theory of the emotions best explains the data to come.

These cells are found in the substantia nigra pars compacta and the ventral tegmental area.
References


The phenomenology of alexithymia as a clue to the intentionality of emotion

Ralph D. Ellis

1. Introduction

Alexithymia is a paradoxical form of psychopathology – the inability to feel one’s own emotions, or as Lambie (forthcoming) puts it, to feel them “as emotions” (see also Ellis forthcoming). There can be little debating whether this paradoxical condition is a reality. Sundararajan (personal communication) tells of a case of hers in which a man killed his father without being aware that he was angry. He fully realized intellectually that the evidence showed he must have been angry at his father – but he could not feel the anger.

Alexithymia can shed light on two main interconnecting themes of this volume. First, the designation of psychopathology includes not only a value judgment, but in fact a judgment that there is something wrong with the valuational aspect of a person’s functioning. A clinical diagnosis is in part a valuational judgment of another person's capacity for making valuational judgments. The alexithymic may kill someone “for no good reason,” we say, leaving open the possibility that in other circumstances killing might be justified, or at least “rational.” The alexithymic performs poorly justified actions, and even performs them without feeling the anger that normally might elucidate the valuational motivations for the actions. For this reason, the alexithymic does not even have a capacity for deliberation regarding the actions in question. The main motivation in favor of the action is hidden from any introspective access, and this is considered a defect in valuational judgment. So one of the themes of this volume, well illustrated by alexithymia, is that psychopathology is at least in part a valuational disturbance, and our judgment of it as abnormal is also valuational.

The other interconnected theme is that emotions and motivations themselves, whether pathological or normal, are inherently valuational, in two ways: valuation both causes and is caused by emotion. On the one hand, I may feel angry at a person (or even an object, as when I stump my toe on a chair) because the person or
object got in the way of an action I was trying to perform. In this case, I value acting in that way or I value getting the results of that action, and therefore feel an emotion toward something that either thwarts or facilitates the motivated action. On the other hand, I may also value acting in a certain way because I already felt a certain way in the first place. In this case, the emotion takes the action itself as its primary intentional object, rather than some person or situation that can thwart or facilitate the action.

2. Two kinds of alexithymia

Often, the connections between value and emotion run in both directions. Emotions often may simply consist of the valuing of certain patterns of action that I want to execute – as I discussed in Curious Emotions (Ellis 2005); yet at the same time, at least in many cases, I may want to act in those ways because I felt a certain emotion (boredom, curiosity, interest, a desire to explore, etc.).

This point leads to one of the central lessons that I believe alexithymia can teach us about the valuational intentionality of the emotional life. Emotions and emotional feelings are not merely interoceptions or proprioceptions of internal body states coupled with associated perceptions and perceptual images. Alexithymia shows that Damasio (2003) is wrong if he really means to suggest (as he does claim to mean) that this is the whole story of the intentionality of emotion. In many cases, the alexithymic has a quite vivid interoception, associated with certain perceptual events or images, but still cannot interpret the interoception as an emotion. The person may feel that thinking about sex causes a stomach ache, for example. The alexithymic may indeed have the proprioception/interoception and the associated imagery, but not be able to know what the feeling is about – i.e., not understand its intentionality, or even feel it as an emotion. In a less extreme case, the person may feel the emotion as emotion, but remain extremely confused as to what it is about. This shows that, in addition to interoception and associated perceptual imagery, still another ingredient is needed in order to feel emotions with an indication as to their intentional meaning.

I argued in Curious Emotions that the missing ingredient here is sensorimotor imagery that could indicate what actions the organism wants to perform. As Newton (2000) suggests, I know what I feel by “knowing what I want to do.” This thesis is fundamentally in agreement with Panksepp’s (1998, esp. Chapter 8) view that emotion begins with activation of the brain’s action circuits, rooted in the midbrain and PAG areas, and that the motivation to act is independent of learning through reinforcement of any consummatory drives. It is for this reason that, in Panksepp’s view, the organism can be endogenously motivated to explore its environment,
without this exploratory tendency leading to any consummatory reward or having been previously learned through reinforcement of any more basic drives.

So it may be that emotion requires sensorimotor as well as interoceptive imagery, and that these are entirely different kinds of imagery. Damasio tells us that interoceptive imagery results from signals received through afferent pathways, just like any perception of an external event. Sensorimotor imagery, on the other hand, is the imagery of the action that I intend to perform, and rises to the conscious level by means of the activation of efferent pathways. Jeannerod (1994, 1997) has studied sensorimotor imagery (or action imagery) extensively, and suggests the sharp differences between it and the mere receiving of perceptual or interoceptive sensations.

It is not surprising, in light of this distinction, that there can also be two different kinds of alexithymia. Some alexithymics have intense interoceptive feelings, but do not experience them as emotions or as having any emotional meaning. Others are just the opposite: they are motivated to take action, and do form the vivid imagery of the action they want to perform (as in Sundararajan's patient, who did perform the action of killing his father, and therefore must first have formed the imagery of the action he intended to carry out) – but at the same time the alexithymic may lack the interoceptive imagery that would let him know that he is feeling an emotion. In this second type of case, the emotion is acted upon, but there is not feeling of an emotional qualia. The person lacks a sense of what emotion is being felt, or in an extreme case even that any emotion is being felt.

3. Alexithymic phenomenology

Let's begin with some phenomenology of the experience of alexithymia. In this way, there may be some hope of dispelling some of the mystery around the paradoxical nature of the phenomenon. What can it mean to have an emotion, which presumably is an intentional, mental state (not just a physical event with no intentional meaning) and yet not be mentally aware that one is in any mental state? It may be that we can all identify to some extent with the “alexithymia of everyday life.” We all know what it is like to have an emotion, but not to be currently feeling it to much of an extent (for example, when we are in grief over the death of a loved one, but have temporarily put it out of our mind). And we also know what it is like to have an emotion while denying that we have it. When a spouse asks us whether we are angry, it may be that the most typical response is to deny that we are angry. At the time, we may even believe ourselves not to be angry. So it may be that we can get a better understanding of the paradox of alexithymia by connecting it to our own experience.
I acknowledge that getting at the phenomenology of such a paradoxical phenomenon is difficult. The most obvious problem is that I am not alexithymic myself, and should be cautious in comparing my own experience to that of the true alexithymic. And if I were alexithymic, this in itself would hardly qualify me to do a phenomenological description of my emotional life! In general, we have to be reluctant to put too much stock in any one person’s description of something as murky and seemingly amorphous as emotional intentionality and reflective awareness. Even people who are not alexithymic do not tend to be very good at describing the meaning and intentionality of their affective lives. But ironically, this fact in itself may give us a place to start. We all have something in common with the alexithymic. So why not begin by considering some ways in which certain aspects of normal emotional experience can be suggestive in understanding what it might be like to be alexithymic. I’ll also speculate a little as to the theoretical implications of these similarities.

First, we need a coherent definition of alexithymia. Most psychologists would agree that alexithymia is not usually just a matter of knowing what you feel, yet remaining unable to label it. In asking what alexithymia can teach us about the phenomenology and the neural correlates of emotional consciousness, it is important to remember that alexithymia is more than merely an inability to name or categorize one’s own felt affects. The inability to “talk about” one’s own affects is a metaphor for being unable to directly sense them in consciousness to a full extent, and the analogy from the Greek nomenclature is only approximate; it is not as if the alexithymic actually fully feels the affects in consciousness, but cannot describe or label them. So the question arises: What does it mean to feel things consciously as opposed to having an affective state at a less than fully conscious level? Correlatively, how are unconscious affects possible? In some sense, it must be that the emotion or affect that the alexithymic more or less fails to feel is still there, despite the person’s failure to be very aware of it.

