

(This page is intentionally left blank)

**On the nature and sources of normativity:
Normativity as grounded in affective human nature**

Søren Wenstøp

June, 2015

NHH – Norwegian School of Economics

For Yujie, Edwin, and Olivia

i. Acknowledgments

I wish to thank all the persons who have contributed in big or small ways to make it possible for me to complete this work. First I wish to thank my advisor, Knut Ims, for his helpful suggestions and guidance. I also wish to give special thanks to those who have provided helpful comments: Bjørn Berdal (general), Magne Dybvig (philosophy), Kjell Grønhaug (social science and management), Ove Jacobsen (business ethics and economics), Sigmund Karterud (psychiatry and neuroscience), Kai Leitemo (economics and psychology), Jaak Panksepp (neuroscience and psychology), and Lars Jacob Pedersen (applied ethics).

Other people I wish to thank for their friendship, intellectual stimulation, and interesting discussions include: Kristian Alm, Bettina Berdal, Simon Blackburn, Peter Booth, Johannes Brinkmann, John Broome, Øyvind Bøhren, Alexander Cappelen, Caroline Ditlev-Simonsen, Andreas Falkenberg, Siri Gullestad, Jørn Halvorsen, Heidi Høivik, Bernadette Hörman, Nigel Iyer, Wanjun Jiang, Svein Johansen, Håvard Koppang, Mark Kriger, Øyvind Kvalnes, Odd Langholm, Xiaohui Liang, Lasse Lien, Eric Lofquist, Alte Midttun, Adekola Oyenuga, Jens Erik Paulsen, Georg Rabl, Jørgen Randers, Daniel Rees, Hannes Reindel, Aksel Rokkan, Joachim Sandberg, Knut Seip, Marcus Selart, Per Anders Sunde, Lars Thue, Sigurd Troye, Bertil Tungodden, Fredrik Wenstøp, Kristian Wenstøp, and Bin Yang.

I would also like to direct a special thanks to my son, Edwin, and my two sweet girls, Yujie and Olivia, who shapes my life as much as they do. Finally, I wish to thank my father, Fred, for his commentary and intellectual stimulation along the way, and my mother, Line, for her motherly and grandmotherly love and care, and innumerable cups of coffee.

ii. Preface

The motivation for writing this dissertation grew out of a dual interest in *understanding* the nature of normativity inherent in the psychological perspective from which ethical decisions are made, and a wish to *improve* ethical decision-making based on this understanding. The first part of this motivation was science-driven, the latter practical with normative ambitions.

Starting out, I was looking at the practical domain of business and strategy. Early on I was intrigued and somewhat provoked by the fact that making strategic decisions normally entailed a moral dimension that sometimes was not explicitly addressed. For example, I saw a glaring contrast between influential management perspectives such as that represented by Michael Porter's famous contention that corporations *ought to fight for every bit of power*, not just with competing firms, but with everyone in the external environment including its own customers, on the one side, and a genuine *concern* and *respect* for other people on the other. The interest in the dual aspects of normativity – the scientific aspect and the normative aspect – culminated in adoption of the research objectives as presented in the thesis.

The business setting provided an arena for reflection around practical moral decisions. Realizing that there were more fundamental issues concerning situated human morality underneath the veneer of business life, however, my attention was soon directed at finding a way to square the descriptive reality of human biology and psychology with the normativity of moral values and moral reasoning more generally. I was particularly interested in where normativity starts, that is, where it comes from. What are its foundations?

The quest for better understanding the normative foundations of ethics took me several places. My first inclination was to look for the normative foundations of ethics in moral philosophy. There I have been influenced by several contemporary figures. In 2004, after reading the works of John Broome, I contacted the author and was kindly invited to The

Swedish Collegium for Advanced Study in the Social Sciences (SCASSS) at the University of Uppsala in Sweden to talk with him in person. Later, I had the opportunity to talk personally with another famous contemporary philosopher, the late Richard Rorty, at The Academy of Management conference in Honolulu in 2005. Finally, I had the opportunity to complete a PhD course held by Simon Blackburn, also a famous contemporary philosopher, at the University of Oslo.

Each of the famous philosophers I encountered made an impression on me, but a very different one. Broome left me with the ambivalently feeling deep admiration and slight disappointment. Rorty left me feeling intrigued but at the same time provoked. Blackburn, meanwhile, left me feeling encouraged yet somewhat perplexed. As a result of my searching quest over the years, I feel that I have gained a reasonably good overview of the main debates about normativity, as well as what it would take to settle them.

Despite several longstanding disagreements that remain to be resolved, moral philosophy and meta-ethics in particular, provided a number of useful insights and ways of conceptualizing ethics. As the current study presumably will bear out, the philosophical positions I defend are in broad strokes in line with the philosophical thinking of Simon Blackburn. However, concluded that, on its own, moral philosophy could not provide a fully satisfactory and definite way of grounding normative ethics.

It became clear that it was vital to base my understanding of normativity also in the sciences. My next inclination was to look into the literature of psychology, especially the psychology of emotion. Rune Lines at NHH gave me many helpful suggestions as to what to read, including important contributions from his own work. When I delved into psychology, I discovered a multitude of theories and positions partly overlapping and partly in conflict, and

sometimes aligned with the positions I knew from philosophy. The overall picture was that of a field rift with conflict.

While psychology provided interesting empirical and theoretical input, I found the literature fragmented and heavily theory reliant, and far less than philosophy, directly concerned with questions of ethics and normativity. Generally, it seemed to focus on processes and relations between conceptual constructs, showing less interest in the content of those processes. Finally, it seemed empirically incomplete in key areas. Every scientific pursuit will of course be empirically incomplete, otherwise there would not be any point in further pursuit, but as several biologists and neuroscientists point out, psychologists frequently do not adequately incorporate knowledge provided by the life sciences.

The area that was to become most important in my quest for understanding ethics was the human brain. I realized that a firm grasp of normative ethics (i.e. what it is, and how it comes to say what it says) has to relate to appropriate facts about what happens inside the brain. Hence, my quest took a turn toward the life sciences, and in particular neuroscience, which deals with the ‘meat of the mind’ head on. My understanding has benefitted immensely from reading works from a number of prominent neuroscientists, and as the thesis will show, I have found the works of Mark Solms and Jaak Panksepp especially compelling. Moreover, I have been fortunate to benefit from the advice of Jaak Panksepp and Sigmund Karterud, as competent neuroscientists, in my literature review of emotions.

In my view, each of the three perspectives – philosophy, psychology, and neuroscience – provides valuable insights. However, it is above all *integrating* these perspectives and seeking their commensurable collaboration that matters. Hence, I decided to undertake the challenging task of putting them into an orderly and unified perspective, with

the aim of understanding the normative foundations of business ethics decisions. The result is this thesis.

Søren Wenstøp,

June 22nd 2015

iii. Table of contents

<i>Acknowledgments</i>	<u>Introduction</u>	iv
<i>Preface</i>	<u>Introduction</u>	v
<i>Chapter 1</i>	<u>Introduction</u>	001
<i>Chapter 2</i>	<u>Theories of emotions and affect</u>	068
<i>Chapter 3</i>	<u>The brain-mind argument</u>	150
<i>Chapter 4</i>	<u>The emotion location argument</u>	187
<i>Chapter 5</i>	<u>The affective nature argument</u>	222
<i>Chapter 6</i>	<u>The emotional decision argument</u>	253
<i>Chapter 7</i>	<u>The dual relativist position</u>	286
<i>Chapter 8</i>	<u>The case against objective values</u>	313
<i>Chapter 9</i>	<u>The emotivist position</u>	336
<i>Chapter 10</i>	<u>Summary and conclusions</u>	404
<i>iii. Appendix</i>	<u>List of tables and figures</u>	425

Chapter 1: Introduction¹

1.1 Opening remarks

The overall aim of this project is to explore the foundations of human morality, or what is often referred to more formally as “ethics”. Consistent with widely accepted terminology in philosophy, I deploy the notion of *normativity* as the central concept of ethics. Thus, every branch of ethics implicitly or explicitly revolves around normativity, including fields that “apply” ethics to specific domains of practice such as business ethics and other forms of professional ethics. Throughout the thesis, I will use business ethics to exemplify applied ethics. I venture to show that understanding normativity has important practical implications, albeit indirect ones. I will not make any clear distinction between “ethics” and “morality” but think of “ethics” more as systematic thought about “morality”, so that ethics relates more to “theory” whereas morality relates more to “practice” and “contextual situatedness”. Thought of this in this manner, I shall be mainly concerned with *morality* and its normative core.

Exploring the foundations of morality presents us with the challenge of making sense of normativity. Why is such sense-making important? Not all agree that it is essential. One alternative option is to simply settle on an *a priori* theory of normativity and proceed “doing ethics” with that theory as given. According to this view, no empirical backing is needed and we can proceed conceptually using the term “normativity”. As I will show, this is common practice in moral philosophy. Another alternative would be simply to let the notion of normativity remain indeterminate, unarticulated, and unexplained. This approach appears to be relatively common in applied ethics. For example a decision-maker could attempt to solve a practical moral dilemma, perhaps by applying a set of pre-articulated “ethical principles”

¹ I would like to give a special thanks to Knut Ims, Kjell Grøhaug, Lars Jacob Pedersen, and Bjørn Berdal for helpful commentary on this chapter.

(principles of justice, fairness, equality, loyalty, openness, etc.) or “codes of conduct”, or else by deploying “judgment” or “intuition”, but without any clear conception of what it means to have solved the dilemma nor what ultimately grounds this solution normatively.

Finally, some think that we do not need to make sense of “normativity” because this notion is ill-conceived, subjective, superfluous, or superficial. Hence, according to this perspective, nothing scientifically meaningful can be made of normativity. This approach is relatively common among various groups of scientists, including many, perhaps most, biologists and neuroscientists. The former two options are overtly “non-foundationalist”, whereas the third option is “eliminativist” about normativity. I argue that each of these options should be rejected. A more detailed discussion is presented below (Ch. 1.2); here I merely highlight a few key points.

Finding each of the three options just described insufficient, I suggest that it is meaningful to aim to establish a scientifically informed understanding of normativity. It seems worthwhile to search for the normative foundations of morality insofar as we recognize that there indeed is some force exerting itself on us in the manner the notion of normativity seem to suggest. This is something that at least deserves to be explored, and if this is a force that exerts itself on us it seems that we need to consider the scientific facts about how we are constituted psychologically and biologically, as well as what this force consists in.

It seems appropriate to approach this type of exploration in a holistic manner that integrates relevant empirical evidence from various theoretical strands. After all, the psychological and biological nature of human beings is multi-faceted, and our relations to the external social and physical world are complex. A foundational approach to normativity is fraught with difficulties and complexities. This may make its pursuit daunting and off-putting. However, these non-foundationalist approaches seem insufficient and unsatisfactory.

One of the reasons why the alternative non-foundationalist options seem less appealing is that they leave empirical questions about normativity unaddressed. This is striking in view of the fact that there *does* seem to be an underlying *psychological perspective* from which normative thought, talk, and action emerge. The eliminative approach, of course, does no better than non-foundationalism because it simply ignores the issue. It remains a legitimate empirical question *how* the emergence of normativity happens. Hence, I argue that people concerned with ethics need to do justice to the complexities of human psychology involved in normative practice. The only reasonable way to do this seems to be to look at human psychology and its biological underpinnings, and what this implies for social interaction. Common moral practices such as issuing normative judgements and following ethical rules, often with non-trivial social consequences, without giving sufficient thought to the nature of these activities, may seem thoughtless and in some cases moralistic. If you like, call the sense that something is amiss the “stickiness” of ethics.

In order to determine if anything can be said scientifically about ethics, the psychological perspective of normativity must be taken seriously. If that is our aim, moreover, we are naturally led to look carefully into the biological underpinnings – the phylogenetic as well as the ontogenetic basis – of human psychology. In the exploration of the foundations of morality, specifically its normative core, I aim to pursue this avenue of inquiry. I will start out by discussing the central notion of “normativity”.

1.2 Normativity

1.2.1 *The philosophy of normativity*

Explicit discussion of the nature of “normativity” has almost exclusively been the enterprise of moral philosophy (e.g., Broome, 2013; 2004; Gibbard, 2012; 1992; Parfit, 2011b; Thomson, 2008; Copp, 2007; 1995; Wedgwood, 2007; Crisp, 2006; Dancy, 2004; Kagan, 1998; Harman, 1997; Korsgaard, 1996; McNaughton, 1991; Brink, 1989), although treatment of the subject is also seen in other areas like sociology (e.g., Durkheim, 1912), economics (Hausman and McPherson, 2006; Broome, 1999), and law (e.g., Klatt, 2008, Kelsen, 1967), and is implied in various approaches to applied ethics. Sociological, economic, and legal approaches to normativity need to establish what underwrites “norms”, “obligations”, and “rights” *qua* normative since these notions are not self-justificatory (Turner, 2010), and the conceptual work of establishing normativity is typically a philosophical undertaking. As a result, the debate surrounding the nature normativity has largely taken place *within* philosophy.

The virtual monopolization by philosophy may strike one as somewhat surprising in view of the fact that normativity has direct practical relevance in our deliberating lives. Following different normative theories will often lead to dramatically different results. Moreover, having a normative outlook – about what is “good”, “right”, “beautiful”, “appropriate”, “the thing to do”, and so on – seems to characterize *every* sentient human being; e.g., the old woman next door, the hair dresser, the high school teacher, the business professional, the politician, and not only adults but also children (Schmidt et al., 2011).

According to Turner (2010: 5), the concept of normativity has gained ground in philosophy in part as a result of “a more or less self-conscious attempt to take back ground lost to social-science explanation”, in which philosophers seek to stake out an area of human

life unamenable to causal explanation. This of course raises the interesting question of whether anything *can* escape causal explanation. In this section I will look briefly at what philosophers write about normativity. I shall suggest that while much of this is meaningful and useful, something remains to be said about normativity in order to appreciate its relation to human nature and human experience.

What is normativity? Etymologically, “normativity” and “normative” stem from the old Latin term *norma*, which refers to a “carpenters square”, a right-angled device used to measure for construction purposes. The original meaning of “normativity”, therefore, relates to something like “standard of correctness” or “measure of correctness”, which has been metaphorically extrapolated into theory of law and ethics. However, it should be pointed out in other languages there are terms that are semantically more or less equivalent to normativity, and some cultures have long philosophical traditions built around these corresponding notions. For example in Chinese philosophy the metaphorical notion *dao* (“the path”) can be seen as roughly equivalent to “normativity”. Undoubtedly examples such as this can be found in other cultures as well. While is a variety of contemporary philosophical definitions of “normativity” their essence is similar. A common feature of these definitions is that they concern the evaluative domain related, at least implicitly, to the *core notions* “good”, “should”, and “ought” (Broome, 2013: Ch. 2).

On a wider conceptualization, normativity concerns such notions as “values”, “norms”, “oughts”, “requirements”, “reasons”, “justifications”, and “rationality” (Robertson, 2009: Ch. 1). One might also include concepts such as “rights” (Rawls, 1993), “duties” (Korsgaard, 1996), and “deserts” (Montague, 1995: 11-24), or still wider take normativity to cover “meaning” (Gibbard, 2012; Whiting, 2007). The common denominator of these notions is that they are *evaluative*. This indicates that normative thought and talk appear to refer to an

evaluative domain of *mental activity*. On this view, it is the evaluative aspect of specific forms of mental activity that makes it appropriate to classify them as normative.