On the other hand, we also don’t want to say that the alexithymic does not even have an emotion or affective state. That would seem to imply that the bodily condition in question does not even rise to the level of an emotion or affect, an intentional and mental phenomenon, and is actually only a physiological condition devoid of intentionality and mentality. This would be too simple. Alexithymia is a fascinating kind of case precisely because the alexithymic seems both to have an emotion, and at the same time to feel that no emotion is present. And of course this dilemma opens up all the same paradoxes that arise anytime we broach any instance of “the unconscious” – yet in clinical and personal experience, unconscious emotional functioning can often be complex, intricate, and enmeshed in cognitive and social phenomena on a level that would make no sense without purposeful, deliberate, thought-out, and intelligent, albeit unconscious functioning.
For this reason, I think there may be as many similarities as differences between normal emotional experience and alexithymia. If I ask myself at some random time “how I feel,” the answer is usually something like: It seems to be going pretty well; or, I feel somewhat frustrated because it isn’t going so well. But what counts as “going well” depends on what my immediate goals are, and the immediate goals depend on longer term goals: but who among us really can specify with any clarity what our ultimate goals in life are, or why we want what we want? Moreover, when I say that it is or is not “going well,” the specific goals in question may not be in my mind at all, and it might be quite a task to say what they are if I were asked. In the case of goals that involve simply maintaining the self-organizational balance of the body, it may be impossible in principle for me to describe them or have much conscious access to them.

Let’s look at some of the similarities we can find between our own experience and that of an alexithymic. I think most of us can easily relate to a situation in which we do know what we feel, but the intensity with which it registers in our consciousness is very low, and out of proportion with its importance for us. There is nothing inherently abnormal about this phenomenon. Suppose someone has died, and you are in grief. You may divert your attention to watching a silly comic movie to relieve the sadness, and during that time, the sadness may still be there, but is being felt at such a low level of intensity that you are hardly aware of it, until the movie is over and you become aware that it was there all along.

Another similarity is that we are often confused about what a feeling is about, even if we do feel it intensely. When my cat Buster disappeared, and it became obvious that he must have been killed, I was intensely grief-stricken. Later I realized that part of what the grief was about was that Buster’s brother Xander had been unusually close to him, since the two of them had survived in the wild together as motherless kittens. Much of the grief that I felt was Xander’s grief. Was the grief “about” Buster’s death, or about Xander’s grief? The more we ask such questions, the more we realize that almost any emotion can be thought of as being about a vast array of intentional contents of which we may not even be aware while we are feeling the emotion. When I have a small run-in with somebody who wants to bully me, how much is the feeling about this specific bully, and how much is about everyone who has ever bullied me, and indeed the very structure of reality in which people are allowed to get away with bullying each other?

A closely related issue is that not only the intensity and the intentional reference of an emotion can be mistaken, but we can also be in the dark as to its aims. Alternatively, we often fail even to recognize that a feeling has aims. We say “he made me angry,” as if anger were merely a state that we are caused to be in, the way a perceptual object causes us to be in a perceptual state. We often ignore the fact that, before we could be angry, we had to first have some aims, which then got
thwarted by the person who stood in our way, resulting in anger. To say that the anger is “at” the person who got in our way is already to assume that there is something the person got in the way of – some ongoing project that has an aim, or a complicated network of aims. And if so, then the aims that got thwarted are at least as much the intentional object of the emotion as is the person who thwarted the aims. Thus pop psychologists – in my view correctly – often advise people that someone else can’t “make” us angry; we make the anger ourselves, or not, depending on the ways in which we choose to deal with obstacles to our aims.

In many cases, when someone bullies us in some trivial way, the incident can trigger a violent emotion, way out of proportion to the importance of the incident, because the incident only symbolizes a difficult and unresolved issue in our lives. The redneck in the big gas-guzzling pickup truck who arrogantly cuts in front of me, showing off the shotgun in his gunrack as he does so, may provoke intense emotion if the incident occurs just after an ultra-conservative like Bush wins an election. The emotion in this case has more to do with the fact that Bush won the election. The traffic incident was trivial.

In fact, the intentional referent of an affective feeling may more often than not function only or primarily as a symbolization. Artists know this. Rather than try to force the viewer to feel X or Y, they offer a fertile system of elements well suited to be used to symbolize a broad range of different feelings that different viewers might be needing to explore at different times. When I listen to Tchaikovsky’s Fifth Symphony at two different times, what it means may be, and in fact typically is, somewhat different on the two occasions. If I try to force it to mean exactly what it did before, the attempt generally fails.

When I was a child, I wanted nothing more than to play the sax, and for several years I practiced six or eight hours a day. Why did I feel this way toward the sax, and what was it I was feeling? The sax must have symbolized a great deal to me – not only the excitement of a new project, the opportunity to exercise skills, an open-ended opportunity to do some really interesting thing that was only a vague notion, and many other things. We often don’t even know and in some instances can never know what the symbolic referent is that is symbolized by what we take to be the intentional object of an emotion or affect. Notice that the same questions could be asked if it were soccer or baseball that aroused my passion rather than the sax; it is not only in artistic contexts that this symbolization function is important for understanding the intentionality of emotions. The intentionality of emotions just is not as simple as that of perception.

So in normal experiencing, not only can (1) the intensity, (2) the intentional reference, and (3) the aim-orientatedness of an emotion all very often be mistaken, but we can also be unclear as to whether, or to what extent, (4) the apparent
intentional referent is really functioning only as an opportunity to symbolize something, and if so, what it is being used to symbolize.

4. Phenomenological distinctions and biological mechanisms

It is often pointed out that many of the same physiological activities are intensified for a wide range of emotions. Heart rate and adrenalin may be increased during sexual excitement, anger, fear, the rapture of experiencing a great artwork, the fulfillment of exercising skills while playing an athletic game, or a hundred other intentional referents. So there is reason to suspect that the mechanism for registering the intensity of an affective condition may be completely different from the mechanism for recognizing its intentional referent or meaning. Many neuropsychologists and philosophers of emotion treat interoception as if it were the be all and end all of feeling emotions. Cut off the interoception, they say, and you have the emotion but not the correlative conscious feeling. On that view, an emotion is just a bodily state, and a feeling is just the consciousness of the corresponding emotion – the registering of it in a higher place in the brain. Of the many problems with this analysis, one is that it depends on the insight that interoception is important in registering the intensity of feeling, but ignores the point that this registering of intensity may not have much to do with how we register the intentional meaning of the experience.

There is a tendency in Western philosophy and psychology to think of consciousness (and also the mind) as a sort of readout screen on which information gets displayed. I.e., consciousness is a passive receiver of information that comes into it, either through the senses, or from the viscera of the body, or from some sort of memory storage facility, or from cortical loops that have formed to further refine and process the information. In all these instances, consciousness is thought of as the end point of a causal chain, and we therefore tend to think of the mechanism that transmits the information to consciousness in terms of afferent nerve pathways – pathways that go inward from the extremities or the viscera, to a central processing unit of some sort. This tendency existed long before the computer metaphor became popular. It existed even before Cartesian dualism; there were dualisms long before Descartes, and perception is much simpler to think about than emotion, so it served as a paradigm for the latter.

But if we don’t know how we feel only through perception and interoception, then how else could we know? Besides afferent nerve pathways, there is also an elaborate efferent system, whose main ultimate purpose is action. (Even the so-called “interneurons” in the brain, which in one sense are neither efferent nor afferent, still play an important role in this efferent system, and thus function for
practical purposes as if they were efferent.) Enactivists, or “embodiment” theorists of consciousness and mind, suggest that conscious and mental operations are a modification of this action system. As Natika Newton (2000) argued in detail in her *Foundations of Understanding*, we know the world by imagining how we could act in relation to it. And we imagine an action by sending the relevant action commands to the premotor, motor, supplementary motor area (SMA), and other such areas of the cortex, and at the same time, according to Jeannerod’s (1994, 1997) neuropsychological studies, we also inhibit the commands. The inhibition of the action command produces the sensorimotor image of ourselves performing the action. A sensorimotor image is different from a perceptual image. We do not form a visual image of our body at a distance, in third person, performing the action. We imagine what it would *feel* like to perform the action.

And yet, this imagining what the action would feel like is not just the imagining of a possible interoception or proprioception. Those would be afferent processes, and would be accounted for by the usual passive-receiving model of the mind. To imagine performing an action requires sending *efferent* signals, just as we would if we were to actually perform the action, while at the same time inhibiting the signals at certain points in the arc of the efferent system – usually, in the vicinity of the motor, premotor or SMA. In some cases, as Anton Lethin (2002, 2004) points out, the signal partly gets to the extremities and is partly inhibited not only by motor cortex, but also by spinal interneurons – and this is why when I hear or imagine a musical melody, my fingers may slightly move to form the sax fingerings for the melody.

Let’s suppose for a moment that this enactive approach to consciousness has some validity. It would fit well with the observation that, before we can be angry “at” somebody, we first had to be engaged in the process of trying to *do* something; it was the thwarting of what we were trying to do that “made” us angry. To the extent that I know what it was I was trying to do, I can give a richer phenomenological description of what I feel when that process gets thwarted. If something is wrong with the efferent action-command system, I may lack understanding of the action component of emotions, and therefore I may know *that* I feel something with a great deal of intensity, because I receive the interoceptive data from my viscera. Yet in this case I don’t know much about *what* I feel, or why, because I don’t know what the feelings are really about – I have only a superficial knowledge of what triggered them.

On the other hand, the efferent system through which I send action commands, as well as the frontal systems that can facilitate inhibition, may be well functioning, yet the afferent interoception from the viscera may be largely blocked due to some neurological lesion or malfunction. In that case, I may be well aware of *what* I feel, but may falsely believe the feeling to be only minimally important,
and to have extremely low intensity. Instead of feeling so angry with my father that I could kill him, I feel only the mildest degree of irritation or antipathy – so inconsequential that the feeling hardly even registers in awareness. I know that the idea of killing him is mildly appealing, but because of the interoceptive malfunction, I don't know that this desire is in fact intense and overwhelming. On the enactivist view, then, there are two components to knowing how I feel: an interoceptive component, which registers visceral sensations, and a sensorimotor component, which sends action commands and enables me to understand the feeling in terms of action imagery.

5. Two alexithymias, again

Louise Sundararajan (2006) reports that, in her clinical practice, she likes to distinguish between two different types of alexithymics. The first type can feel whatever it is they feel, but have trouble interpreting, labeling, or understanding it. The extreme example is the “somatizer,” who interprets emotional feelings as if they had no affective meaning at all, but instead were only physical sensations such as an upset stomach. The second type insists that nothing at all is being felt – not even a feeling of being generally upset, agitated, or physiologically aroused – and yet behaves as motivated by the corresponding emotional processes.

It may be that these two types of alexithymia involve different brain abnormalities. (I don't commit myself to a nature-nurture position here; a brain abnormality can result from behavioral habit as well as the other way around.) In the case where the patient cannot feel the affect at all, or feels it only very mildly, the problem may involve a cutting off of the afferent signals from interoception in some way. In the case where the patient feels something intensely but is not good at interpreting its meaning, what is missing may involve the efferent action command system in one way or another. In this case also, there will be little understanding of the aims of the affective feeling, because on the enactive account, the intentional meaning of any emotion or feeling depends on its aims – what it is trying to do. Natika Newton argues extensively, in fact, that knowing what I feel is essentially a matter of knowing what I want to do.

This second type of alexithymia is probably the one that is primarily addressed by the Rogers-Gendlin “experiencing scale” (see Gendlin 1962/1998, 1963, 1981), in which psychotherapy clients are rated on a scale of one to seven according to how well in touch they are with their feelings, how well they understand them, and how well they are able to deal with them. All three of these affective skills tend to correlate, but Rogers and Gendlin provide different criteria for identifying a person’s position on the scale, depending on various aspects of emotional self-aware-
ness. As the clients move through the therapeutic process, their “success” as judged by independent raters correlates with their score on the Experiencing Scale. Stage one on the scale sounds much like the type of alexithymic who is not aware that there is any emotion present at all, let alone what the emotion is or what it means.

As we progress toward the middle of the scale, what is happening sounds more like the other type of alexithymic – the one who feels something, but cannot understand what the feeling is about.

Rogers and Gendlin insist that good emotional self-awareness is not a matter of being able to accurately attribute what we feel to a rational or causal explanation. I mentioned earlier Sundararajan’s patient who had killed his father. Subsequent discussions led the man to understand that it made perfectly good sense that he must have been angry at his father, and that this was what motivated the killing. But he still could not feel the anger. Emotional self-awareness is not simply a matter of an intellectual understanding that such-and-such must be the causal explanation or intentional correlate of what I feel. When patients speculate as to what they “must be feeling,” Rogers and Gendlin do not rate this kind of locution high on the Experiencing Scale.

Recent research suggests that a good bit of behavior that Gendlin and Rogers would associate with poor emotional self-awareness correlates with prefrontal inactivity or malfunction. People who have trouble with impulse control, for example, have impaired prefrontal activity. This is presumably because the prefrontal lobe is rich in inhibitory neurotransmitters and is richly interconnected with the motor areas where an efferent action command would need to be inhibited in order for an action image to be formed, as opposed to simply physically performing the action. Impulse control means that before actually executing an action, we first imagine the action – form the idea of it by imaging it – and then consider whether to go through with it or not. Presumably, a person with a severe lack of impulse control would not be able to hit a good curve ball; as soon as he entertained the idea of swinging at the pitch, he would actually be swinging at it. However, many people seem to lack certain kinds of impulse control and yet are good at others. Some good athletes, for example, are also drug addicts and are generally un-self-disciplined when off the field. This would seem to suggest that habit formation can affect prefrontal activity in subtle and intricate ways; prefrontal activity is not a unitary, monolithic thing, but involves complex relations with other brain areas.
6. Conclusion

If what has been said so far has any usefulness, then it would seriously call into question one of the most popular approaches to the intentionality of emotion (for example, that of Damasio as well as “cognitivists” such as Ben Ze’ev, 2000, 2002) – which attributes conscious affective feelings almost completely to interoception. On my account, this interceptive process is only half the story, and probably the least important half. By contrast to this “passive receiving of interoception” approach, I would advocate a more “enactive” one.

I distinguish between interoception, which (as Damasio says) is mostly afferent, and sensorimotor action imagery (of the kind Jeannerod discusses), which is mostly efferent. The intentional meaning of emotions and feelings is partly due to the kinds of actions they aim at, not just the “input” that is received from the viscera. I.e., the importance of the input, both visceral and perceptual, is in terms of how it promises either to facilitate or thwart the body’s self-organizational action tendencies – with the aims of the actions being ultimately geared toward maintaining dynamical patterns of activity characterized not only by self-organizational balance but also a suitably high energy level to subserve the non-consummatory needs of the organism, such as play, curiosity, nurturance and empathy. The intentionality of emotion and feeling must be both in terms of interoception and sensorimotor action imagery. The alexithymic may be able to consciously register interoceptive sensations (for example, in “somatization” – the original classic type of alexithymia), but the condition prevents understanding the aims and intentions of the affect, because the enactive component has become blocked or short circuited, so that what the body wants to do is unable to register in action imagery; and this implies that the subject is unable to “focus” on the intentional meaning of the affect, in the sense used by Gendlin and Rogers in their “experiencing scale,” which attempts to assess the extent to which people not only are “in touch” with their feelings, but also understand their intentional meaning.

References


1. Introduction

“Many scientific and philosophical studies of the emotions attempt to restrict their focus to the factual dimensions of emotions alone. This strategy often appears to be based on the assumption that the factual aspects of emotions can be studied in isolation from their more normative and evaluative dimensions. But can emotions really be studied as factual objects alone, or do they involve an ineliminable evaluative dimension? Can fact and value in emotion be separated and studied in isolation? Is there a scientifically or philosophically valid distinction that can be drawn between fact and value in emotion?”

(Charland, L. and Zachar, P. in Call for Papers for present volume).