However, a fiercely disputed topic in moral philosophy and epistemology is whether normativity is in fact *internal* to the mind and thus mind-dependent, or *external* and independent of the mind. A central issue in this debate is what role feelings, affects, and emotions play in relation to morality. “The division between those theorists who think feelings are essential to morality and those who think emotions are incidental is perhaps the most central rift in moral philosophy” (Prinz, 2007: 13). Thus, there appears to be two major camps (Goldman, 2009; Falk, 1947-1948); those that believe normativity is internal, i.e. “the internalists”, and those that think it is external, i.e. “the externalists.” These positions transpose into internalist accounts of *reasons of action* (e.g., Hubin, 2001) and their externalist counterparts (e.g., Searle, 2001).

There are different shadings of internalism and externalism; some philosophers, for example, propose that normativity makes its appearance internally in the mind, but depends on and results from a specifiable set of external facts, properties, or principles, so that you can be an internalist about normativity but externalist about its sources. But even on these latter versions *normative authority* is conceived of as external, and that is the main issue. Falk (1947-1948) succinctly summarizes the externalist view as follows:

“They propose, not unnaturally, that when someone ‘ought’ or ‘has a duty’ he is subject to some manner of demand, made on him without regard to his desires; and they imply that this demand issues essentially from outside the agent: that whether made by a deity or society or the ‘situation’ (if this means anything), it has an objective existence of its own depending in no way on anything peculiar to the agent’s psychological constitution” (Falk, 1947-1948: 132).

A crucial question in this debate is whether the “measure of correctness” that morality involves is internal and subjective (e.g., Blackburn, 2010; Goldman, 2009; Hubin, 2001) – as Protagoras said: “man is the measure” – or whether it is external and objective (e.g., Broome, 2013; Parfit, 2011ab; Copp, 2007; Searle, 2001). Typically, moral externalists take cognition, often in the form of beliefs or “intuitions” about the external world, to discern moral correctness in it. The world is, as one might put it, “enchanted” with normativity. On this view, ethics as a philosophical activity is concerned with establishing the *truth* of these beliefs or intuitions, analogously with the quest for empirical truth in science; ultimately, normativity does not depend on the existence of human beings. In contrast, the moral internalists normally take values, sentiments, desires, and emotions – special features of human nature – to be proper source of normativity (Bagnoli, 2011: 62; Blackburn, 1998); it suggests that there “could be a way of doing moral philosophy that started with the ways in which we experience our ethical life” (Williams, 1985: 93).

Where is normativity implicated? Normativity features in everyday decisions, judgments, plans, assessments, and practical reasoning about what to do (Robertson, 2009: Ch. 1). Hence, normativity has a central place in relation to both forward-looking activities such as decision-making, and backward-looking activities such as evaluation and assessment. Normativity is perhaps most obvious in explicit activities that require attention and conscious awareness. I shall henceforth refer to such cases as “explicit normativity”. However, mental activities that do not enter awareness should not be ruled out, because psychological research shows that many “decisions” and evaluations are made either without or before awareness (Berlin, 2011; Bargh and Morsella, 2008). Since decisions and evaluations often are taken to be inherently normative (Hausman and McPherson, 2006: Ch. 13) I shall refer to tacitly made decisions and evaluations as “implicit normativity”. In general, normativity is relevant in every matter that concerns hands-on “practical ethics”.

Practical ethics is involved in “applying” or “implementing” normative theories and perspectives, thereby transforming them into moral practice. This activity implies endorsement of specific normative theories or perspectives to the neglect of others. Therefore it also seems to involve (implicit or explicit) *choice* of theories or perspectives, and also some understanding of – or beliefs about – what this choice involves, although agents may be fully or partially unaware of why they actually make the choices we make and have the perspectives we have. It can be argued that “the distinction between normative ethics and applied ethics does not rest upon any kind of sharp line” (Kagan, 1998: 3).

Moral philosophy makes two central distinctions that serve as demarcation of the theoretical territory of normativity. First, there is the well-known distinction between *descriptive* ethics and *normative* ethics. Descriptive ethics is concerned with the activity of accounting for actual ethical practices, and as thus it branches into disciplines such as social psychology, sociology, and culture studies. Normative ethics, in contrast, is not primarily concerned with actual ethical practices, but with what they could be and above all what they *should* be. Normative ethics asks question such as: What norms are morally valid? What principles should we follow? What standards should we comply with?

Second, moral philosophy makes a distinction between *meta-ethics* and *normative* ethics (Cartwright, 2010: 408; Smith, 2004; Kagan, 1998: 5; Kohlberg, 1984: 277). Normative ethics represents the *first-order* perspectives concerned with questions of “ought” (e.g., what ought to be done) and “good” or “bad” (e.g., what is evaluated as appropriate or inappropriate), sometimes aided by other evaluative notions such as “virtue” and “vice”. This first-order perspective can be taken as a *first-person* perspective insofar as it refers to the perspective of a moral agent (Harman, 1977). Normative ethics can therefore be seen as taking a stand *within* ethics.

Meta-ethics is a *second-order* perspective *about* normative ethics. It essentially concerns the undertaking of understanding, explaining, and justifying normative ethics. In terms of “values” meta-ethics can be viewed as “the philosophical study and examination of the epistemology and ontology of moral values” (Cartwright, 2010: 408). Because of the interconnectedness between normative ethics and meta-ethics, philosophers typically shift between these types of perspective in their analyses and discussions (e.g., Crisp, 2006). In order to understand precisely what is argued in philosophy it is often crucial to have this distinction in mind, although some philosophers retain a sharp analytical distinction while others do not.

Normativity can be conceived of as the central notion of ethics or morality (Brinkmann, 2010: 1; Thomson, 2008; Wedgwood, 2007; Copp, 1995), and it is often taken for granted that moral claims are inherently normative (Copp, 1995: 11). However, the concept of normativity covers *more* ground than just ethics (Finley, 2010), because it also serves a central function in theories of *rationality* and *aesthetics*. In epistemology, as noted, normativity has also been more widely extended to a theory of meaning (Gibbard, 2012). My focus will be on normative as practically oriented; which is to say directed at decision and action.

Since aesthetics is typically not practically oriented, I will consider it outside the scope of my discussion. Aesthetics *may* in some cases be practically oriented, however, in which case I would view it as overlapping with the moral domain. For example, while it is normally morally unproblematic for different people to have different aesthetical taste in music, such difference in taste may under unfavorable circumstances be converted into a moral problem if for example people are to collectively decide what music to play; does the majority have a moral right to always decide on pop music? It can be noted, that moral judgments and aesthetic judgments seem closely analogous (Haidt, 2008: 69), even for very basic aesthetic

experiences such as that of color (Goldman, 2009: 25); the difference between ethics and aesthetics seem to lie essentially in their practical orientation.

In contrast to aesthetics but like ethics, rationality is a practically oriented subject. Rationality concerns what should be done, or the status of what has been done. Whether rationality is genuinely normative or just derivatively so, however, is disputed territory. In a seminal paper Kolodny (2005) proposes that rationality is only *apparently* normative, whereas several other prominent philosophers have presented rationality as normative for our actions and decisions, for example in the form of axioms and requirements to which we are obliged to conform (e.g., Broome, 2005). Relatedly, some philosophers have argued that *logic* is normative (e.g., Field, 2009), whereas others have argued that this type of view mistakenly conflates the internal reasoning from the agents perspective with externally imposed ideals of inference (Harman, 1986). According to Harman (1986: 155) reasoning is subject to the practical constraints of “feasibility” and “practicality” presented by an array of human limitations, whereas logic is not; hence reasoning serves as a normatively *guiding* perspective whereas logic does not.

In my discussion, I will essentially leave the philosophical discussion concerning rationality at side (although I will briefly return to the topic in Ch. 6). For clarity, however, I will outline a tentative position concerning rationality that consists of three points. Firstly, I shall hold that whatever can be said under the heading of “rationality” concerns how normative thinking is *structured*. This will naturally concern such things as following rules of consistency and coherence, but I will not claim that these concerns necessarily have any independent normative force or authority.

Secondly, I will hold that for there to be any point in structuring normativity, there must be first be some *normative content* in place on which to impose this structure. For

example, normative rules are normative in virtue of the concerns to which they correspond; these concerns are their content. It does not seem sensible to hold that rules and principles are normative unless some sort of concern can be cited (Goldman, 2009; Frankfurt, 1988); for why should anyone care about rules unless their breaching, bending, or compliance affects something we care about?

Finally, I take the questions concerning the nature and sources of normativity as relating to its content. Where does this content come from? Hence, issues relating to “rationality” and structure will be seen as secondary to issues concerning normative content. My focus, then, will be on the part of normativity that relates specifically to *ethics*, where the conception of normativity is one where ethics is not conflated with the structural issues of “rationality”. I will now look at normativity more specifically in relation to the various theories of moral philosophy.

1.2.2 Branches of normative thought in moral philosophy

Particularists about ethics (e.g., Dancy, 2004: 1) think that ethics reasoning can be done perfectly well without resorting to rules or abstract principles of thought. “Normativity, according to the particularist, is not a matter of the application of rules, if by rules we mean something articulable, something that is independent of the context” (Dancy, 2004: 198). According to particularist meta-ethics the content of normativity has to be *extracted* in a more direct manner from the particular context at hand, rather than according to decrees from general rules. Dancy (2004) defends a version of particularism that is formulated in terms of what we have “reasons” to do, and how various reasons *contribute* to determining what to do, based on an intuitive ability to judge what has value or have relevant normative “properties”.

On this particular account, the sources of normativity are “out there” in the texture of the external world (the “enchanted” world alluded to in the section above). Particularists are

not committed to this type of externalism about value, however, since the position seems equally compatible with internalism about value (McKeever and Ridge, 2005). The difference here lies in the role of the notion of *truth* in ethics. *Externalists* about value hold that morality tracks “normative truth”, whereas *internalists* deny this or hold that such a notion of normative truth is unnecessary (and moreover indefensible).

Whether there can be any *truth* to normative claims is widely debated in moral philosophy (Prinz, 2007; Smith, 2004; Mackie, 1977). This is an important meta-ethical question to which I will devote some space (especially in Ch. 7 and Ch. 8). After all, we often talk *as if* normative statements are true, as for example when we say “this is good”. The surface of linguistic practices, however, does not imply that there is an underlying fact-of-the-matter that is referred to (Blackburn, 1998; 1993). Alternatively, a fact-of-the-matter might be referred to, and believed in, but still be an *empty* reference because it does not actually refer to anything real (Mackie, 1977). Since beliefs in moral truths (objective values, commands, rights, duties, principles, etc.) can have powerful practical implications, these need to be taken seriously. This is a philosophically but also practically grounded reason for trying to understand what normativity is.

An example of what practical implications normative moral positions might have is this: Imagine an executive who believes that there is a divine command placed upon his shoulders to punish homosexuality (for its disrespect of the divine). Imagine that he feels normatively compelled by this belief to action (he cares strongly for the will of the divine). Now imagine that you are in a job interview with this executive about a job you really want, and he asks you about your attitudes towards homosexuality and your beliefs about the nature of homosexuals. What do you say?

Moral philosophy points to some important questions about our moral practices. Central among these are the questions about what normativity is, and where it comes from. It does matter how we resolve these questions because, one way or another, the result will challenge our moral convictions. Numerous attempts to answer these fundamental questions have been proposed, and below are a few examples to illustrate that point.

Some attempts are oriented around the idea of *properties* (e.g., Copp, 2007; Brink, 1989; Moore, 1903). Copp (2007; 1995: 27), for example, proposes that there “moral properties” (moral realism) and sees normativity as “second-order properties” that spring out of the first order moral ones. A naturalist version moral realism is proposed by Brink (1989: 158), who holds that “moral properties are constituted by, but not identical with, natural properties”. Moore (1903), in an seminal contribution to modern moral philosophy, proposes the non-naturalist idea that normativity could be captured in terms of “non-natural properties” that nevertheless “supervene” on nature; properties that are not empirically detectable and only discernable through “moral intuition” (of some people, himself included).

Others (e.g., Rawls, 1993; 1980; Gauthier, 1986; Glover, 1984) are oriented around the idea of socially *constructing* morality. Rawls (1993) proposes that normatively valid concepts such as “justice” can be arrived at through “social construction” on the basis of a fundamental theory of rights. Similarly, many other constructivist theories of normativity are rationalistic in the sense that they look at, or try to determine, what an ideal rational agent would endorse or judge. Sometimes they can be characterized in terms of what there is “objective reason” to do or judge (e.g., Parfit: 2011b: 426-453). Gauthier (1986) suggests that normativity can be constructed by agreement and the act of consent. Glover (1984) proposes that a subjective thought experiment might roundaboutly settle what sort of people and values there ought to be, and thereby inform us about our shared normativity.

Similarly, Habermas (1984), drawing in part on Kant, argues that normativity arises out of rational *discourse* that takes place under idealized conditions among adequately competent and mutually respecting partakers to that discourse, resulting in principles that are rationally acceptable and binding universally. Hence, agents *co-create* normativity by relating primarily to a *social world* – the discourse community – rather than the physical world as such. Moreover, the *conditions* for this social construction of normativity are themselves supposed to be a product of social construction. According to Habermas (1984) the social context is a “normative context”, and this normative context serves as a normative authority for determining which social relations and doings are morally justified. A normative context, according to Habermas, can only exist once ideal conditions for discourse are met.

Many of the attempted answers above appear incomplete or resting heavily on philosophical assumptions, but that does not mean that the frameworks are necessarily unimportant normatively and practically speaking. It may well be, for instance, that specific constructed moral devices (“institutions”, as they are sometimes called) serve important social functions such as constraining individual behavior (Mackie, 1977: 106). My concern is rather that none of these frameworks seem to serve well as *foundations* for normativity, and do not adequately explain the nature and sources of normativity.

Specifically, they do not properly address the psychological and biological facts about *human nature*; in particular the *emotional* aspect of human life. It is increasingly recognized that human morality is intimately related to emotions (e.g., Gibbard, 2012; Prinz, 2007; Haidt, 2001; Williams, 1973: 207-229), as Hume classically argued. Goldie (2012: 76), for example, argues that a crucial role of emotions is in guiding practical reason; which gives emotions a direct normative function.

If morality is human-centered and inextricably emotional, any account of the foundations of normativity that entirely leaves out emotions seems question begging. Many philosophers now recognize that morality is likely to have evolved out of socio-biological dynamics over time (e.g., Kitcher, 2011: 109; Prinz, 2007), and that these dynamic have made their imprint on culture as well as innate psychology and its biological premises. My inclination is to start the inquiry into the nature and sources of normativity by first investigating what science has to say about human nature (see Ch. 2), in particular from the perspective of the life sciences.