In this chapter I would like to present a phenomenological viewpoint on the question of emotions and their relation to “consciousness,” or more accurately, to human experience in general. This view will:

a. argue strongly against the separability of “fact” and “value” (or of “cognition” and “emotion”) in human experience, and

b. suggest that the above has major implications for the types of scientific and applied approaches which will be best suited for an adequate psychological understanding of human beings.

I will then go on to briefly discuss some of the epistemological implications of this viewpoint with respect to some important philosophy of science questions, and
also briefly look at how these phenomenological findings may fit in with some recent thinking coming from other approaches to studying consciousness including that of neuroscience.

2. Martin Heidegger’s phenomenological philosophy: knowledge as a founded mode

Phenomenology is an approach to studying consciousness, human experience, our human way of being, or “existence,” developed (primarily) in 20th century continental philosophy. It seeks to describe the *phenomena* of human consciousness, perceptions, or experiences precisely as they are *experienced*. In other words, it is an attempt to suspend our abstract or theoretical knowledge and judgments about the objects of our experience and to try to describe the phenomena of our experience *just as they appear* to us, as it were, as ‘pre-theoretically’ as possible. From carefully examining the form of how these phenomena are experienced by us, as seen through multiple descriptions of such raw data of perception, one then seeks to tease out some of the general features or *structures* common to all perceptions. In some cases these common or essential features may further be seen to form the philosophical *grounding* or ‘the conditions for the possibility of’ any and all such perceptions.

One of the most famous, thought-provoking, and I believe profound phenomenological discussions of the basic structures of human experience is Martin Heidegger’s so-called “Existential Analytic of Dasein” in his magnum opus: *Being and Time* (1962, originally published in 1927). In this work Heidegger describes in wonderful and highly original detail his descriptions of the basic form of how our perceptions tend to be experienced by us. Interestingly, Heidegger himself was not really all that concerned with such questions as ‘how the mind works’ or ‘the basic structures of consciousness.’ While his teacher, Edmund Husserl (1931; 1960) (founder of the later-called “phenomenological movement”), was more interested in these questions directly, Heidegger was interested primarily in “the question of Being.” Nevertheless, he saw Man as the form of Being whereby the question of Being comes to be asked. And so he concluded that to understand the nature of Being one must start with studying the nature of Man. He subsequently refers to Man solely in terms of Man’s unique and particular “mode of Being;” i.e., as “Being-there” (English translation), or as *Dasein*. We needn’t go into the details of that here, but suffice it to say that even though Heidegger’s project was not meant to

stop with a description of the general structures of human experience, his work in that area is extremely useful for our purposes.

In his phenomenology Heidegger argues that human Being (or our unique human sort of Being) differs from the Being of “things” essentially in two main ways, because:

a. human being is essentially caring Being, and
b. human being is essentially temporal Being.

To skip ahead briefly: In Heidegger’s work on the basic structures of human experience some basic or general features of how we experience our “Being-in-the-World” were described. He came to call the set of these features “the Care-structure.” But Care is obviously an emotional term which thus sets emotion right at the heart of our way of Being. So how did he get there?

Some of this goes back to Heidegger’s descriptions of Dasein’s “primordial” experience of the world as a world of “equipment.” Early in Being and Time Heidegger argues that our primordial (first, initial, original) way of experiencing the world and the objects in it is to perceive the latter always as already within a broad context, or “world,” of “equipment.” He famously argued that our primary, pre-reflective experiences of objects are:

a. that objects are experienced as part of a whole world context; i.e., as being “already in a world,” and that
b. they are perceived originally not in terms of their abstracted scientific properties (height, weight, hardness, etc.) but rather in terms of their usefulnesses as equipment – equipment to be used for our various wants, desires, projects, etc. (whether we are reflectively aware at the moment of these goals and values or not).

He thus describes objects as being apprehended initially only in the mode of being “ready-to-hand” which is the way they are experienced prior to their being available to be seen in the alternative mode of appreciating them in their “present-at-hand” qualities. Here readiness-to-hand (Zuhandenheit) (Heidegger, 1962, p. 98) refers to one’s appreciation of objects solely in terms of their immediate, situational, heavily contextualized, equipment-usage possibilities; while presence-at-hand (Vorhandenheit) (p. 103) refers to understanding them in their more decontextualized, general, abstracted, and scientific properties. The latter (present-at-hand) understanding is thus not seen as primary and can only be appreciated secondarily after an act of distancing and abstraction. For example, a hammer is apprehended first as a ’hitting tool’ (e.g., to bang in the nail in front of me) and only secondarily as a ‘long, thin object of a certain size, shape, and color,’ etc. But it is precisely this latter, abstracted viewpoint that yields what we tend to think of as “scientific” knowledge of or about the object.
So why is this important to us? Well, to start: The term “equipment” cannot be evaluatively neutral. For equipment, in this sense is always equipment for something, or for some task. Thus the very notion of equipment implies and presupposes purpose, and intention. But all such purposes and intentions when examined closely reveal a common structure which includes an imagined (anticipated, desired, or projected) better (or alternatively, worse) state of affairs than that which is currently factually present. Of course such evaluative terms as ‘better’ or ‘worse’ obviously imply that values and value-judgments are being made. And whether such valuations are being made explicitly or implicitly, reflectively (consciously) or pre-reflectively (unconsciously), they are nevertheless always present.

Indeed the primordial way in which the world is perceived by us according to Heidegger is as a world of objects which are experienced as originally understood only in such value-laden ways and in such a purposive or desirous context. Furthermore, in his subsequent discussions of human “temporality” (or how we experience and live out our sense of Being-in-time – a topic too big to describe in this paper) he describes our mode of Being as characterized by our continuously projecting ourselves forward in time (imaginatively) toward a variety of anticipated possibilities (p. 308), but all of which are seen as already valued either positively (as desired), negatively (as dreaded), or in conflicted mixtures of both --but never neutrally. This once more implies that the presence of evaluative components is embedded as deeply as is logically possible within the original organization and nature of perception itself.

Where then does this leave the so-called “factual” elements of such perceptions? Heidegger sees the merely present-at-hand features of things, which constitute the objects of cognitive or “scientific” knowledge, as being only a product of our having taken up a particular secondary stance of abstraction and reflection. While this type of appreciation of objects is obviously of great importance to us, it is equally important not to forget that such ‘objects of knowledge’ are never actually experienced in isolation, completely unbiasedly, or neutrally. For the taking up of the secondary stance required to make such scientific descriptions or knowledge possible does not erase our primary, contextualized--usefulness--value-laden, ready-to-hand mode of experiencing these same objects: it merely adds an additional layer or viewpoint to it. In practice in our age and culture, however, the impressive results of using this particular secondary mode often tend to obscure, hide, deny, or lead us to “forget” the nevertheless important and active presence of that primary one.

2. For more on this see Imagination and Its Pathologies (Hersch, 2003b).
Thus, the “factual” findings of scientific knowledge are seen in this view as being something of a set of after-the-fact abstractions, and ones which can never be purely or merely factual either as they always remain rooted in our primary, pre-abstracted, pre-theoretical, pre-scientific mode of experience as well. This is what Heidegger means when he says that cognitive or scientific “knowledge is a founded mode (p. 86).” And this is part of what in his later work he chastises scientific and particularly “technological thinking” for forgetting (Heidegger, 1976; 2001).

In further following the work of Heidegger and later phenomenologists human perception and experience at any given time comes to be seen as occurring always as within, and as already structured in accordance with, a large network of cares and emotional significances or meanings. Heidegger refers to man’s special mode of Being alternatively as Dasein or as Being-in-the-World. He also refers to his description of the more detailed aspects of Being-in-the-World as the Care-structure (Heidegger, 1962, p. 237). These cares and meanings may also be seen to carry certain emotional valences at any particular moment, although these valences fluctuate with time, re-interpretations, and different cares or projects.

Let us jump ahead a bit now in terms of where the philosophical elaboration of this Care-structure winds up and how it may then be applied in the psychological fields.

3. The care-structure and a phenomenological psychological theory

In other work of my own (Hersch, 2003a) I have tried to develop many of the findings of Heidegger’s existential analytic of Dasein, as augmented by the work of such subsequent phenomenologists as Maurice Merleau-Ponty (1962) and Jean-Paul Sartre (1966) for application in the more psychological fields (e.g., psychology, psychiatry, psychotherapy, psychoanalysis). In doing so I have compiled a list of some key features of human existence which need to be dealt with by a phenomenologically-informed psychological theory. The phenomenological-psychological model developed in that work came to be known as the Beams-of-Light-through-Time model. Human experience is likened in it to a beam of light, but one which projects not just through space but through time as well.3 In that work, “six essential features of human experience” have been identified.4 They are described as follows:

a. Relatedness to a World. Human experience is always experienced as taking place within the context of already being in a world.

3. For a detailed discussion of this please see Hersch, 2003a.
b. **Temporality.** Human experience is always experienced as part of an ongoing, necessarily incomplete process of becoming, one which is situated in time, in the full multi-dimensional sense of time where past, present and future co-imply each other as a totality.

c. **Interpretiveness and Perspective.** Human experience (including perception) is always actively interpretive; that is, it is always interpreting and re-interpreting its previously interpreted situations, at a variety of levels of awareness at any given time, and doing so at all times from its own unique spatial, temporal, and above all, meaning-full (care-full) perspective.

d. **Care-fullness.** Human experience is always filled with, and organized in accordance with our cares (or that which we care about). Thus, the presence of some degree of emotion, imagination, and anticipation is an essential feature of every act of human perception.

e. **Embodiment.** All human perception is phenomenologically experienced from within an embodied perspective, that of the lived body.

f. **Being-With-Others.** The world is experienced as essentially social, that is, as peopled with Others. Furthermore, all human experience is necessarily embedded within an interpersonal, cultural, and social-historical context.

4. **Implications**

4.1 The primacy of care

If we look at human experience in terms of the Being-In-The-World approach we can see that this different understanding of our way of Being may require different ways of looking at human psychology. Of particular significance will be the issues of the primacy of the above-mentioned qualities: the primacy of relatedness to a world, temporality, interpretiveness and perspective, care-fullness (including emotion, imagination, and anticipation), embodiment, and Being-with-Others. These are seen here as being inextricable, integral features of each of our perceptions, and also of all of our subsequent reflections and re-interpretations.

When seeking a single simple term to summarize many of these features I have turned to the term “care.” My use of the word care takes on a broader meaning than does Heidegger’s mainly angst-ridden, worrisome “Sorge” (German for care). My usage has more to do with ‘whatever we care about,’ which perhaps even more obviously puts our emotions right at the heart of perception or experience. This is so whether they be positive, negative, or mixed emotions.

Our world(s) are seen then as always organized and structured in terms of our ‘cares,’ or what we care about. Our goals, projects, wants, desires, values, beliefs,
interpretations, understandings, and our attributions of significance and meanings are all reflective of our various cares.

French philosopher Maurice Merleau-Ponty devoted his major work to describing *The Phenomenology of Perception* (1962, originally published in 1945) and used the term *The Primacy of Perception* in the title of a subsequent lecture / paper on this same topic (1947). But it is one’s cares (or “meanings” in Merleau-Ponty’s terms) which serve to organize and structure even perception as it happens. It is the current ‘set of cares at a given moment’ which can be seen to grant the particular, individualized, metaphoric coloring(s) to one’s world at any given time.

It is also the fluidity of our specific set of momentary cares that we see reflected in the fluidity of our perceptions and the continuously reinterpreted quality of the flow of our experience. Perhaps then, it would be most proper to speak here of the *primacy of care* or the primacy of caring rather than just of the primacy of perception, as there is no perception which is not ‘already organized’ by our set of cares, meanings, and significances. That is, in human experience, caring comes first. Humans always care about something or, more accurately, many things. Caring is omnipresent in human experience. This is a fundamental starting point in any psychology which is to be consistent with the phenomenological underpinnings discussed above.

Furthermore, this omnipresence of caring is perhaps the most fundamental feature of our existence which truly distinguishes our peculiarly human mode of Being from that of inanimate objects or things. Things don’t care. As such, methods for dealing with things will be more straightforward. With things we may proceed in a rather ‘factual,’ or fact collecting way and need not concern ourselves with the current status of the cares of such objects of investigation. On the other hand, in the human world, the most central feature in our understanding of any given individual will be found in our ability to appreciate the status of his/her cares, in his/her current situation or life-context.

Descriptions of human beings which lack an appreciation of the emotional context of their cares are thus seen here as being of very limited value. These may be valid in terms of describing certain features of non-living bodies (which are ‘things’ in this sense). But to rely too exclusively on data gathered solely on the largely de-contextualized, third-person, thing-like aspects of human brains and bodies – no matter how carefully these are visualized and measured – will always remain an *inadequate* methodology when it comes to understanding human life. This is because it neglects the crucial caring, temporal, and other experiential

---

5. Color metaphors are used extensively elsewhere in the Beams-of-Light-Through-Time model. Please see Hersch 2003a (especially pages 133-140) for further details.
aspects of life. Indeed, it is precisely the ‘omnipresence and primacy of care’ – which so distinguishes human life – that such an approach seems to miss entirely.

As such, the epistemic value of a ‘scientific neuroscience alone’ will be quite limited, and may be akin to trying to locate a position on a map when given only the longitude or the latitude, but not both. It is not, of course, that longitude is important and latitude is not (or vice versa) but that neither provides an adequate reference in itself. And neither will the most detailed information about the thing-like qualities of our bodies (including the findings of some very sophisticated, impressive and potentially important data being gathered by recent neuroscience) in the absence of a concurrent appreciation of the lived, emotional, evaluative, caring stuff that is human experience.

For example: A description of some biological markers of stress and anxiety, such as increased activation along the hypothalamic-pituitary-adrenal axis accompanied by an increased galvanic skin response, increased gastric acid secretion in the stomach, and perhaps even increased metabolism noted in the locus coeruleus in the brain may all be interesting findings. But they alone will tell us little about whether the person in which they are occurring is terrified because she is remembering having a gun pointed at her in the past, is worried about an upcoming public speech she is to give, or is anxious about the current health of her child. Since we really can’t hope to know such things without the experiencing person’s telling us, our ‘scientific’ data alone lacks information which is critical to understanding the situation as a whole. And as such, its data – no matter how technically sophisticated it may be – is of limited value when it remains so isolated from its experiential context.

Human experience is here seen as being essentially multi-dimensional in terms of its involving the manifold of qualities listed above as being inherent to the general ‘structure of our experience.’ It is also seen as being multi-facetted in terms of its being simultaneously organized, patterned, and structured in accordance with the interplay of a whole gamut of different types of cares. And our temporality further dictates that the multiplicity of interacting and interactional cares which helps to constitute our “life-world” is also continuously changing and developing over time. For these reasons it appears that complex, multi-factorial, field or systems types of models, and ones which include a particular emphasis on experiential data, will be necessary to adequately understand, describe, or deal with human psychology.

4.2 The primacy of emotion

Let us return now to our discussion of such concepts as have been noted to be of central importance to an adequately human psychology. Above, I have already
asserted the primacy and omnipresence of care. When we look at the term *care*, however, we readily see that it is an emotional term. That is, care itself implies that emotion is present. It would be hard to conceive of caring about anything without at least some emotional feeling or significance being inherent to that caring. Thus, when I speak of the primacy of care I am also implying the primacy of emotion. That is, I am speaking of the inextricable presence of emotion (or feelings), as an integral aspect of all human perceptions or experiences as well.