One strand of moral philosophy that *does* take such facts seriously is “expressivism” or “emotivism”. This position suggests that the normative aspects of morality can be explained in terms of psychological and biological notions such as “passions”, “sentiments”, “affects”, “feelings”, and “emotions” (Blackburn, 1998 1-47; D’Arms and Jacobson, 1994; Gibbard, 1992), and that expression and acceptance of norms is a “natural, biological, phenomenon” (Gibbard, 1992: 61). Similarly, Stevenson (1937) argued that an expression such as ‘stealing is wrong’ first and foremost expresses a negative emotional attitude toward theft, but also an encouragement for others to share, heed, and respect this attitude. Hence, on this view, the usage of moral terms reflects not facts in the external world, but internal emotionally based attitudes that serve outwardly directed social functions.

Perhaps most straightforwardly among the philosophical meta-ethical views, expressivism (i.e. emotivism) “reduces the meaning of normative terms to psychology” (Gibbard, 1992: 25-26). More generally, perhaps moral decisions and behavior can be interpreted as reflecting emotional states of mind as well, at least to some extent and under certain conditions. With this perspective in mind, we can now turn to looking at normativity, not as an external notion that is independent of human nature and the subjective experiences of our minds, but as a *psychological* notion.

1.2.3 The psychology of normativity

Gibbard (1992: 24) notes that philosophers often “divorce normative studies from psychology and biology” which “limit their inquiries to what can be said *a priori*” (see e.g., Wallace, 2006; Smith, 2004). This is in many respects unfortunate. For one thing, philosophers have, despite putting in considerable amount of sustained intellectual effort, been unsuccessful in establishing the foundations of ethics on an *a priori* basis (Gibbard, 1992: 25). While some philosophers nevertheless appear to believe that they have succeeded here, they face a considerable amount of pending criticism that must be properly addressed. For another thing, what is taken as *a priori* differs considerably among different accounts of normativity, and thus there is a pervasive lack of unity in philosophical discourse about normativity under the *a priori* conception.

However, some philosophers have a more intimate relationship with the empirical science of human nature. Northoff (2004: 42), for example, sees “philosophy not as an *a priori* propaedeutic or groundwork for science, but as continuous with science”, hence holds there is “no external vantage point, no first philosophy”. By reducing normativity to psychology, the *emotivist* position (detailed in Ch. 9) makes normativity, in the first instance, internal and subjective; a matter of the human mind. This may seem like an attractive starting point for thinking about normativity, insofar as it seems hard to deny that moral views of a given human being must be seen from a *perspective*, and perspectives are inherently *psychological* and *subjective*. Already centuries ago, Hume (1740/1978: 469) recognized the basis of human morality in the internal subjectivity of the mind, arguing that virtue and vice “may be compar’d to sounds, colours, heat and cold, which, according to modern philosophy, are not qualities in the objects but perceptions of the mind”.

Still, many philosophers have been discontent with locating normativity in psychological subjectivity in the manner that emotivists and other internalists about values

propose. Central in this flow of criticism is the implication that accepting subjective psychology comes at the cost of having to abandon the notion of “moral truth” or “moral objectivity” as the core idea of normativity (e.g., Parfit, 2011a; 2011b; Smith, 2004; Nagel, 1986). One might fear that life would be meaningless without moral truth (Rosenberg and Sommers, 2003). Another influential line of criticism is the charge that subjectivity makes emotivism necessarily “viciously circular” (MacIntyre, 1981: 12). However, the two forms of resistance against the subjectivity of normativity can be met by relatively simple arguments.

The most straightforward reply to the first criticism is to agree, but to add that ethics and practical moral life do not actually need the notions of “moral truth” and “objectivity” (Blackburn, 2010); we can get along fine without, and we are equipped to *find* our own subjective meaning in life. Thus although the criticism is *correct*, the consequences turn out to be acceptable. Blackburn (2010) likens the relentless pursuit after moral truth to the search after the “Holy Grail”; basically a search in vain for something that is not there. Emotivists can argue that not only is moral truth and objectivity unnecessary, but they also distract dangerously from what really matters; our concerns, subjective human values, empathic feelings towards one another, biological needs, and so on.

The second criticism above can be met by pointing out that the alleged vicious circularity inherent in the emotivist position is not there. In this case, the analysis of the critics is *mistaken*. Circularity *would* indeed have been problematic in a fundamental way. For circularity to be that case, however, a logical chain of arguments must end up at the same place it started, and so the argument would get us nowhere. Since emotivism has an available explanation of subjective normativity that can be followed down to the specific realities of human affective nature, it *does* have a natural end point. Hence, there is no vicious circularity. The subjectivity of normativity can be seen as part of human nature just as much conscious experience of which it is an instance.

The discussion above suggests that there is room for attempting to offer an explanation of what normativity is, and where it comes from, in an empirically directed and scientifically informed manner. I do not mean that empirical science necessarily can explain everything, but it could certainly be able to explain something. The discussion also suggests that the theoretical perspective on normativity as internally based in the human mind is a viable one. Hence, I propose to look for the sources of normativity inside human nature rather than outside it.

The closest analogies to normativity in scientific discourse appear to be found in the concepts of “affect” and “consciousness”. Similar to the manner in which many argue that normativity is irreducibly subjective, it is widely argued that affect and consciousness are inherently and irreducibly subjective (e.g., Solms and Turnbull, 2002; Chalmers, 2003; 1995). This makes each of these notions at the same time not fully available to scientific scrutiny yet necessary for a holistic theory of human nature. These subjective phenomena impose a dualistic wedge between what can be experienced internally and what can be observed externally.

The idea that there is connection between consciousness and normativity is not new. From a neo-Kantian angle, Korsgaard (1996) proposes that the source of normativity is “self-consciousness”. According to Korsgaard (1996: 46) normative concepts exist in virtue of the fact that human beings are “self-conscious rational animals, capable of reflection about what we ought to believe and do”. According to this view, cold rationality makes us capable of realizing what duties we have; and it is this realization, if it is forthcoming, that makes us human. This is a theory of the sources of normativity that makes no room for emotions.

Haidt (2008; 2001) has presented a convincing and vivid case for dismissing the central importance of rationality in grounding morality, viewing rationality merely as “the

wagging tail of the emotional dog”. Scientific evidence abundantly suggests that much of what goes on in the brain happens outside the relatively narrow focus of conscious attention and awareness, (Berlin, 2011; Bargh and Morsella, 2008), including important decisions (Dijkserhuis, 2004); the very idea of “rational agency” has been suggested to be illusory (Frith, 2007; Wegner, 2002) based on current evidence about how the brain works.

Moreover, deeper levels (subcortically based) affective consciousness, of which we are unaware, may play a central role in shaping our feelings, thoughts, and actions (Panksepp, 2005). This perspective suggests that there are different layers to the meaning of what it is to be a person, and that our sense of agency is only a part of this affectively based personhood. Affective consciousness is a central topic to which I will return (see Ch. 5).

Several moral theorists view empathy as central to human morality. As Decety and Cowell (2014: 337) puts it in quite general terms: “Morality and empathy are fundamental components of human nature across cultures”. The capacity for “empathy” is often taken as a central component of human moral psychology (e.g., Churchland, 2011), especially for explaining *altruistic* attitudes and behavior. While empathy is generally taken to further pro-social behavior and bonding, it is not always a direct route to morality (Decety and Cowell, 2014), and undoubtedly the relation between morality and empathy depends on precisely how the somewhat nebulous and pluralistic term “morality” is construed. It remains clear, however, that the capacity for empathetic feelings provides us with an important dimension along which to classify moral agents.

Introducing the term “empathy” (*emföhlung*), Lipps (1903) proposed it as a special capacity for “inner imitation” of the actions of others. Batson (1991: 58) defines empathy more narrowly as “an other-oriented emotional reaction to seeing someone suffer”. Thus conceived, human suffering and the empathy it evokes places *emotions* at the center stage of

moral psychology. The lack of empathy (e.g., in sociopaths and psychopaths) is frequently linked to what is generally considered morally erratic behavior (e.g., Hare, 1991). The explanatory role of emotional concepts such as empathy exemplifies a close connection between normative psychology and emotions. This suggests that to carefully consider the role of human emotions is a prudent step towards the goal of establishing a theory of normativity. It suggests that, if we want to understand normativity, we must first understand emotions.

Recent evidence suggests that human capacities for *empathy* are hardwired, with its own subcortical neural circuits (Beckes et al., 2013). This study suggests that threats, which generally activate the fear circuits (see Ch. 4), are dealt with in markedly different ways depending on who is threatened. The interesting point is that threats to people who are considered strangers are dealt with differently from threats to people who are considered friends or family. Such threats are dealt with in a manner that is closely similar to a threat to oneself. This suggests that the *care circuitry* (see Ch. 4) instills long-lasting effects in decision-makers. Social bonding, in effect, blurs the distinction between oneself and others.

Having situated normativity in the mind and related it to emotions, to search for the foundations of normativity seems primarily an *empirical* quest, as opposed to a purely philosophical and conceptual one. This means that the disciplinary perspectives on normativity have to be widened considerably. Psychology deals scientifically with issues concerning the subjective mind, albeit mostly in a generalized way. Neuroscience, meanwhile, can provide scientific insights into the chemical, electrochemical, and physical processes that underpin the psychology of the mind. In order to understand what normativity is and where it comes from, I will accordingly proceed by using the term “normative” in a *psychological* sense.

1.2.4 The biology of normativity

Psychology deals with matters of the subjective mind. As a psychological concept, normativity is therefore appropriately considered the domain of psychology, more specifically moral psychology. However, it is widely recognized that the human mind is directly dependent on the biological facts that underpins it (Panksepp and Biven, 2012; Narvaesz, 2008; Solms and Turnbull, 2002), which is to say the human brain. If we accept this point (addressed in more detail in Ch. 3), it follows that psychology, as a discipline, stands in a special dependence relation to factual input from the disciplines that study the brain, principally neuroscience (Panksepp, 2007).

Insofar as normativity is a proper mental phenomenon, biology and neuroscience could contribute by presenting scientific explanations of normativity, at least partial explanations. Early examples of this line of thinking include Brickner (1944) who proposed that *human values* could be given a neuroscientific explanation. Brickner (1944: 225) argued that values are “a product which has originated in the nervous system and has been filtered through the brain”. This presents a biological “bottom-up” and “inside-out” view of the genesis of values in humans, which is not to deny that also cultural and contextual “top-down” “outside-in” factors may play an important role in shaping the values of a given individual. This suggests that human values are constructed in the human brain. The interest in the neurological underpinnings of human values continues after Brickner, with increasing theoretical and methodological sophistication (e.g., Zahn et al., 2009).

Recently, the interests in the connections between moral psychology and neuroscience have been strengthened from more sophisticated knowledge about the human brain combined with an increasing amount of evidence from experimental psychology. Blair et al. (2006: 13) proposes that “there are multiple, partially separable neuro-cognitive architectures that

mediate specific aspects of morality: social convention, care-based morality, disgust-based morality and fairness/just”, and argue that all aspects of human morality involve affect. Greene (2003), meanwhile, proposes that utilitarians and deontologists exhibit different patterns of neuronal firing, and that the latter group tend to make more “emotional” moral judgments while utilitarians draw more upon “cognitive” reasoning.

Seminally, Damasio (1994) proposed that the involvement of emotions is crucial in order to make moral decisions, and researchers have since investigated the role of the prefrontal cortex in moral decision-making in more detail based on clinical research (e.g., Ciaramelli and di Pallegriano, 2011). According to Casebeer and Churchland (2003), current neurological evidence indicates that normative reasoning is at least partially dissociable from general forms of problem solving reasoning. They suggest that unlike other forms of cognitive reasoning, “good moral cognition is shot through with emotion” (Casebeer and Churchland, 2003: 1716).

An understanding of emotions as inherently affective has increasingly gained ground (Panksepp and Biven, 2012; Damasio, 2010; Solms and Turnbull, 2002; Panksepp, 1998), giving rise to a new *paradigm* within neuroscientific research called *affective neuroscience*. Panksepp (1998), who coined the term “affective neuroscience” (Panksepp, 1992), argues that the evidence currently points towards a sub-cortical architecture of basic affective emotions, marking a shift away from cortico-centric and cognitive-centric accounts of emotions. Damasio (2010), emphasizing the sub-cortical brain more than he previously did, now defends a similar as Panksepp concerning the role of emotions. This affective turn was anticipated by affect theory in social psychology, notably developed by Tomkins (1963; 1962). For an extensive review of the literature of emotions, see the next chapter (Ch. 2).

Affective neuroscience combines insights from experimental animal research, clinical research on human patients, as well as experimental human research. Specific neuroscientific imaging techniques including functional magnetic resonance imaging (fMRI), positron emission tomography (PET), magnetoencephalography (MEG), electroencephalography (EEG), facial electromyography (EMG), transcranial magnetic stimulation (TMS), event-related potentials (ERPs), galvanic skin response (GSR), electrocardiograms, electromyograms, endocrinology, and studies of brain lesion patients (Panksepp and Biven, 2012; Damasio, 2010; Cacioppo and Berntson, 2009; Panksepp, 1998; Davidson and Sutton, 1995).

The implications of the paradigmatic shift towards affective neuroscience for moral theory and for understanding normativity are not yet well explored. There are certain indications that some of the basic insights presented by affective neuroscience are starting to gain ground more generally, also in the direction of better understanding normativity. According to Goodenough and Prehn (2004: 1716) “current work applying neuroscience to normative thinking has largely rejected the Kant/Kohlberg conception of normative judgment as properly seated in the realm of affect-free, rational, conscious thought”. Still, however, most of moral philosophy, psychology, as well as neuroscience appear to lean mainly towards cognitive views of emotion (Haidt, 2008; Panksepp, 1998), notably including the perspectives such as those of Kant and Kohlberg.

Evidence from neuroscience, as I venture to show, provides important but limited insights concerning human morality and normativity. Some of the philosophical problems concerning normativity cannot be directly addressed by means of neuroscientific techniques alone. However, sometimes neuroscience can provide important resources even for addressing philosophical conundrums outside its disciplinary grasp. For example, while neuroscience may not be able to access the *experiential qualia* of normativity, since this is inherently

subjective, it may be able to illuminate the circumstances that are objectively necessary for normative subjective experience to take place. This scientific endeavor would not involve inference of normative conclusions (“ought”) from descriptive premises (“is”) – as David Hume famously discouraged and Gerald Moore echoed centuries later by formulating his “naturalistic fallacy”. Instead, scientific endeavor involves saying something *about* the nature of normative perspectives from *outside* them. This is important because it has become increasingly clear that subjective experience must be included in any scientific explanation of how the mind works (Modell, 2003).

We can grant that scientists cannot completely escape their own subjectivity, and that science as it is conducted by humans is never absolutely “value free”; my point is that normativity as a phenomena can still perfectly well serve as an object of scientific enquiry, even though its subjective content is not directly accessible to scientific methods. The widely held contention that normativity is “naturalistically problematic” (Christensen, 2012: 104), accordingly, needs qualification. It may not be problematic to look for descriptive underpinnings of normativity as sometimes suggested; while normative perspectives issues in “ought”, the activity of describing and explaining them does not. Hume’s sensible warnings about inferring “is” from “ought” are intact; they are simply not trespassed against as sometimes alleged.