But this is by no means a neutral or non-evaluative sort of emotion. Rather the sort of emotion that is rooted in caring is evaluative at all times. By definition, such caring is the opposite of indifference. Thus to speak of a purely factual, or value-free emotion sounds absurd. It misses the whole point of the *experiential* phenomenon in question. Now certainly one can speak of the neuroanatomical structures and chemicals which are active and describe the biological expressions or supports for such experiences when they occur, but that is not at all the same as discussing the experiences themselves. Even grammatically one cannot describe first-person phenomena appropriately with third-person language. Such first-person statements as “I love…” or “I dread…” simply cannot be captured from the third-person perspective. Third-person statements like “she loves…” or “she dreads…” do not convey either the immediacy or the certainty of the former statements.

Although at first glance this may not seem to be a very radical point, it is indeed quite inconsistent with a variety of psychological viewpoints which, either explicitly or implicitly, maintain that there is a definite separation between the so-called ‘rational’ and ‘emotional’ spheres of the human psyche. Such abstract splits between such terms as ‘reason and emotion,’ ‘cognition and affect’ and other, I believe misguided, dichotomies as those of ‘mind versus body’ and ‘subjectivity versus objectivity’ have proven troublesome to both philosophy and science for many years.

4.3 Reason and emotion

The ‘facts only’ approach of some contemporary research on emotion or on emotion and its relationship to cognition seems to be a contemporary example of a mistaken separation or dichotomization of the notions of “fact” and “value.”

Heidegger’s “Being-In-The-World” paradigm, in contrast, defined the Being of Dasein (Man) as Care, and saw scientific knowing as being merely a “founded” (secondary) mode of isolating and relating to objects in an abstracted way, rather than as the primary way in which we relate to the world. Our ‘cares’ necessarily involve emotions: how we feel about the things, situations, and others of our past, present, and anticipated possible futures, describes our cares. And our cares, as they are lived out in the world, serve to structure our experience. As such it is
argued that from this viewpoint everything in human life is emotional. That is not to say that reason, intellect, or cognition do not occur or are illusions, but it does argue two points about them that have by no means been universally accepted by other psychological theories (or by some theories of emotion). These are:

a. that neither ‘Reason,’ nor a purely receptive or passive cognitive perception, precedes emotion in human experience.

Rather, our emotion-laden cares serve to structure what is noticed/ attended to/ or made foreground, and what is not focused on/ not attended to/ neglected or pushed into the background in our perceptions at any given time. As such, even in our subsequent attempts to deliberately extend and emphasize the more reflective, relatively rational, cognitive, or scientific aspects of our experiences, our perceptions nevertheless remain thoroughly rooted in their intrinsically emotional grounding.

And

b. that human reasoning or cognitive knowing as such can, in principle, never be ‘pure,’ completely uncontaminated by emotion, or entirely isolated from it. And that to imagine that it could be would be to misunderstand the nature and groundedness of this phenomenon.

The first of these statements will not be new or problematic to most psychodynamic, developmental or evolutionary psychological theories but it is quite contradictory to several, mainly ‘Cognitivist’ or ‘Rationalist,’ approaches which tend to think of ‘Reason’ and ‘Emotion’ as radically separate faculties of Mind which can be studied in isolation, or which may be combined in a merely additive manner.

4.4 Fact and value in emotion

The great increase in knowledge and research in Cognitive Science and Neuroscience of late has led many to look at cognition as if it could be primary and pure. Even some Cognitive Psychology writings geared towards psychotherapy make it sound as if emotions are merely secondary phenomena determined by the contents of our thoughts and cognitions (“distorted” though they may be). See for instance David Burns’ statement that “your thought actually creates the emotion” in the first chapter of his enormously popular book Feeling Good: The New Mood Therapy (Burns, 1980, p.12).

Similarly one can sometimes find in the Artificial Intelligence literature talk of emotion as if it were a mere epiphenomenon which could simply be added on at some point to a primary, purely calculative machine. Our culture’s fascination with
the incredibly rapidly-developing field of computers and cybernetics often both reflects and continues this sort of belief. The somewhat elusive, but seemingly plausible, ‘emotion-chip’ addition for robots or computers in science fiction is perhaps the pinnacle of the emotion-as-an-epiphenomenon line of thought.

Of course biologists, and evolutionists in particular, see such an argument as being precisely backwards (e.g., writers like Damasio (1994) & Panksepp (1998) argue strongly for the deep-rootedness of emotion in the evolutionary biology of the brain). While a merely calculative approach may be fine in dealing with computer systems and other machines which are not ‘caring creatures,’ it cannot be carried through into the human sciences without seriously misrepresenting the basic nature of our subject matter. Human experience is always caring and as such even human memory is always emotional. The idea of looking at emotion in an abstracted, purely factual way divorced from experience and denying the evaluative aspects inherent to emotion makes little sense when considered in terms of the primacy of care argument above. The primacy and centrality of emotion and an appreciation of the emotion-laden nature of our perceptions and responses in our lived-world(s) are necessary concepts for a psychological theory which seeks consistency with a paradigm based on such ideas as those of Being-In-The-World or the Beams-of-Light-Through-Time model.

As a clinician this feels like a good requirement to make of our theories. People do not act simply ‘reasonably’ or ‘emotionally.’ They reason in the context of their feelings. Psychoanalyst/philosopher Donna Orange has discussed this more integrated concept of reason and emotion in her 1995 book, appropriately titled: Emotional Understanding. A contextualized understanding with an appreciation of the inextricably mixed nature of the evaluative-factual-emotive components to our existence is crucial.

4.5 Epistemological and methodological implications

In looking at the phenomenology of human experience one never actually finds any facts without values (or for that matter values without any factual-like basis at all --even when errors and misinterpretations dominate, as in delusions). Likewise there is no purely factual perception – rather it is always interpretive, particularly in terms of our cares and our futurality. This, of course, strongly argues against any attempts to describe so-called purely cognitive, purely affective, or purely factual aspects of human experience.

However it is important to recognize that this does not in any way deny the existence, presence, or importance of such philosophically (or ontologically) realist elements at all. It merely argues that in our perception, experiences, or best descriptions of these there will always be present an inherently interpretive,
emotion, care-filled, evaluative quality to them. This amounts to a sort of “uncertainty principle” as applied to the limits of factuality or of any pure, cognitive rationality in the psychological sphere. It is here maintained that such ‘pure cognitions,’ ‘pure knowledge,’ and ‘pure uninterpreted facts’ necessarily fall outside the boundaries of the epistemological horizons of ordinary human experience. It seems reasonable to then also acknowledge these as similarly lying beyond the epistemological boundaries of our theories and methods specific to the study of human psychology. As these abstractions are not humanly accessible (by definition), and may in fact be quite misleading, they are not likely to be useful constructs in our field and would be best abandoned for our purposes.

One question which may be raised at this point is: From where would we even get the idea of the separability of the factual from the evaluative? Certainly not from our day to day experiences. The very idea of their separability comes not from our emotional experiences, but rather from an abstract theory permeated with the idealized and wishful objectivistic assumptions that we can be far more neutral and ‘objective’ than we really are in practice. Nevertheless, it is a theory which is historically well-entrenched in several traditions within modern, western philosophy and science.

The scientific-technological investigative approach which seeks to isolate one particular variable from all others in order to study it on its own has been enormously fruitful in a variety of fields. It works best on matters in which an atomistic view is suitable; i.e., where the close analysis of the smallest parts or constituents of something yield a great deal of understanding about the thing as a whole. However, in the context of dealing with essentially interactional, relational and field-like phenomena replete with emergent properties, the atomistic approach may be severely lacking and inappropriate. I would argue that this is the case when dealing with the topic of human emotions. That is because emotions may be best appreciated as being features (among others) occurring as part of the basic structure of human experience. To attempt to divorce them from that inherently inter-related context is thus to miss much of their essence.

How then is science possible if knowledge is a founded mode and there are no pure uninterpreted facts? I would respond to this by saying that what makes science possible or what makes a scientific pursuit truly scientific is not the presence

---

6. They are indeed abstractions and not true phenomena, for they never actually present themselves as experiences.

7. Even in psychoanalytic theory an overestimation of our possibilities for neutrality and ‘objectivity’ has been problematic, for instance, in the literature on “transference.” For a discussion of this please see Philosophically-Informed Psychotherapy and the Concept of Transference (Hersch, 2006).
of either pure, uninterpreted facts, or of absolute certainties, but rather it is the consistent and rigorous attempt to be true to the phenomena in question. In first developing his phenomenological method, Husserl was trying to “bracket off” or put in abeyance as many of our assumptions as possible precisely to “let the phenomena present themselves…precisely as they present themselves” (or appear to consciousness) (Husserl, 1931; 1956). This seems to me to be closer to the spirit of what it means to be scientific than any attempts to impose external and somewhat arbitrary criteria (e.g., quantitative measurability) on a phenomenon that may not be best approached or understood in that way.

We can probably do our science best in the psychological field precisely by being true to the phenomenon in question (living, caring, temporal, human Being) and acknowledging the care-laden, value-laden nature of even our scientific activities. To deny these, or worse yet, to try to reduce the human phenomena in question to something they are not (uncaring, atemporal, thing-like Being) just because we have easier tools to use with that sort of object, seems to me to be poor science indeed – as such a dubiously scientific approach values its tools and methods more than it respects the phenomena it seeks to understand.

And just as the field of physics survives as a science even after acknowledging the epistemological limits dictated to it by Heisenberg’s uncertainty principle (Heisenberg, 1958) there is still some ‘relative’ objectivity to be had in the study of human psychology as well, but it is a matter of degree and can never be either complete or absolute. This approach is not a mere subjectivism nor is it a radical ontological relativism. But it does wholeheartedly reject a radical objectivism (one which asserts the epistemological accessibility of pure, empirical, uninterpreted facts).

### 4.6 How neuroscience fits in with this

We must keep in mind that the arguments above are not at all opposed to third-person, physicalist, or quantitative studies of brains and bodies per se. Rather the argument is against either relying exclusively on that perspective as being sufficient or adequate to understand human life on its own, or to over-interpret or over-value the findings of that sort of research as if it told us more of ‘the whole story’ than it actually can do. Neuroscience and other ‘factual, scientific pursuits’ can be very important sources of information for us, but they need to preserve their humility as well.

Oddly enough, the phenomenological model described above is actually quite congruent with some recent thinking in the neuroscience literature, and is not without some corroborative support there.
Even though they originate from a distinctly non-phenomenological starting point, neuroscience authors such as Solms and Turnbull, Damasio, Edelman, and Panksepp have come up with thoughts about the nature or structure of human emotions and consciousness which seem to echo the *six essential features of human experience* described by my phenomenological model.

Mark Solms and Oliver Turnbull (2002) are neuroscientists who seek to bridge some of the conceptual gap between *The Brain and the Inner World* as their title says. They describe “the function of consciousness” as “that which let’s you know how you feel.” But they go on to describe this as “…intrinsically evaluative. It imparts value.” (p. 91). This is akin to saying it is “Care-filled” in my model. They go on to describe consciousness as further being “not only what you feel, it is what you feel *about* something,” (p. 92) which sounds a lot like the *intentionality* of consciousness in phenomenological writings.

They then go on to agree with noted neuroscientist Antonio Damasio who describes something like the “relatedness to a world” phenomenon on my list. Damasio states that consciousness “…consists of fluctuating *couplings* of the current state of the self with the current state of the object world” (1999).

The features of “temporality” and “interpretiveness” from my list also find support in the works of Gerald Edelman (e.g., 1989), another major neuroscientist, who states that: “We all automatically reconstruct the reality we perceive from models we have stored in our memories…We adults *project* our expectations…onto the world all the time, and in this way we largely *construct* rather than perceive…the world around us” (quoted in Solms & Turnbull, p.155).

What I’ve called “embodiment” on my list is something that Damasio also very much emphasizes throughout his work. His most popular book is called *Descartes’ Error* (1994) and the main “error” described is that of conceptually splitting off the mind from the body. Damasio’s work, like my own (which gets its phenomenological inspiration in this regard from the works of Maurice Merleau-Ponty), emphasizes the radically embodied nature of all mind, consciousness, rationality, or human experience. He also argues strenuously against the abstract separation of reason from emotion, finding basically that it just doesn’t work that way in real, human life.

Finally, to complete my list, there is some support for the “Being-with-Others” concept to be gleaned from the neuroscience literature as well. The work of Jaak Panksepp (e.g., in *Affective Neuroscience*, 1998) talks about very basic emotional and motivational systems whose origins are thought to reside deep within the structures of the human brain as reflective of early mammalian evolution. But even the most basic of these biological “systems” and “circuits” (as he calls them) are denoted with names like “the RAGE system” and “the PANIC system” (p. 54). In referring to this latter system Panksepp says that “To be a mammal is to be born
socially dependent,” and that this panic has to do primarily with attachment, separation, and the distress caused by the latter. This obviously fits well with phenomenological notions about the Being-With-Others structure to human life which see it as being intrinsically social. Panksepp goes on to describe other “more sophisticated special-purpose socioemotional systems,” giving them obviously interpersonal names like LUST, CARE, and PLAY as well.

While Panksepp’s work is seeking primarily to elicit “basic neural substrates” for such social phenomena as “the social contract (i.e., the possibilities for love and bonding)” (p.54), and is thus not first-person, phenomenological, or experiential at all, this is an example where his third-person scientific approaches can still lead to results which are more corroborative than contradictory to those of phenomenology. Indeed these two perspectives may prove to be complementary to each other and even more valuable when taken together, if done in a mutually respectful manner.

5. Discussion and conclusions

Phenomenology attempts to accurately describe phenomena of human experience precisely in the ways in which they are ordinarily, or initially, experienced. Various phenomenologists have found human experiencing to be a very complex, holistic, interactional, and highly contextualized phenomenon, and one which is inherently evaluative and emotional to its core. Even perception itself is seen to be quite an active process which is continuously being (re-)organized, structured, and influenced at any given moment by our various cares (e.g., wants and fears) and imaginative anticipations of possible future states.

Traditional natural scientific endeavors generally aim at a different mode of experience, which nevertheless remains “founded” or grounded in our initial, everyday ways of experiencing. In this type of scientific work we seek to focus in more specifically on particular items (or variables) and try to study their more isolated but generalizable qualities. To do this we try to tease out as many of the immediate situational / context-dependent details of what we are observing in order to get more of an abstracted sense of these variables, as it were, in a relatively de-contextualized way, or as we imagine they would look ‘on their own.’ We tend to label what is left as their more “factual” elements. Such methods of studying particular, relatively isolated variables have been enormously fruitful in the study of a variety of things. But this is only one approach. And in the case of studying human experience or human emotions it appears that the very act of isolating or of de-situating or de-contextualizing destroys, or at best radically oversimplifies and misrepresents, the very phenomena in question. The methods which were so
useful in our analyses of things are seen to be ill-suited to dealing with the phenomena of human experience. This is especially so when one appreciates the latter as being thoroughly permeated with the presence of the organizing cares and anticipatory temporality that things lack.

This situation is something like that of a man learning to walk on stilts, thus greatly enhancing his abilities insofar as he is involved with tasks requiring him to reach to greater heights. Our typical natural scientific methodologies are analogous to the stilts in this example in that they help us to reach certain goals that would otherwise be inaccessible. We may even get so used to using these tools that we forget or overlook the fact that our knowing how to walk on stilts is thoroughly grounded in our more originary skills of knowing how to stand and how to walk at all (i.e., the skill of walking on stilts is thus a “founded” mode). But when the task at hand is changed from one of reaching to one of jumping or riding a bicycle it becomes clear that using the familiar stilts-tool will no longer be appropriate. Furthermore, in order to learn the new tasks well we will have to return to the grounding skills of standing and walking as a basis for learning these new ones. Similarly if we are to study human emotions in a valid and meaningful manner we will have to leave behind the familiar but stilt-like methods designed primarily for the study of inanimate things. To do justice to these human phenomena will require that further attention be paid to their experiential or phenomenological presentations.