However, there is another issue. As Turner (2010: 26) notes, the “problems of normativity parallels many other problems in philosophy involving dualisms and exotic objects. The mind-body dualism, with its focus on the irreducibility of consciousness and *qualia*, for example, is similar in structure”. Crucially, normativity seems to depend on *subjective consciousness*. This dependency brings up thorny questions, such as the question of how *subjective conscious experience* can arise out of mere physical reality, appropriately dubbed “the hard problem” by Chalmers (1995). However, these problems are not particular

to explaining normativity, since they crop up in the explanation of every aspect of the mind that is subjectively experienced. However, it may be the case – and I shall later suggest it is – that normativity is a fundamental characteristic of consciousness. This is a topic to which I will return (see Ch. 3 and Ch. 5).

Although the supposed logical obstacle presented by the is-ought-gap to scientific investigation of human “morality” or “normativity” appears to be invalid, the scientific inquiry into the nature and causes of normativity are fraught with many kinds of difficulties and unknowns. As noted, some central difficulties dovetail with those surrounding consciousness. Attempts to locate both morality and consciousness in the human brain have been speculatively pursued by various biologists, psychologists, and philosophers over the centuries since the enlightenment. However, the perspectives on morality represented by enlightenment thinkers have not “failed” as MacIntyre (1981: Ch. 5) proclaims; instead they have needed time to mature. In particular, scientific progress in neuroscience over the last few decades has provided a substitution of empirical facts for pure speculation. This suggests a certain urgency regarding the task of finding ways of conceptualizing the emerging wealth of facts, and ways of integrating theory as it develops.

1.2.5 Animal nature and morality

It is not uncommon to think of morality as unique and special to the human species. This belief, as far as it goes, can provide reason to think that morality is due to our unique biological constitution. In turn, this leads us to see morality partly as a result of the evolution of our species. Since human beings interact with their social and cultural surroundings, however, these factors will also play their part in shaping the morality of any given individual. But it seems clear that the biology evolution has endowed us with is basic prerequisites for all our socio-cultural activities.

When we consider the evolution of human morality, we may also be lead to recognizing that it is *not* obviously the case that all of our morality is particular to humans. After all, we have only been a separate species from the other primates for roughly six million years; which is a remarkably short time-span in evolutionary terms. As Darwin (1906: 193) keenly reminded us: “The difference in mind between man and the higher animals, great as it is, certainly is one of degree and not kind”.

Given the abundant evidence, it seems undeniable that non-human primates, especially the greater primates such as the chimpanzees, have sophisticated interactive moral lives (de Waal, 2009; 1996). This indicates that a basic psychological sense of normativity is unlikely to be uniquely human; hence also unlikely to result from those mental capabilities that are special to humans. One feature of morality that we appear to have in common with the other primates is our strong bonding and caring relations within family, group, and kin. It does not stop there; *every* basic emotional pattern of behavior appears to be remarkably similar between humans and other primates (Panksepp, 1998; MacLean, 1990). As de Waal (1996) shows, other primates engage in empathic behavior, reconciliation after aggression, and reciprocal fairness behaviors; this involves the interplay of basic emotions such as care, joy, sadness, and anger.

Based on experimental research on animals, there is a case for arguing that negatively or positively valenced feelings always accompany emotions (Panksepp, 1998). If this is correct, it makes emotions conveyers of “good” and “bad” experiences; the most plausible explanation of which is that emotions *feel* good or bad (Panksepp, 2011). Prompted by emotional reactions, good and bad experiences can, in turn, be attributed to perceived objects in the external world (or imagined objects in our inner mental world), thus causing them to be perceived *as* good or bad. This suggests the possibly of an explanation for how emotions can

account for, or at least contribute to, our moral perceptions or intuitions about “good” and “bad”.

According to de Waal (2009: 7) it is reasonable to think that “the building blocks of morality are evolutionarily ancient”. Panksepp (2011; 1998: Ch. 3), meanwhile, presents abundant evidence showing that there are evolutionarily ancient emotional systems in the subcortical brain, which humans appear to share with all other mammals. Emotional display and interplay plays a central in the moral behavior that de Waal describes, and this occurs in humans as well as other primates. There are certain differences in the morality of men and women, which are also documented by neuroscientists (Pfaff, 2011: Ch. 6), such as men being more prone to aggression than women are. However, I will ignore them for simplicity.

Instead I will focus on the morality that is shared among healthy human beings, and its various practical implications. Most of all I shall focus on understanding and explain human morality. It has emerged from the discussion above that, in order to explain human morality, Darwin represents an important lead in recognizing our animal nature and its evolutionary origins. Without this perspective we will be out of touch with the current state human nature and how we got to be this way.

1.2.6 An encouraging suggestion

Some prominent contemporary philosophers take a bleak view of the current state of human morality and the role of emotions as fundamental. Identifying *emotivism* and the “emotivist culture” of modern society as the central target of criticism, MacIntyre (1981) proposes that the current state of morality can be described by his “disquieting suggestion”. To illustrate the point, he first he invites us to imagine the occurrence of a disorderly state of science (in some respects uncannily similar to the ongoing situation in northern Nigeria under siege by a religious militia called Boko Haram):

“Imagine that the natural sciences were to suffer the effects of a catastrophe. A series of environmental disasters are blamed by the general public on the scientists. Widespread riots occur, laboratories are burnt down, physicists are lynched, books and instruments are destroyed. Finally a Know-Nothing political movement takes power and successfully abolishes science teaching in schools and universities, imprisoning and executing the remaining scientists.” *After Virtue*, MacIntyre (1981: 1).

MacIntyre uses this story as an analogy for what has happened to morality since its alleged heydays in ancient Greece, exemplified especially by the Aristotelian moral views and ways of life. According to MacIntyre (1981) modern morality is in a state of disorder and confusion, and he conceptualizes this by presenting the image of the “disquieting suggestion” as means of explanation:

“The hypothesis which I wish to advance is that in the actual world which we inhabit the language of morality is in the same state of grave disorder as the language of natural science in the imaginary world which I described. What we possess, if this view is true, are the fragments of a conceptual scheme, parts which now lack those contexts from which their significance derived.” *After Virtue*, MacIntyre (1981: 2).

Over the following four chapters of his book, MacIntyre makes it clear that he suspects the influence of the enlightenment and subsequently emotivist theory as the main culprits of “corrupting” and “distorting” morality. In effect, he argues that these intellectually driven influences aimed at analyzing, understanding, and explaining morality causes impotence in the application of moral concepts. His key point seems to be that emotivism and enlightenment thinkers lack a sociological account of morality as “socially embodied” (MacIntyre, 1981: Ch. 3).

The objective correctness that MacIntyre (1981) believes his particular narrative possesses, however, can be questioned. But some of the key elements of his story are undoubtedly true. For example, it seems correct to suppose that the linguistic concepts inherited from the past are connected with new social contexts, and that they thereby lose part of their original meaning. However, the idea that there was a morally superior fit between moral concepts of the past as they were applied to *their* social contexts, compared to the application of moral concepts today in *our* social context, is speculative. How much do we really know about the way of life as it was lived and experienced in ancient Greece? How glorious were the moral lives for the individuals living them? How certain can we be that morality has gone downhill since? Indeed, one of MacIntyre's (1981) own key points is that we now only understand fragments of that past.

In Chapter 9 I detail what emotivism is committed to, and furthermore show that MacIntyre's specific charge that emotivism is circular and has to be rejected fails. Here, I return to the primacy MacIntyre (1981: 22) places on morality as "socially embodied", which he claims make a sociological approach necessary for moral philosophy. By the manner in which he uses this notion, he seems to take as a premise for moral philosophy that morality can only properly be seen as inherent to a social context. I wish to challenge this view, by presenting an alternative. Following the emphasis on biology and psychology in the sections above with, I propose that in the *first order* morality is literally embodied biologically and that only in the *second order* morality is embedded socially. Hence, I shift the focus from the external context to the internal mental and biological features of the individual in order to explain morality.

MacIntyre's "disquieting suggestion" can now be appropriately paraphrased. I propose that the central problem of contemporary morality is *not* that its moral language is in disorder, and *neither* is the problem that people adhere to an emotivist mental model (if this were the

case). Rather, the problem with morality as it is enacted personally and professionally is that its mental models of the *role of emotions* are to a disquieting extent distorted and ignored. The conceptual problem – which undoubtedly can cause practical problems – is the extent to which our personal and professional moral lives are *estranged* from the emotional animal nature at the heart of our experiencing selves. While we appear to be good at portraying ourselves as objects, we seem prone to ignore ourselves as emotional subject. Ironically, therefore, if my account of the problem is correct, the emphasis that emotivism places on emotions as a basis for moral talk, thought, and action is in *deficit*. Over the remaining chapters of the thesis I explore this line of thought in more detail.

Several prominent philosophers have been interested in human nature and drawn links between morality and emotions, including Hume, Nietzsche, and Bergson. Bergson (1935: 82), for example, considers a biological understanding of emotions central to morality and summarizes his views by writing: “let us say in conclusion that all morality, be it pressure or aspirations, is in essence biological”. With increasing scientific insights about the human brain and its emotional structures, perhaps a more detailed biologically based account of normativity can be worked out. Even though it almost certain to be incomplete it may well turn out provide crucial and corrective insight into the nature of human morality.

As a working hypothesis, I propose that biologically based emotions can help us to explain normativity. According to the theoretical position just stipulated (Ch. 1.2.5), the hypothesis that human morality could have emotional foundations seems reasonable. This is what I intend to explore in more detail. A focal area in this exploration will be the relation between normativity and emotions. My *encouraging suggestion* is that such an exploration is now possible and meaningful due to scientific advances providing better understanding of the emotionality of human nature, notably elucidated by the perspectives and techniques of

affective neuroscience. We should not fear a neurobiological approach but instead embrace it and attempt to identify its limitations as well as prospects.

1.3 Research question and research objectives

1.3.1 Research question

In accordance with the research focus I arrived at in the previous section, I propose the following research question as a guide for this study: *Is it possible to explain the phenomena of normativity based on the current state of scientific knowledge about emotions, and if this is possible, what would the resulting explanatory theory look like?* This research question has two parts. Firstly, it invites an exploration of the phenomena of normativity in terms of empirical evidence about emotions. Exploring and explaining the nature of emotions is therefore a central task. Secondly, and conditionally, it challenges me to show or explain what an emotion based explanation of normativity would look like. This involves identifying the main ingredients necessary for the development of an emotion based theory of normativity.

Normativity can, as noted, be presented as the central concept of any reasonable theory of ethics. I take this to be a relatively uncontroversial underlying premise for understanding the research question, although it should be pointed out that some dismiss the usefulness of deploying the concept of normativity on the grounds that it does not fit with a scientific worldview (e.g., Turner, 2010). I think this eliminative move is unappealing because I conceive of science as something holistic that should cover everything that belongs to reality, including the realities of the mind. All talk, thought, feeling of “ought”, “good”, “right”, and so on, implies normativity, and it seems difficult to make any sense of ethics or morality if we dispense of these notions. By the same lights, any explanatory theory of normativity will contribute to and give shape to an explanation of ethics.

If we accept the argument that normativity should be seen as a psychological notion with biological underpinnings, as I proposed in the previous section that we should, we can identify some of the features of meta-ethics. Meta-ethics – the theoretical perspective that explains ethics – must be empirically oriented and it must be a form of epistemology. It must be epistemological insofar as it deals with the mind, but it must also be empirically based so as to capture the biological premises of the mind. This meta-ethics must show how we think morally as well as the mechanisms that allow us think in these ways. It must therefore be broad enough to explain not only moral language, but also the mental states and dynamics that underpin moral decisions, behavior, and language.

In contrast to many popular philosophical treatments of meta-ethics, a relevant understanding of human nature seems essential. Human nature is central for understanding normative ethics (Bradie, 1994: 11), not least due to the fact mental states and dynamics, including those that support normative perspectives, are embedded in the biology of human nature and its social context. “As soon as we learn more about the neuroscientific basis of ethical reasoning, as well as what underlies self-representation and self-awareness, we may revise our ethical concepts” (Roskies, 2002: 22). Human nature is therefore a natural starting point for sound conceptual development of meta-ethical theory, and I have pointed out that emotionality is a central aspect of human nature.

1.3.2 Research objectives

Given the vast array of different and mutually incompatible views that have been proposed, it is challenging to come to terms with what “ethics” should be taken to mean (see Ch. 7; Kagan, 1998: Ch. 1). Possibly, the single modest point that can be generally agreed upon is that “practical” or “applied” ethics is *normative*; it is in the last instance about what *ought to be done*. While this initially does not seem to get us very far in terms of

understanding normativity, it is nevertheless a start, for now we can proceed by asking relevant questions. *What* is the nature of normativity? *Where* does normativity come from?

In the spirit of these basic questions, I have chosen to focus on the set of research objectives formulated below. They are inter-related in such a way that their joint pursuit is meaningful. If we want to find out *where* the phenomena of normativity come from, we first have to determine *what* they are. These quests are clearly interwoven. The research objectives respectively correspond to the tandem I call “the location problem” and “the nature problem” (see *Fig. 1*, Ch. 1.4). Emotions, I have suggested, may be of key importance for addressing the nature problem. If this is assessment correct emotion or emotionality represents a viable inroad for understanding normative phenomena.

There are two basic components of understanding normativity. The first component is the understanding of the *nature* of the mental states or perspectives. If we unpack this component according to the themes discussed above (Ch. 1.2.4), we arrive at a question of whether the nature of normativity is best seen as *cognitive* or *affective*. Researchers of the mind in several fields (philosophers, psychologists, psychiatrists, and neuroscientists) often albeit not unanimously hold that there is a meaningful and important difference between cognition and affect (this is specified in more detail in Ch. 2).

However, there is also a tension between affect and cognition that has proved difficult to resolve. In Freud’s (1927: 46) words: “One the one hand, you have to admit that men cannot be guided by their intelligence, they are ruled by passions and their instinctual demands. But on the other hand, you propose to replace the effective basis of their obedience to civilization with a rational one”, and continues “it seems that it must be either the one thing or the other”. A crucial question is whether normative authority lies with the intellectual part of cognition or with the desires and drives of the affects. If normativity is to be explained in

terms of emotions we should assume that the nature of normativity to follow the nature of emotions regarding the question of affect versus cognition.

The second component is to understand of *where* normative mental states or perspectives come from; what are the causal mechanisms by which they appear in our minds? Of special interest in this regard, is the subsidiary question of where normativity makes its first appearance; is it in the parts of the mind associated with cognition, or with the parts associated with affect (two distinct versions of internalism)? Alternatively, could it be that normativity first appeared outside the mind (externalism)? To sum up, the overarching set of research objectives and critical sub-questions are:

1. Determine the nature of emotions

- 1.1 Are emotions cognitive or affective?

- 1.2 What is the function of emotions?

- 1.3 What are the sources of emotions?

2. Determine the nature of normativity (the nature problem)

- 2.1 Can normativity be sufficiently explained in terms of emotions?

3. Determine the sources of normativity (the location problem)

- 3.1 Does seeing normativity in terms of emotions solve the location problem?

The discussion in the section above (Ch. 1.2) outlined what I mean by an understanding of normativity, and the type of focus needed in order to approach it. In particular, I argued that there is a need for understanding normativity in terms of the psychology of the subjective mind, including the biological facts that underpin the mind. The phenomena of emotion and consciousness, in this respect, represent parallel aspects of the subjective mind. These two

aspects of the mind present themselves as reasonable areas of inquiry in the quest for understanding normativity. Whenever we are moved to do something, emotion, consciousness, and normativity all represent possible explanatory accounts. This raises an important question as to how one might translate between these explanations.

On the face of it, these three explanations may be either competing or complimentary. Perhaps this is best seen as a case where one perspective does not seem to exclude the others, which would indicate complementarity. One may, for example, be moved into action by a (normative) moral outlook which is emotional, and which furthermore depends on consciousness. However, there still remains a question of what is to be considered the “best” explanation, which represents an element of competition. The ability to conceptually translate between explanations should reduce this competition. Furthermore, it may be possible to establish a *layered* explanatory account of moral action. For example, consciousness may explain emotion, emotion normativity, and in turn normativity may explain moral decisions. For these reasons I propose to keep the explanatory possibilities relatively open.

The location problem can putatively be approached from either a scientific or a philosophical vantage point. By scientific, I mean that theoretical progress must be measured against established theories in proportion to the relative strength of empirical evidence, so that any theoretical position and proposition strives to be empirically adequate. I aim to integrate both philosophical and scientific approaches, in a manner in which the philosophical account have to conform to scientific evidence. In this perspective, philosophy serves to extend what can be said on the basis of empirical evidence. In classical (e.g., Moore, 1903; Sidgwick, 1874/1907) as well as contemporary philosophy (e.g., Parfit, 2011b: 58-73) there has been some opposition against a foundationalist approach that seeks to ground ethics and normativity in biology. Facing up to the challenge, one of my research ambitions is to explore the foundationalist option.

The research objectives above seek understanding. To *understand* something presupposes that there is some fact of the matter to be understood. I have suggested already that normativity should be seen as a *psychological* notion, corresponding to mental phenomena of experienced subjective perspectives. This offers an approach towards a factual understanding of what these subjective mental phenomena are, and how they come about.

If we accept the psychological notion of normativity, it follows that it is meaningful to talk of correct and incorrect usages, hence also more or less accurate usages, of the terms “normative” and “normativity”. Thus, normativity can be understood correctly or incorrectly, and a correct understanding can be more or less accurate. There can be conceptualizations of normativity, for example, that do not correspond to anything, rendering them empty of meaning, which is what Turner (2010) feared was ubiquitous to talk of normativity.

According to the view outlined above, what are to be explained in relation to normativity are, in the first instance, internally experienced subjective “perspectives” or psychological “states of mind”. A theory that does not explain them fails to explain normativity, and hence also morality. Of course, it remains to be explored whether this perspective of normativity works. Alternative understandings of normativity, hence ethics, also have to be explored (see Ch. 7, Ch. 8, and Ch. 9).

Much of the meticulous empirical work needed for a biological account of ethics and normativity appears to be accomplished. For example, Panksepp (1998) has rigorously mapped much of the neurochemical and neuroanatomical emotional infrastructure in humans and other animals (see Ch. 4). The question is whether this scientific research has proceeded far enough, and whether it is sufficient, for reasonable theory building to take place.

It needs to be stressed, however, that I do not suggest that biology, neuroscience, and so on, can provide decision-makers with anything like justified normative theory about what to

do. Again, I am more concerned with how we might obtain a better *understanding* of normative theories and perspectives. Hence, I hope to equip decision-makers with an improved understanding of (1) what normativity is, (2) where normativity comes from, (3) the limits this implies for moral justification, and derivatively (4) what makes certain practical lines of reasoning misguided and potentially harmful. Meta-ethics is not directly normative, but the possibility of indirect normative implications needs to be explored. Potentially therefore, practical normative decision-making becomes more focused, and less prone to certain kinds of error, but without any preconceived solutions to practical moral decision-problems and dilemmas.

The question of which foundations normative ethics rests on is something ethics-practitioners often leave to the philosophers. It is interesting to note, then, that prominent philosophers such as Anscombe (1958: 1) hold that “it is not profitable for us at present to do moral philosophy; that should be laid aside at any rate until we have an adequate philosophy of psychology, in which we are conspicuously lacking”. Now, while this statement was made over half a century ago, the position that psychological insight is a prerequisite for sensibly doing moral philosophy stands. Moreover, the question of whether we *now* have an adequate philosophy of psychology has renewed relevance, not least because of the advancement of neuroscience between then and now. I interpret Anscombe to mean that we need an *empirically adequate* philosophy of psychology. Call this “Anscombe’s challenge”. To meet Anscombe’s challenge, I draw upon relevant perspectives from the current body of scientific evidence.

The reliance of practical ethics on philosophy, and in turn philosophy of psychology, suggests that an inter-disciplinary approach is relevant to understanding normativity and its practical upshots. Moreover, if psychological theory is about the processes that happen inside the human brain (a question treated in Ch. 3), their empirical structure seem to need to

conform to these processes, as well as how the brain relates to the external social world. This points in the direction of neuroscience as a foundationally important part of a unified interdisciplinary understanding of the sources of normativity.

I should mention an important limitation of my research. Some argue that normativity is based on religion or the supernatural or that it has mystical origins (e.g., Schopenhauer, 1818). Instead of going into mystical and religious interpretations of normativity, I would rather work with conceptions of normativity and ethics that are strictly secular. Thus and will offer no explicit treatment of religiously based views. Essentially, I view normativity and ethics as psychological phenomena of human experience, underpinned by facts about the human mind and brain (Gibbard, 2012; 1992; Haidt, 2008; 2001; Prinz, 2007; Blair et al., 2006; Goodenough and Prehn, 2004; Hume, 1740). This will be my basic understanding of normativity.

1.3.3 Specification of criteria

In this section I set up rough criteria by which to determine whether normativity can be adequately explained in terms of emotions. In essence, a sufficient explanation requires that emotions can perform the roles represented by normativity conceived of as mental phenomena. Above (Ch. 1.2) I argued that normativity should correspond to states of mind, and that these states of mind have corresponding biological underpinnings. Since states of mind are dynamic rather than static we should widen this correspondence to include mental processes.

Based on research in neuroscience over the last few decades we can assume that the biological underpinnings of the mind can be characterized in terms of “neural correlates” (Koch, 2004: 16-18), and it seems plausible to assume that “[a]ll mental processes, even the most complex psychological processes, derive from operations of the brain” (Kandel, 1998:

460). This presents a picture in which “what we commonly call the mind is a range of functions carried out by the brain” (Kandel, 1998: 460).

I furthermore noted that normative mental phenomena involve subjective experiences and perspectives of “ought”, “good”, “right” and similar notions, concerning a sense of “what ought to be” as opposed to “what is”. This means that normativity is *evaluative*. Moreover, it can be evaluative retrospectively, about the present, or prospectively about expected future events. Normativity is not limited to the real and the probable, but can also take items of imagination and fiction as objects. For example, we can imagine a world without war and poverty in an approving manner without finding the prospect the least bit realistic.

These normative experiences and perspectives have to cover not only *positive*, but also *negative* valence. Broome (1999: 162-173) argues that ethical terms such as “good” and “bad” should be seen as relational and that they reduces to what he calls “betterness”; questions of what is better than what. Similarly, the idea that something is “right” seems to imply that there is also something comparable that can be “wrong”; this is especially evident in cases of mutually exclusive alternative courses of action in practical contexts that require choice; if one decision alternative can be unambiguously identified as the right one to elect this will imply that the other alternative are the wrong ones. However, we should allow for the possibility that something, such as a prospective action, nonetheless can be concurrently appraised as negative in some respects whilst positive in other respects.

Normativity involves a variety of qualitatively different experience. It seems questionable to assume that they can be measured on a common scale of utility without being depleted in terms of experiential richness (that is not, however, to claim that the practice of utility comparison is necessarily meaningless). We should also keep in mind that not only does a given person have different normative experiences, but also different people, and so we

need to recognize that there are problems for direct comparison between the subjectivities of different individuals. Normative experiences can be characterized as have different evaluative “colorations” or “flavors”. Indeed, shaped by the particularities of each new context and how it is apprehended, each new normative experience may be appropriately seen as unique, although perhaps often familiar given a degree of similarity between various experiences.

Normative experiences of which we are aware and normative perspectives on practical matters are *about* something. They involve object representations to which normative attitudes are attributed or projected. For example, in the phrase “I am confident that this was the right decision”, the “decision” is the relevant represented object along with whatever further specifiable objects this decision implicates. We are not necessarily aware of all our normative experiences, since it has become evident that much of our mental life takes place outside conscious awareness and escapes attention (Bargh and Chartrand, 1999; see Ch. 1.2.3).

As almost unanimously accepted in moral philosophy, normativity *motivates* action. While normative experiences and perspectives are not always sufficiently strong to cause action (or cause agents to act), it does so incline. Normativity denotes motivating phenomena, and it seems plausible that the central function of normative experience and perspectives is to motivate and to guide action.

In order to serve as an adequate explanation for normativity, therefore, emotions have to be able to meet the requirements above. The list provided is not necessarily exhaustive; there may turn out be additional requirements or more refined versions of the ones cited. I have included the ones I consider most central and uncontroversial. If a stricter and more encompassing set of requirements could be identified, that would be an advantage, but I found the cost of doing that would be that I would take aboard too much controversy. An assessment

of the status emotions according to the criteria stated here is to be found at the end of the chapter review theories of affect and emotions (see Ch. 2.7.5). To recap, the requirements that emotions need to meet in order to explain or be representative of normativity are as follows (taken as valid for healthy adult human beings):

- (1) Emotions must, at least in principle, be describable at the level of *neurobiology*. Emotional expressions must correlate with connective activity between nerve cells.
- (2) In the face of moral issues, subjective experiences and perspectives of “ought”, “good”, “right” (etc.) must be accompanied emotional activity. *Subjective normative experience must correlate with neurobiological activity* associated with emotions.
- (3) Emotions must be *evaluative*, meaning that they are subjectively non-neutral. They must render objects to which they attach attractive or unattractive in some way.
- (4) Emotions must cover positive as well as negative *valence*, such that they are capable of pointing either in the direction of “good”, “right”, and “ought” or “bad”, “wrong”, and “ought not”. However it should be allowed for ambivalence, such as in the case of mild fear combined with satisfaction derived from controlling this fear.
- (5) There must be emotions that are about the various features of moral issues as they are conceptualized. There must be at least some emotions that are capable of being directed at *objects*. Conversely, it should be possible that emotions are more fleeting with unspecified targets.
- (6) Emotions must be *motivational*. Emotions must be capable of being experienced as motivating action (whether or not behaviorally efficacious). For every practical moral issue there must be emotions that exert psychological pressure for or against action.

(7) No cases of normative experience or normative perspective holding can happen without at least a minimal involvement of emotions.

1.4 The structure of the thesis

The thesis is structured around the main arguments and positions, as illustrated below (*Fig. 1*). While there is a line of reasoning running through the chapters, such that prior chapters generally build on the conclusions of former chapters, the progression should not be seen as a strict stepwise procedure. Sometimes later chapters give more precise meaning to previous chapters.

Generally, the former chapters (especially Ch. 2 and Ch. 4) lays the foundation for later chapters by discussing more science oriented matters – by which I mean that the former chapters to a greater extent emphasize robust empirically evidence. Conversely, later chapters address more philosophical and practical matters, culminating in a defense of an emotivist meta-ethical position. Accordingly, the emotivist position I call “neuroemotivism” (Ch. 9.4) will be a product of empirical foundations and the philosophical theorizing necessary for providing a holistic conception of normativity. Empirical evidence is taken as premises for philosophical thought; while philosophy has the capacity to move beyond what empirical evidence can shed light on, it has to be empirically adequate vis-à-vis this evidence. In addition to the chapters highlighted in *Fig. 1* below, there is a concluding chapter summarizing the thesis.

Chapter 2 reviews theories of affect and emotion, focusing on important and influential contributions from psychology, neuroscience, and philosophy. It starts out with a presentation of the different views and proceeds to make a summary and discussion of them. Finally, the main insights from theory of emotion are integrated into a coherent theoretical

platform. The review results in an integrated theoretical position which proposes (1) that there are *fundamental biologically based emotions*, and (2) that their nature is *affective*.

Chapter 3 discusses the relation between the mind and the brain. Several prominent views about the mind-brain-relation are presented. Important empirical observations and theoretical ideas are deployed to make an assessment of the different positions. The outcome of this assessment is a verdict in favor of a position called dual-aspect monism (a position originating with Spinoza). *Dual-aspect monism* suggests that while the activity of the mind depends on and is caused by activity in the brain, the mind and the brain do not fully reduce to each other because these notions represent fundamentally different aspects (one that is subjective and another which is objective) of the brain.

Chapter 4 presents evidence showing that emotions originate in specific subcortical areas in the human brain. Each of the emotional centers that are presently known to science with a high degree of confidence is presented. Chapter 4 is important in that it anchors emotions concretely in specific facts about the brain. This ensures a robust correspondence with evidence that grounds the integrated project as a whole in natural facts.

Chapter 5 argues that it is meaningful to see normativity in terms of affective emotions. It picks up the dual-aspect monist position from Chapter 3, arguing that the internal aspect represents a psychological and affective-normative perspective. One of the key arguments that sediments this idea is that consciousness is layered, and that the fundamental layer is an affective consciousness. Thus, affective consciousness and distinct primary affects provide the content of what humans experience as normative. Neuroanatomically affective consciousness underpins the basic emotional circuits (that were presented in Ch. 4).

Chapter 6 argues that decisions require appropriate emotional input. A central idea is that, without appropriate input from affective emotions, the normative basis of decisions will

be basically arbitrary (although often systematic). It is argued that emotions cannot generally be chosen consciously, that they are fundamentally non-rational, but that they nevertheless support and furnish “rational” or high-quality decision-processes. It is recognized that most of the processes that take place in decision-making are outside conscious awareness, including emotional regulation.

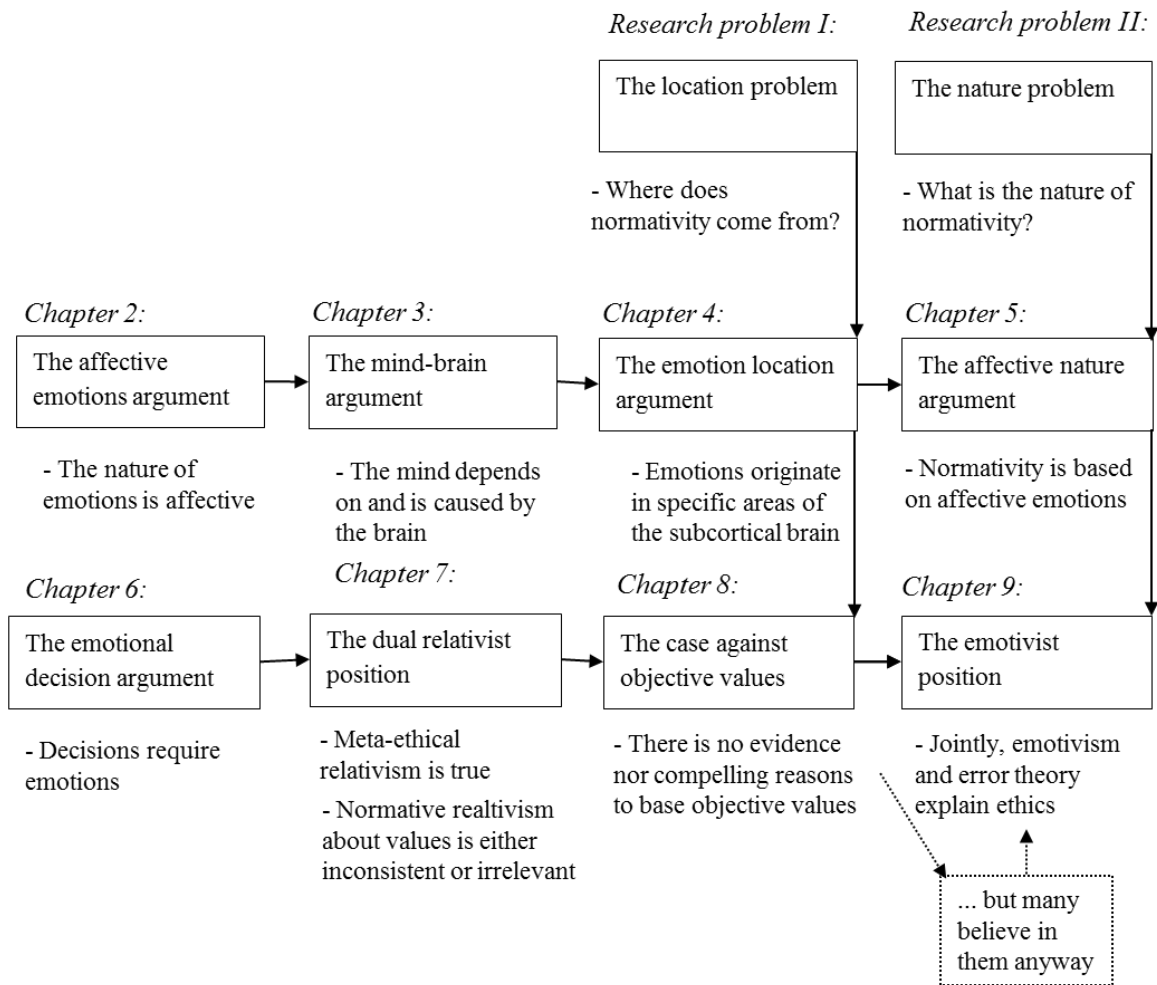
Chapter 7 defends a dual position concerning relativism about values. First, relativist theories are classified as either meta-ethical theories (concerning epistemology and understanding) or normative theories (about practical recommendations). Each set of theories is subsequently assessed and compared with relevant alternative theoretical perspectives. It is argued that *meta-ethical relativism* explains moral thought, talk, and action, while meta-ethical error theory explains cognitive errors. Meanwhile, four versions of *normative relativism* are identified and rejected as a basis for normative decision-making. It is argued that normative ethics needs to be based on personal values in order to ensure that affective content is intact in practical reasoning. While I remain methodologically agnostic about what personal values ought to “say” substantively in practical cases and contexts, I argue that they ought to say something, and that in explicit moral deliberation, this something ought to be anchored in a teleological structure that emphasizes second-order values (desires) over first-order desires.

Chapter 8 presents the “one-stone-two-birds” argument that, on the one hand, there is neither evidence nor any compelling reason to believe that there are objective values. In contrast, there *is* a compelling theory that locates values internally and connects them to a hierarchy of subjective desires, which in turn are experiential upshots of affective emotional processes in the human brain.

Chapter 9 presents a short historical literature review of emotivism, and proceeds to update this view based on what science informs us about human nature, notably including the nature of emotions. This chapter ties together many of the threads from preceding chapters. An updated version of emotivism is defended as a basis for understanding normativity and for practical ethics in terms of implications of this understanding. I propose a theory that I have chosen to call *neuroemotivism*. Neuroemotivism emphasizes the broader picture of human nature and the related empirical evidence that underpins meta-ethics.

Fig. 1 illustrates the main arguments and positions and shows which chapters discuss them. In order to show clearly what problems the arguments aim to resolve or illuminate, I have also included the auxiliary research problems 1 and 2 (slightly abbreviated) in the figure. The arrows in *Fig. 1* are there to highlight the flow of reasoning that runs through the chapters. There is also a concluding tenth chapter that is omitted in the figure for the sake of simplicity.

Fig. 1 – Outline of the main argumentative structure:



1.5 The social dimension

1.5.1 The socialized individual

The focus of the thesis is chiefly on the perspective of the individual decision-maker, and with the individual decision-maker in mind. This is because, ultimately, decisions are attributable to individuals. Social entities, such as for instance organizations, do not have a mind of their own, because “all the facts about the group mind supervene on beliefs and attitudes of the members” (Blackburn, 2010: 70). Hence, “morality pertains to individual human beings and their interpersonal relations, and only derivatively to social entities” (Miller, 2010: 19).

However, to having the individual in mind, including the individual’s mind in mind, in discussions of human morality and decision-making, implies, for very fundamental reasons, dealing with the *social nature* and *interactions* of the individual. One reason is that human beings are fundamentally social creatures. The neuroscientist Vittorio Gallese puts it this way:

“Any time we meet someone, we are implicitly aware of his/her similarity us, because we literally embody it. The very same neural substrate activated when actions are executed or emotions and sensations are subjectively experienced, is also activated when the same actions, emotions, and sensations are executed or experienced by others.” (Gallese, 2009: 524).

In this passage, Gallese points to the fact that, already from birth, our brains are hardwired to interact socially, and does so largely in subconscious fashion. According to Gallese (2009), humans have special “mirror-neurons” that are engaged in “embodied simulation” of objects our external world, thereby enabling empathy and grounding our social identification and interconnectedness with others. Thus, the social precursors of morality are to some extent built into our brain structures.

While the minds from which moral decisions are made belong to the individuals who make the decisions, these are social minds operating in social contexts. Hence, the methodological individualism espoused in the current study is a *social* individualism. It can appropriately be seen as a “relational individualism” (Miller, 2010: 18), emphasizing the densely social and relational context within which individual decision-makers interact. Relational individualism “does not adhere to the implausible doctrine that social entities can be reduced to the individual human beings who constitute them” (Miller, 2010: 18), nor does it assume that decision-makers operate in social isolation.

The fact that the experiences of a person take place in her own mind does not mean that they are not social. As Han and Northoff (2008: 646) note: “Our brains and minds are shaped by our experiences, which mainly occur in the context of the culture in which we develop and live.” Highlighting the relation between emotions and culture, Prinz (2007: 143) argues as follows: “Moral responses require moral emotions, and moral emotions may emerge through enculturation”. This gives culture and socialization a crucial role in shaping the emotions of individuals.

The focus on the individual is important for a discussion of moral responsibility in the practical context. It is of psychological importance because it sheds light on how persons relate to the decision they make or contribute to making. It is also of practical importance because the application of the concept of individual moral responsibility can serve as a tool for cutting through more nebulous and contingent notions of moral responsibility such as “corporate responsibility”, “collective responsibility”, and “legal responsibility”. The Danish philosopher, Knut Løgstrup, argued that responsibility is primarily attributable to persons, not collective groups or organizations (Alm, 2012: 68-69). Citing the notion of collective responsibility, in particular, does not erase personal responsibility. It is not difficult to see that

if it did, this would shape the incentives of agents and decision-makers in organizations and have encompassing practical moral implications.

1.5.2 The social context

It is essentially the social context that makes it appropriate to conceive of decisions as moral. As later detailed, individuals are generally social in that they care about others. Adam Smith (1759) highlighted this point, in his *Theory of Moral Sentiments*, in two morally significant ways; we care about others in themselves (the basis for altruism), and also about the social responses of others (the basis for reciprocity). On top of these two psychological moral features, social relations shape our moral lives in fundamental ways, and in part determine how our propensity for altruism and reciprocity plays out. Individuals lead highly social, relational, and affective lives (Damasio, 2000). This is also the case in professional ethics, for example business life, where, according to Adam Smith our sense of fairness and justice is shaped by the psychological principle of reciprocity (de Vries, 1989).

The social context creates *normative tension* that bears on decisions. Decision-makers are often pulled in different directions, and this is a social basis for perceiving some decisions as involving dilemmas, or as morally controversial. Different stakeholders in the social context have different thing at stake – needs, wants, desires, concerns – that can be frustrated or promoted by decisions.

The things at stake can be concrete, such as for example pecuniary losses, or loss of expected life years; they can also be more abstract, such as the perceived loss of professional reputation. Often decisions entail concrete physical as well as abstract psychological impacts. For example, it can be appropriate to aggregate such concrete things, such as the physical contribution to global climate change represented by additional units of carbon dioxide emitted into the atmosphere, with concerns like avoiding damaging someone's reputation. As

Oderberg (2013: 3) puts it: “There is a tension between the reasonable desire not to be judgmental of other people’s behaviour or character, and the moral necessity of making negative judgments in some cases”. This suggests that socially complex decisions typically involves normative tension between relational concerns and general concerns, as well as between concerns about concrete consequences and more abstract psychological effects.

There is a related way to distinguish two broad types of normative concerns that decision-makers may have in the social context; the distinction (Broome, 1999) between a concern for the *size* of impacts (the amount of good) and its *distributive* implications (the distribution of good). When a decision is strictly private, impacting only the agent herself, only the amount of good matter. But in the socially dense context things are typically socially complex, and the *size* of impacts as well as their social *distributions* seems morally relevant. Such decision problems are especially prone to normative tension because meeting distributive concerns will normally be at the expense of the total amount of good a decision can produce.

1.6 Interdisciplinary method

I aim to offer a seamlessly interdisciplinary perspective. It is widely recognized that an interdisciplinary approach is important for understanding ethics, although a few territorial moral philosophers object to this view. After all, ethics is *applied* to virtually every practical occupation and professional discipline that require moral judgment; law, medicine, health-care, economics, finance, accounting, engineering, marketing, and so on. Since ethics is not the primary a field of expertise in these disciplines, they often import theories from other fields, in particular moral philosophy. For example, many seminal contributions to business ethics have been made by applying moral philosophy to business ethics (Bryne, 2002), and business ethics also import theory and evidence from psychology to bear on moral issues. As

argued above, understanding the normative core of ethics requires an understanding of human emotions that their place in the human mind. This understanding also requires an integrated disciplinary approach (Bernstein, 2011: 13; Panksepp, 1998: Ch. 1).

1.6.1 The forms of integration

Disciplinary integration can take different forms. “Vertically” it can be performed at different *levels of theory*, and “horizontally” with different *theoretical scopes*. The thesis focuses vertically primarily on *the foundational level*, because one of the main aims is to understand the *sources* human normativity. Establishing an empirically adequate understanding of relevant biological and psychological aspects of human nature is a central scientific concern.

The foundational level can be usefully distinguished from a higher “level of social construction”. Social construction includes everything that is, or can be, experienced as a normative reason or force that either comes about, or have come about, due to processes of social interaction; objects that we in virtue of mutual recognition treat *as if* “objective” facts and values (Putnam, 1985). In this sense “moral artefacts” and societal-level “value systems” can be said to be “socially constructed”. Examples include the giving of promises, the making of laws, the signing of contracts, the institution of property rights, and norms of culturally defined group of people. Frederick (1995: 290) proposes an equivalent distinction which sees the human sense of morality as a level beneath and prior to socio-cultural ethical systems and principles, which are subsequently placed on top of this older, evolutionary based, level represented by moral sense.

My understanding of science is to a considerable extent inspired by van Fraassen’s (1980: 12) constructive empiricism: “Science aims to give us theories which are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate”, where “empirical adequacy” essentially means that what is *said* about the observable things

and events in the world is true according to evidence. Faced with a complex and widely discussed topic such as “emotions”, it would seem question begging to include the empirically input from one specific discipline only, as a basis for determining empirical adequacy. This suggests that, in principle, a relative wide theoretical scope is appropriate. Perhaps more important still, the scope should be held with a certain methodological openness so that perspectives from outside can readily be invited in.

The theoretical scope should be chosen so as to be wide enough to safeguard crucial empirical and theoretical contributions for the overall integration of theory. At the same time, the research scope cannot be too wide for then it will be impossible to handle within the constrained space of the thesis. This does not mean that empirical and theoretical inputs outside the chosen scope are considered of little value; it is merely a pragmatic choice based on the fact that a scope has to be set. To make up for this, the thesis is written in a manner that invites further disciplinary integration.

1.6.2 The pillars of integration

The three disciplinary pillars of the thesis are, as indicated earlier this chapter, philosophy, psychology, and neuroscience; and the focus is on their treatment of emotions and related concepts such as “values” and “desires”, and the conflicts between “cognitive” and “affective” perspectives. It can be objected that it is not prudent or economical to treat several overlapping notions, but I consider this necessary because successful disciplinary integration is likely to depend in part on connection and integration of related concepts, as this putatively enhances inter-disciplinary commensurability. I try to avoid incommensurability of theories caused by breakdown of communication across disciplinary division lines or paradigms (Kuhn, 2002: Ch. 2) due to lack of common terminology or shared concepts.

Other disciplinary pillars could have been chosen, but some of these would fall outside my own areas of expertise, and not all related fields seem equally central to address my research objectives and research problems. Other interdisciplinary approaches, however, overlap considerably with the integration project pursued here. For example, Northoff (2011: 15) operates with four disciplinary pillars to support his integrated “neuropsychanalysis”; psychology/neuroscience, psychoanalysis, philosophy, and psychiatry. While I leave psychoanalysis and psychiatry out of the discussion, it should be kept in mind that these disciplines remain relevant; as such they represent potent avenues for future research.

Much of my focus is on evidence from neuroscience, in particular affective neuroscience. Due to the hard fact quality of some of this evidence, it occupies a substantial place within the overall integrated approach. Affective neuroscience represents an empirically driven *paradigmatic shift* that is shaping neuroscience from the inside (Panksepp, 1998: 3-7). As others have pointed out, the advances in understanding emotion are likely to have wider theoretical implications: “A revolution in the science of emotion has emerged in recent decades, with the potential to create a paradigm shift in decision theories” (Lerner et al., 2015: 799).

This is by no means to say that affective neuroscience operates in an intellectual vacuum, nor that it is suited to address all the questions we need to address in order to understand the normativity of practical ethics. My aim is to explore the nature and sources of normativity. Besides the perspective of affective neuroscience, this aim is also assisted by philosophizing from the qualitative feel of subjective experience (e.g., Bergson, 1935), intersubjective exchange of thought, and psychological approaches directed at understanding the moral mind (e.g., Kahane, 2013; Haidt, 2008). For this reason, I argue for the necessity of a *holistic and interdisciplinary approach*. As an interdisciplinary undertaking, I have found it helpful to include also psychology and philosophy as disciplinary perspectives.

The facilitation of interdisciplinary communication and integration is another central contribution of my study. I contribute to breaking down disciplinary boundaries that I view as artificial and counterproductive with respect to the prospect of arriving at an understanding of normativity and human nature.

While interdisciplinary approaches are not uncommon in practical ethics research, most of these have been preoccupied with the interdisciplinary challenges of applying “ethics”, or with the implementation of “ethics”. Much less focus appears to have been directed at understanding the normative foundations of practical ethics from an interdisciplinary theoretical perspective. Mostly, questions about normative foundations appear to have been “outsourced” to moral philosophy. While moral philosophy has much to offer in terms of reasoning and conceptualization, it mostly discusses normativity in a manner that does adequately account for the complexities of human nature and how normativity emerges in the human mind as a distinctive type of *experience*. I argue that psychological and neurological insight is essential for understanding normativity. The disciplinary integration that I propose represents a theoretical constellation aimed at establishing a basic understanding of normativity, and thereby the normative foundations of ethics and moral practice.

1.7 Reading guidance

A brief note on how to read the thesis is in order. As far as possible the chapters of the thesis are ordered along a chain of arguments, as illustrated by argument structure (see Fig. 1). There should, I hope, be a visible and compelling logic behind the chosen order of the chapters. Generally, the scientific arguments are presented early, the philosophical refinement and integration somewhere in the middle, and the practical upshots toward the end. At the same time, I found it impossible to create a strict order in which each argument successively

lead on to the next, steadily progressing from beginning to end. Many of the arguments in the different chapters harbor mutual interdependencies that point both ways. Furthermore, there is unavoidably some overlap in the subject matter presented in the different chapters.

The table of content will therefore be an especially useful tool for navigating and coming to terms with the complete set of arguments. Each main argument and position has a designated chapter, and I have attempted to make as informative headings as possible. In order to minimize repetition, I have extensively cross-referenced the chapters of the thesis, deferring certain arguments to their most appropriate chapters. For example, sometimes an argument is outlined in one chapter and detailed in another.

My final suggestion is that the reader to some extent considers the thesis a web of arguments, rather than merely a sequential chain of reasoning. I suspect the reader will find it helpful to understand not only later chapters in the light of early chapters, but also the reverse. I have attempted to make an appropriate structure with a supply of pointers to facilitate following the cross-referenced material as a coherent, integrated, theoretical unity. In this regard, it can also be helpful to keep in mind that the emotivist position (Ch. 9) roughly defines the center of the web. Whether read sequentially or not, I end up in the position that I call neuroemotivism.

References chapter 1:

- Alm, K. (2012) *Yrkesetikk: Utfordringer for næringsliv og finans*. Oslo: Universitetsforlaget.
- Anscombe, G. E. M. (1958) Modern moral philosophy. *Philosophy* **33** (124): 1-16.
- Bagnoli, C. (2011) Emotions and the categorical authority of moral reason. In C. Bagnoli (Ed.) *Morality and the emotions*. Oxford: Oxford University Press.
- Bargh, J. and E. Morsella (2008) The unconscious mind. *Perspectives on Psychological Science* **3** (1): 73-79.
- Bargh, J. A. and Chartrand, T. L. (1999) The unbearable automaticity of being. *American Psychologist* **54** (7): 462-479.
- Batson, C. D. (1991) *The altruism question: Toward a social-psychological answer*. Hillsdale: Lawrence Erlbaum Associates.
- Beckes, L., Coan, J. A., and Hasselmo K. (2013) Familiarity promotes the blurring of self and other in the neural representation of threat. *Social Cognitive & Affective Neuroscience* **8** (6): 670-677.
- Bergson, H. (1977) *The Two Sources of Morality and Religion*. (R. A. Audra and C. Brereton, Trans.). Paris: University of Notre Dame Press. (Original work published in 1935).
- Berlin, H. (2011) The neural basis of the dynamic unconscious. *Neuropsychoanalysis* **13** (1): 5-31.
- Bernstein, W. M. (2011) *A basic theory of neuropsychoanalysis*. London: Karnac Books.

- Blackburn, S. (2010) *Practical tortoise raising and other philosophical essays*. New York: Oxford University Press.
- Blackburn, S. (1998) *Ruling passions: A theory of practical reasoning*. Oxford: Oxford University Press.
- Blackburn, S. (1993) *Essays in quasi-realism*. New York: Oxford University Press.
- Blair, J., Marsh, A. A. Finger, E. Blair K. S., and Luo J. (2006) Neuro-cognitive systems involved in morality. *Philosophical Explorations* **9** (1): 13 – 27.
- Bradie, M. (1994) *The secret chain: Evolution and ethics*. New York: SUNY Press.
- Brickner, R. M. (1944) Man and his values considered neurologically. *Journal of Philosophy* **41** (9): 225-243.
- Brink, D. O. (1989) *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.
- Brinkman, S. (2010) *Psychology as a moral science: Perspectives on normativity*. New York: Springer.
- Broome, J. (2013) *Rationality through reasoning*. Oxford: Wiley-Blackwell.
- Broome, J. (2005) Does rationality give us reasons?. *Philosophical Issues* **15**: 321-337.
- Broome, J. (2004) *Weighing lives*. Oxford: Oxford University Press.
- Broome, J. (1999) *Ethics out of economics*. Cambridge: Cambridge University Press.
- Bryne, E. F. (2002) Business ethics: A helpful hybrid in search of integrity. *Journal of business ethics* **32** (2): 121-133.

- Cacioppo, J. T. and Berntson, G. G. (2009) *Handbook of neuroscience for the behavioral Sciences*. New York: John Wiley & Sons.
- Casebeer, W. D. and P. S. Churchland (2003) The neural mechanisms of moral cognition: A multiple-aspect approach to moral judgment and decision-making. *Biology and Philosophy* **18**: 169–194.
- Cartwright, J. (2010) Naturalising ethics: The implications of Darwinism for the study of moral philosophy. *Science and Education* **19**: 407–443.
- Ciaramelli, E. and di Pellegrino, G. (2011) Ventromedial prefrontal cortex and the future of morality. *Emotion Review* **3** (3): 308-309.
- Chalmers, D. (2003) Consciousness and its place in nature. In Stich, S. P. and Warfield, T. A. (eds.) *Blackwell guide to the philosophy of mind*. Malden: Blackwell: 102-142.
- Chalmers, D. (1995) Facing up to the problem of consciousness. *Journal of Consciousness Studies* **2** (3): 200–219.
- Churchland, P. S. (2011) *Brain trust: What neuroscience tells us about morality*. Princeton: Princeton University Press.
- Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs, P., and Skuce A. (2013) Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters* **8** (2): 1-7.
- Copp, D. (2007) *Morality in a natural world: Selected essays in metaethics*. Cambridge: Cambridge University Press.
- Copp, D. (1995) *Morality, normativity, and society*. New York: Oxford University Press.
- Crisp, R. (2006) *Reasons and the good*. Oxford: Oxford University Press.

- Damasio, A. R. (2010) *Self comes to mind: Constructing the conscious brain*. New York: Random House.
- Damasio, A. R. (2000) *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.
- Damasio, A. R. (1994) *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- Dancy, J. (2004) *Ethics without principles*. Oxford: Oxford University Press.
- D'Arms, J. and Jacobson, D. (1994) Expressivism, morality, and the emotions. *Ethics* **104**: 739-763.
- Davidson, R. J. and Sutton, S. K. (1995) Affective neuroscience: The emergence of a discipline. *Current Opinion in Neurobiology* **5**: 217-224.
- Darwin, C. (1906) *The descent of man and selection in relation to sex (2nd ed.)*. London: John Murray.
- Decety, J. and Cowell, J. M. (2014) The complex relation between morality and empathy. *Trends in Cognitive Science* **18** (7): 337-339.
- De Vries, P. H. (1989) Adam Smith's 'theory' of justice: Business ethics themes in the Wealth of Nations. *Business & Professional Ethics Journal* **3** (1): 37-55.
- de Waal, F. B. M. (2009) *Primates and philosophers: How morality evolved*. Princeton: Princeton University Press.
- de Waal, F. B. M. (1996) *Good natured: The origins of right and wrong in humans and other animals*. Cambridge: Harvard University Press.

- Dijksterhuis, A. (2004) Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality and Social Psychology* **87** (5): 586–598.
- Ditlev-Simonsen, C. D. and Wenstøp, S. (2011) Companies' ethical commitment: An analysis of the rhetoric in CSR reports. *Issues in Social and Environmental Accounting* **5** (1/2): 65–81.
- Durkheim, E. (1912/1995) *The elementary forms of religious life*. Fields, K. E. (trans.). New York: Free Press.
- Etzioni, A. (1988) *The moral dimension: Toward a new economics*. New York: The Free Press.
- Falk, W. D. (1947-1948) "Ought" and motivation. *Proceedings of the Aristotelian Society* **48**: 111–138.
- Field, H. (2009) The normative role of logic. *Proceedings of the Aristotelian Society* **83**: 251–268.
- Finlay, S. (2010) Recent work on normativity. *Analysis* **70**: 331–346.
- Frankfurt, H. G. (1988) *The importance of what we care about: Philosophical essays*. Cambridge: Cambridge University Press.
- Frederick, W. C. (1995) *Values, nature and culture in the American corporation*. New York: Oxford University Press.
- Freud, S. (1927) The future of an illusion. In W. D. Robson-Scott (Trans.) *The standard edition of the complete psychological works of Sigmund Freud, V. XXI (1927-1931)*:

- The Future of an Illusion, Civilization and its Discontents, and Other Works* (pp. 1-56). London: Hogarth Press.
- Frith, C. (2007) *Making up the mind: How the brain creates our mental world*. New Jersey: John Wiley & Sons.
- Gallese, V. (2009) Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanalytic Dialogues* **19**: 519–536.
- Gauthier, D. (1986) *Morals by agreement*. Oxford: Oxford University Press.
- Gibbard, A. (2012) *Meaning and normativity*. Oxford: Oxford University Press.
- Gibbard, A. (1992) *Wise choices, apt feelings: A theory of normative judgment*. Oxford: Oxford University Press.
- Glover, J. (1984) *What sort of people should there be?*. London: Penguin.
- Goldie, P. (2012) *The mess inside: Narrative, emotion, and the mind*. Oxford: Oxford University Press.
- Goldman, A. H. (2009) *Reasons from within: Desires and values*. Oxford: Oxford university press.
- Goodenough, O. R. and Prehn, K. (2004) A neuroscientific approach to normative judgment in law and justice. *Philosophical Transactions of the Royal Society of London* **359**: 1709–1726.
- Greene, J. (2003), From neural ‘is’ to moral ‘ought’: What are the moral implications of neuroscientific moral psychology. *Nature Reviews Neuroscience* **4**: 847–850.

- Habermas, J. (1984) *The theory of communicative action: Vol. I: Reason and the rationalization of society*. McCarthy T. (trans.). Boston: Beacon.
- Haidt, J. (2008) From philosophical thinking to psychological empiricism. *Perspectives on Psychological Science* **3** (1): 65-72.
- Haidt, J. (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* **108** (4): 814-834.
- Han, S. and Northoff, G. (2008) Culture-sensitive neural substrates of human cognition: A transcultural neuroimaging approach. *Nature Reviews Neuroscience* **9**: 646–654.
- Hare R. D. (1991) *The Hare psychopathy checklist: Revised*. Toronto: Multi-Health Systems.
- Harman, G. (1997) *The nature of morality*. New York: Oxford University Press.
- Harman, G. (1986) *Change in view: Principles of reasoning*. Cambridge: MIT Press.
- Harman, G. (1977) *The nature of morality: An introduction to ethics*. New York: Oxford University Press.
- Hausman, D. and McPherson M. (2006) *Economic analysis, moral philosophy, and public policy*. Cambridge: Cambridge University Press.
- Hubin. D. C. (2001) The groundless normativity of instrumental rationality. *Journal of Philosophy* **98** (9): 445–468.
- Hume, D. (1740/1978) *A treatise of human nature*. London: Clarendon Press.
- Kahane, G. (2013) The armchair and the trolley: An argument for experimental ethics. *Philosophical Studies* **162** :421–445.
- Kagan, S. (1998) *Normative ethics*. Boulder: Westview Press.

- Kandel, E. R. (1998) A new intellectual framework for psychiatry. *American Journal of Psychiatry* **155** (4): 457-469.
- Kelsen, H. (1967) *The pure theory of law*. Berkeley: California University Press.
- Klatt, M. (2008) *Making the law explicit: The normativity of legal argumentation*. Oxford: Hart Publishing.
- Kitcher, P. (2011) *The ethical project*. Cambridge: Harvard University Press.
- Koch, C. (2004) *The quest for consciousness: A neurobiological approach*. Englewood: Roberts & Company Publishers.
- Kohlberg, L. (1984) *The psychology of moral development: The nature and validity of moral stages*. New York : Harper & Row.
- Kolodny, N. (2005) Why be rational?. *Mind* **114** (455): 509–563.
- Korsgaard, C. M. (1996) *The sources of normativity*. Cambridge: Cambridge University Press.
- Kuhn, T. (2002) The road since structure: Philosophical essays, 1970-1993, with an autobiographical interview. *Chicago: University of Chicago Press*.
- Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. S. (2015) Emotion and decision making. *Annual Review of Psychology* **66**: 799–823.
- Lipps, T. (1903) *Einfuehlung, innere nachahmung und organenempfindung*. Leipzig: W. Engelmann.

- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., and Barrett, L. F. (2012) The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences* **35** (3): 121–143.
- MacIntyre, A. (1981) *After virtue: A study in moral theory*. Paris: University of Notre Dame Press.
- Mackie, J. L. (1977) *Ethics: Inventing right and wrong*. London: Penguin Books.
- MacLean, P. D. (1990) *The triune brain in evolution: Role in paleocerebral functions*. New York: Plenum Press.
- McKeever, S. and Ridge, M. (2005) The many moral particularisms. *The Canadian Journal of Philosophy* **35**: 83–106.
- McNaughton, D. (1991) *Moral vision: An introduction to ethics*. New York: Wiley.
- Miller, S. (2010) *The moral foundations of social institutions: A philosophical study*. Cambridge: Cambridge University Press.
- Modell, A. H. (2003) *Imagination and the meaningful brain*. Cambridge: The MIT Press.
- Montague, P. (1995) *Punishment as societal-defense*. Rowman & Littlefield.
- Moore, G. E. (1903) *Principia Ethica*. Cambridge: Cambridge University Press.
- Nagel, T. (1986) *The view from nowhere*. Oxford: Oxford University Press.
- Narvaez, D. (2008) Triune ethics: The neurobiological roots of our multiple moralities. *New Ideas in Psychology* **26**: 95–119.
- Northoff, G. (2011) *Neuropsychanalysis in practice: Brain, self and objects*. Oxford: Oxford university press.

- Northoff, G. (2004) *Philosophy of the brain: The brain problem*. Amsterdam: John Benjamin's Publishing.
- Oderberg, D. S. (2013) The morality of reputation and the judgment of others. *Journal of Practical Ethics* **1** (2): 3–33.
- Panksepp, J. and Biven, L. (2012) *The archeology of mind: Neuroevolutionary origins of human emotions*. New York: Norton.
- Panksepp, J. (2011) Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals. *PloSOne* **6** (9): e21236.
- Panksepp, J. (2007) Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives on Psychological Science* **2**: 281-296.
- Panksepp, J. (2005) Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition* **14**: 30–80.
- Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.
- Panksepp, J. (1992) A critical role of “affective neuroscience” in resolving what is basic about basic emotions. *Psychological Review* **99**: 554–560.
- Parfit, D. (2011a) *On what matters: Volume 1*. Oxford: Oxford University Press.
- Parfit, D. (2011b) *On what matters: Volume 2*. Oxford: Oxford University Press.
- Pfaff, D. W. (2011) *Man and woman: An inside story*. Oxford: Oxford University Press.
- Plutchik, R. (1991) *The emotions*. New York: University Press of America.

- Prinz, J. J. (2007) *The emotional construction of morals*. Oxford: Oxford University Press.
- Putnam, R. A. (1985) Creating facts and values. *Philosophy* **60** (232): 187–204.
- Rawls, J. (1993) *Political Liberalism*. New York: Columbia University Press.
- Rawls, J. (1980) Kantian constructivism in moral theory. *Journal of Philosophy* **77**: 515–572.
- Robertson, S. (2009) Introduction: Normativity, reasons, rationality. In Robertson, S. (ed.). *Spheres of reason: New essays in the philosophy of normativity*. Oxford: Oxford University Press.
- Rosenberg, A. and Sommers, T. (2003) Darwin's nihilistic idea: Evolution and the meaninglessness of life. *Biology and Philosophy* **18**: 653-668.
- Roskies, A. (2002) Neuroethics for the new millennium. *Neuron* **35** (1): 21-23.
- Schmidt, M. F. H., Rakoczy, H., and Tomasello, M. (2011) Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science* **14** (3): 530–539.
- Searle, J. R. (2001) *Rationality in action*. Cambridge: MIT Press.
- Sen, A. (1997) Economics, business principles and moral sentiments. *Business Ethics Quarterly* **7** (3): 5-15.
- Sidgwick, H. (1874/1907) *The methods of ethics* (7th. ed.). London: Macmillan and Company.
- Smith, A. (1759/2010) *The theory of moral sentiments*. New York: Digireads.com Publishing.
- Smith, M. (2004) *Ethics and the a priori: Selected essays on moral psychology and meta-ethics*. Cambridge: Cambridge University Press.
- Solms, M. and Turnbull, O. (2002) *The brain and the inner world*. New York: Other Press.

- Schopenhauer, A. (1818/2008) *The world as will and Presentation*. R. E. Aquila (trans.) and D. Carus (ed.). New York: Longman.
- Stevenson, C. L. (1937) The emotive meaning of ethical terms. *Mind* **46**: 14-31.
- Thomson, J. J. (2008) *Normativity*. Chicago: Open Court.
- Tomkins, S. S. (1962) *Affect imagery consciousness: Volume I: The positive affects*. London: Tavistock.
- Tomkins, S. S. (1963) *Affect imagery consciousness: Volume II: The negative affects*. London: Tavistock.
- Turner, S. P. (2010) *Explaining the normative*. Cambridge: Polity Press.
- van Fraassen, B. C. (1980) *The Scientific Image*. Oxford: Oxford University Press.
- Wallace, R. J. (2006) Normativity and the will: Selected essays on moral psychology and practical reason. *Oxford: Oxford University Press*.
- Wedgwood, R. (2007) *The nature of normativity*. London: Clarendon Press.
- Wegner, D. M. (2002) *The illusion of conscious will*. Cambridge: MIT Press.
- Williams, B. (1985) *Ethics and the limits of philosophy*. London: Routledge.
- Williams, B. (1973) *Problems of the self: Philosophical papers*. Cambridge, Cambridge University Press.
- Whiting, D. (2007) The normativity of meaning defended. *Analysis* **67**: 133-140.
- Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E., and Grafman, J. (2009) The neural basis of human social values: Evidence from functional MRI. *Cerebral Cortex* **19**: 276–283.

Chapter 2: Theories of affect and emotion²

2.1 Introduction to the theory of emotion

Emotions have recently attracted much interest in the fields of philosophy (e.g., Goldie, 2012; Blackburn, 2010: Ch. 6; Solomon, 2007; Robinson, 2005; Prinz, 2004), social science (e.g., Greco and Stenner, 2008; Frijda, 2007; Lazarus, 1991), and the (natural) life sciences (e.g., McNaughton, 1989), notably including neuroscience (e.g., Damasio, 2010; Panksepp, 1998; LeDoux, 1996). In addition, there has been a growing interest in emotions in areas of disciplinary overlap (e.g., Panksepp and Biven, 2012; Prinz, 2007; Solms and Turnbull, 2002). However, emotions have attracted relatively less attention in practical ethics, such as business ethics.

Historically, textual evidence traces the English term “emotion” back to at least 1579, and suggests that it was imported from the French term “émouvoir”, which means “to stir up” (Suchy, 2011: 7). Nowadays, the term is used in a wide range of ways in the science-oriented literature (Siegel, 2009: 147), some of which are specific and substantive, while others more general and unqualified (Izard, 2010: 368).

Recent empirically oriented literature links psychological mental states of emotion with processes in the brain and nervous system (e.g., Panksepp and Biven, 2012; Churchland, 2011; Damasio, 2010; Porges, 2009; Panksepp, 1998; LeDoux, 1996; Izard, 1991; Zajonc, 1980). The manner in which the term “emotion” is used in philosophical discourse demonstrates a range of usage that is at least as wide as in the scientific disciplines outside philosophy. The result is that “there is no consensus on a definition of ‘emotion’” (Izard, 2010: 363; Zachar, 2010; Panksepp, 1998).

² I am indebted to Jaak Panksepp and Sigmund Karterud for generously commenting on this chapter. I would also like to thank Siri Gullestad for her suggestions.

Despite differences concerning the meaning of “emotion,” however, there are some points of theoretical agreement, including an understanding of the *functions* of emotion (Izard, 2010: 368); most psychological studies agree that the central functions of emotion are in *motivating* behavior and *focusing* attention (Izard, 2010), and this way of viewing emotions is also reflected in philosophy (Bagnoli, 2011: 62). Furthermore, there seems to be broad consensus among emotion researchers that there are rapid, automatic, and unconscious connections among emotion, cognition, and action (Izard, 2010: 366).

This review chapter documents that, although we are far away from a consensus regarding emotions in the scientific community as a whole, there is a coherent and scientifically based center of gravity. This center of gravity is grounded in a body of well-established empirical facts regarding the physiology, the causes, and the effects of emotions. However, somewhat paradoxically, the fact that there is such a center of gravity does not indicate that it represents the majority view in emotions research (Panksepp, 1998: Ch. 1).

The reasons for this are undoubtedly many but include at least the following four points; (1) recent evidence contradicts traditionally popular views about emotions that still have many adherents and where the theoretical investment is considerable; (2) relevant empirical facts are not brought to bear on relevant questions about emotions, sometimes because the questions and the relevant facts are located in different disciplines so that the facts are not known widely enough or are insufficiently understood; (3) many different things are meant or implied by the way quite different theoretical perspectives use the term “emotion”; and (4) the interdisciplinary approach that an understanding of emotions seems to demand is both challenging and insufficiently funded (Panksepp and Biven, 2012).

It should also be noted that emotions theory sails under many different flags. In the psychology literature, for example, emotion theory is also known as “motivation theory,”

“stress and coping theory,” and “theory of conflict and avoidance” (Plutchik, 1991: 24). This fact seems likely to hinder the lack of unification of theory and terminology surrounding the concept of emotions and its underlying phenomena.

In a critical assessment of the central theoretical positions across philosophy, psychology, and neuroscience, I identify the distinction between *cognitive* theories and *affective* theories as the main fault-line concerning the nature of emotions. This fault-line runs through all the disciplines reviewed. The question of whether emotions arise from the outside or from the inside is closely aligned with this fault-line, where an outside-in view is associated with cognitivism and an inside-out view is associated with affective theories. It is concluded that the weight of scientific evidence presently appears to be on the side of the affective perspectives on the nature of emotions, even though this still appears to be the minority view.

2.2 Emotions and normativity

Why examine emotions? First, I would like to call attention to a close analogy between how emotions and normativity seem to work psychologically. I drew attention to this in Chapter 1. A similar analogy has been noted by others, such as Etzioni (1988: 93–113), who discusses emotions in terms of “normative-affective factors” that serve as a basis for practical decision-making.

The close analogy is supported by recent evidence. For example, Decety and Cacioppo (2012) find that moral evaluations of harm are almost instantaneous (62 ms), pre-cognitive, emotional, and involve an immediate unreflective distinction between intentional and accidental actions. This indicates that normative assessment and emotion is interwoven. In should be pointed out that, this evidence indicates that certain normative distinctions occur well before the slower cognitive processes of the cortex. This is a striking finding in

suggesting that, at least in some cases, neither emotions nor normative assessment require the involvement of cognition.

The analogy between emotions and normativity has inspired me to look into emotions as a possible *foundation for* normativity, or normativity as a mental *aspect of* emotions. These two ideas intermesh. The practical relevance of emotions for understanding decision-making and behavior in organizations seems clear (Lines, 2005; Fineman, 2000: Ch. 1; Weiss and Cropanzano, 1996). Weiss and Cropanzano (1996: 11) point out that “[t]hings happen to people in work settings and people often react emotionally to these events. These affective experiences have direct influence on behaviors and attitudes.”

I propose that the relevance of emotions may, in part, be due to their inherent experiential normative aspect. Before investigating the relation between emotions and normativity more closely (which I do in Ch. 5), however, it seems prudent to first look into what researchers, scientists, and philosophers offer in the way of understanding the nature of emotions. In this section, I only briefly present the analogy as an explanation for why it is of particular interest to investigate the nature of emotions.

It is widely agreed in the mind sciences that emotions serve as sources of motivation (Panksepp, 1998; Plutchik, 1991). Similarly, it is widely held in philosophy that normative perspectives are motivating (e.g., Broome, 2013). This indicates an overlapping *functional* parallel between normativity and emotions. If emotional motivation is strong enough, and its hindrances are weak enough, it will cause action. Viewed in terms of explicit normativity, sufficient moral or “rational” motivation will cause a moral or “rational” agent to act. Both normative perspectives and emotional feelings can be taken to refer to subjective desires of the agent, although this topic is heavily debated, and both often feature in what agents cite as their reasons for action (Goldman, 2009).

As noted, however, the connection between normativity, on the one hand, and affects and desires, on the other hand, is often not recognized and frequently met with philosophical skepticism and hostility. Similarly, emotions are often given a cognitive interpretation that dismisses kinship with subjective desires. However, it remains a viable theoretical point of view that “to have a desire and to be in the grip of an emotion” can be seen as parallel stories about motivation (Goldman, 2009: 20). This point can be clarified by investigating the nature of human emotions, in particular whether they are appropriately seen as cognitive or affective. Our understanding of normativity can benefit from an improved understanding of emotions.

Without investing too much scientific confidence in intuition, our intuitive grasp of the phenomena of emotion and normativity can at least give us some indication. Intuitively, affect seems to attract or repel human decision-makers much in the same way as normative judgments about “right” and “wrong” do. It is possible that this is no coincidence; philosophical emotivism (see Ch. 9), for example, argues that affective emotions *constitute* or *convey* the normativity of evaluative judgments, including moral judgments (e.g., Blackburn, 1998). There are similar examples among those who take a social constructivist view of emotions, including Prinz (2007), who argues that emotions are both necessary and sufficient for normative moral judgments.

To further illustrate the parallel between affect and normative judgments, consider the different affective attachments we have; while some clearly are fleeting and superficial, others are sustained in deep and durable ways and are considered morally important by the person who has these attachments (such as my affective attachment toward my son). As Goldman (2009: 26) points out, motivation includes not only superficial desires but deeper affective concerns that explain and organize the more superficial ones. In these situations, it is difficult to imagine that the normative judgments of a normal human agent in practical life would routinely and studiously disregard her affective attachments and deeper concerns. For

example, I suppose most people would find it odd for a manager to see no normatively relevant difference whatsoever between how *her* employees in particular would be affected by a decision, compared to how employees in the business community in general would be affected.

Hence, there seems to be a conspicuous functional resemblance between our affects and emotions on the one hand and our normative perspectives on the other. Moreover, the motivational function of normativity and emotions seems to be broadly shared by those who think of normativity and emotions as “cognitive” (e.g., Broad, 1954) as well as by those who hold that the nature of normativity and emotions is “affective” (e.g., Blackburn, 2010; Ayer, 1984). Emotions and normativity seem closely interconnected in the psychology of human beings. It remains more of a pressing question, however, whether normative perspectives and judgments together with emotions should be interpreted as essentially cognitive or affective.

I now turn to look at what the different disciplines inform us about the nature of emotions. The literature on emotions is expansive. While I will not be able to do justice to all the interesting theories that have been proposed across various disciplines, I aim to cover a set of views that are influential, informative, and also relevantly different from each other.

2.3 Contemporary theories of emotion in psychology

Psychology has traditionally been the intellectual epicenter for theories of emotion, although in relative terms it has been more involved in other research areas, notably learning, perception, and cognition. Psychology routinely deals with matters of the human mind, either therapeutically or scientifically, and emotions are an integral part of this picture. The classical philosophical categories of mind (Hilgard, 1980; McDougall, 1908; James, 1892), which date back to Plato, are (1) thinking (cognition), (2) feeling (affect), and (3) effectuation of behavior

(conation). Emotion theory fits into this tripartite taxonomy, evidencing deep-cutting differences with regard to what is taken to be the nature and causes of emotion.

Although there is considerable variety in cognitive theories of emotion, they do share some basic assumptions. Cognitive theories of emotion (e.g., Frijda, 2007; 1986; Lazarus, 1991; 1984a; Roseman, 1984) consider the nature of emotions to be either fully cognitive (strong cognitivism) or partly cognitive (moderate cognitivism) to qualify as “emotions.” They furthermore accept cognitive processes to be causally *necessary* for emotions. These cognitive processes include “appraisal” (e.g., Lazarus, 1991) of the *external* context and “cognitive interpretation” of *internal* sensations (e.g., Posner et al., 2005: 715; Russell, 1980). Many cognitive theories of emotion hold that cognition per definition is explicitly conscious or “declarative.”

Affective theories of emotions stand in stark contrast to the cognitive views. The differences between cognitive and affective theories are not merely terminological but substantive (Griffiths, 1997: 25). This point will become clear when we examine associated processes of the brain as reviewed in the presentation of neurological theories of emotion in this chapter. Affective theories of emotion hold that the *nature* of emotions is fundamentally affective (e.g., Robinson, 2005; Panksepp, 1998; Griffiths, 1997; Ekman and Davidson, 1994; Ekman, 1992; Izard, 1991; Plutchik, 1991; Tomkins, 1991; 1963; 1962; Zajonc, 1984). Moreover, affective theories of emotion propose that central parts of the *causes* of emotions are inherently affective or non-cognitive, and that cognition is not necessary for emotion. There seems to be broad agreement, however, that many of the manifestations of emotions are also shaped by cognitive processes, often to a considerable extent (Panksepp, 1998).

Affective theories differ from each other in several important respects. While some hold that all emotions are non-cognitive (strong non-cognitivism), most of these theories

(moderate non-cognitivism) allow that some emotions are at least partly cognitive (or affective-cognitive). For instance, Panksepp (2007: 281) argues that “most of our everyday emotions are such complex mixtures of primary (feeling), secondary (learning and thinking), and tertiary (thoughts about thought) processes that we can barely see the primary process emotions and affects that contribute to