To attempt then to take the evaluative dimension out of emotion or to try to study ‘only the “factual” aspects of emotion’ in such an isolated manner appears to be an example of a misguided research program. The phenomenon in question becomes lost in the manipulation. The abstract concept of ‘emotion without any evaluative component’ is not emotion at all.

This does not mean that there are no factual aspects to emotion or that there is no point at all to studying anything about emotions in traditional scientific ways. But it does suggest that an over-reliance on that approach will not do justice to the phenomena in question. Such methods cannot provide an adequate understanding of human emotion on their own, and they should not be utilized in so abstracted a way that we forget that the phenomena in question are ultimately experiential ones. We should not forget that all claims as to the validity\(^8\) of our findings must ultimately appeal to the court of lived reality (or what Husserl called “the Lebenswelt” or the life-world).

\(^8\) For a detailed discussion of this concept please see Hersch, 2003a, pages 92-115.
References

## Index

<table>
<thead>
<tr>
<th>A</th>
<th>Adaptiveness</th>
<th>7, 37–38, 40, 46–47, 49, 57, 127</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Affectivity</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Anger</td>
<td>7–8, 53–63, 127, 188</td>
</tr>
<tr>
<td></td>
<td>and gender</td>
<td>doubling effect 7, 57–58</td>
</tr>
<tr>
<td></td>
<td>Alexithymia</td>
<td>5, 11–12, 18–187, 189–191</td>
</tr>
<tr>
<td></td>
<td>Appraisal</td>
<td>10, 55, 80–81, 107, 118–120, 128, 140, 160–164, 172–175, 185–186</td>
</tr>
<tr>
<td></td>
<td>Autonomy</td>
<td>8, 89, 90, 93</td>
</tr>
<tr>
<td></td>
<td>Authenticity</td>
<td>emotional, of emotions 4, 10, 7, 41–47, 49, 43, 44</td>
</tr>
<tr>
<td>B</td>
<td>Borderline Personality Disorder (BPD)</td>
<td>5, 7, 53–63</td>
</tr>
<tr>
<td></td>
<td>Being-in-the-world</td>
<td>198, 201</td>
</tr>
<tr>
<td>C</td>
<td>Capacity</td>
<td>decisional 87–88, 90–91, 96, 180</td>
</tr>
<tr>
<td></td>
<td>Care</td>
<td>primacy of 9, 12, 147, 198–200</td>
</tr>
<tr>
<td></td>
<td>Core relational themes</td>
<td>35, 38</td>
</tr>
<tr>
<td></td>
<td>Consciousness</td>
<td>4, 12, 184, 187</td>
</tr>
<tr>
<td></td>
<td>Competence</td>
<td>mental 87–88, 90–91, 96, 180</td>
</tr>
<tr>
<td></td>
<td></td>
<td>see also capacity</td>
</tr>
<tr>
<td></td>
<td>Crichton, A.</td>
<td>7, 15, 19, 22–24, 27–29, 31</td>
</tr>
<tr>
<td>D</td>
<td>Damasio, A.</td>
<td>3, 12, 161, 182, 191, 206</td>
</tr>
<tr>
<td></td>
<td>Depersonalization</td>
<td>10, 140–141</td>
</tr>
<tr>
<td></td>
<td>Depressive disorders</td>
<td>5, 9, 67, 71–72, 75, 80–82, 82, 89–91, 125–126</td>
</tr>
<tr>
<td></td>
<td>Diagnostic and Statistical Manual of Mental Disorders (DSM)</td>
<td>7, 8, 82, 87</td>
</tr>
<tr>
<td>E</td>
<td>Embodiment</td>
<td>4, 10, 107, 198, 206, 160–164, 166, 188</td>
</tr>
<tr>
<td></td>
<td>Emotion(s)</td>
<td>development 13–118, 126–128, 142, 144, 146, 105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dispositions 7, 56–57, 102, 104–106, 111–113, 124–128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>evaluation 140</td>
</tr>
<tr>
<td></td>
<td></td>
<td>intensity 8, 1, 6, 76, 185–187</td>
</tr>
<tr>
<td></td>
<td></td>
<td>intentionality 10, 11, 57, 77–81, 182, 185–187, 106</td>
</tr>
<tr>
<td></td>
<td></td>
<td>moral evaluation 87–89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>plasticity 9, 101–137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>primary and secondary 10, 57, 101, 114, 142</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reasons 37, 40, 42–45, 47–50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>truth 4, 7, 41, 44–45, 48–50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>warrant 4, 7, 35–38, 40–41, 43–45, 49–50, 57, 59–60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>valence 139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Essentialism 145, 150</td>
</tr>
<tr>
<td>F</td>
<td>Fact-value distinction</td>
<td>1–5, 5–6, 12, 65–66, 139, 180–182</td>
</tr>
<tr>
<td></td>
<td>Feminism</td>
<td>2, 146</td>
</tr>
<tr>
<td></td>
<td>Frijda, N.</td>
<td>108</td>
</tr>
<tr>
<td>G</td>
<td>Gender</td>
<td>60–62, 144, 147–148</td>
</tr>
<tr>
<td>H</td>
<td>Heidegger, M.</td>
<td>12, 194–197, 201</td>
</tr>
<tr>
<td></td>
<td>Human nature</td>
<td>9, 10, 141–142, 145–152</td>
</tr>
<tr>
<td></td>
<td>Hume, D.</td>
<td>1–5, 6, 17–19</td>
</tr>
<tr>
<td>I</td>
<td>Izard, C. E.</td>
<td>115, 120–121</td>
</tr>
<tr>
<td></td>
<td>Infrahumanization</td>
<td>10, 141–145</td>
</tr>
<tr>
<td></td>
<td>Innateness</td>
<td>105, 110–111, 115, 121</td>
</tr>
<tr>
<td>J</td>
<td>James, W.</td>
<td>35</td>
</tr>
<tr>
<td>K</td>
<td>Kagan, J.</td>
<td>141</td>
</tr>
<tr>
<td>L</td>
<td>Lazarus, R.</td>
<td>35–36</td>
</tr>
<tr>
<td></td>
<td>Linehan, M.</td>
<td>54, 56</td>
</tr>
<tr>
<td>M</td>
<td>Modules</td>
<td>121–122</td>
</tr>
<tr>
<td></td>
<td>Moral(s)</td>
<td>emotion 7, 54–55, 58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>evaluation and judgment 6, 8, 102–103, 129, 148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>evaluation in psychiatry 87–89, 97–98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>judgment 157–160, 165</td>
</tr>
<tr>
<td></td>
<td></td>
<td>treatment 6, 16–17, 20, 24–28, 30–31</td>
</tr>
<tr>
<td>N</td>
<td>Narcissism 5, 9, 91–94</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Neuroimaging 4, 10–11, 68, 123–124, 157, 168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normativity 4–5, 8, 11, 54, 58–62, 103–104, 107, 109, 139, 140–141, 144, 148, 150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>Pain</td>
<td></td>
</tr>
<tr>
<td>definition 66–67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>emotional (e-pain) 5, 8, 71, 72–82, 82–83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sensory (s-pain) 8, 71, 72–82, 82–83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mental disorder 82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punkepp, J. 2, 182, 206–207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pinel, P. 6, 19, 25–27, 29–31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychopathology 5–6, 15, 27, 29, 31, 181–182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychotherapy 7, 45–47, 57, 62, 13, 189–190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>see also moral treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punishment 158–159, 173, 176</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russell, J. A. 123</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rolls, E. T. 162–165, 175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panksepp, J. 2, 182, 206–207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia 8, 94–96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russell, J. A. 123</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentimentalism 4, 10, 158–160, 174, 177</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reward 162, 164, 173, 176</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rousseau, J. 18–19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporality 74, 195, 198</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia 8, 94–96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>See also moral treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporality 74, 195, 198</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia 8, 94–96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uptake 54–55, 58–60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In the *Consciousness & Emotion Book Series (C&EB)* the following titles have been published thus far or are scheduled for publication:


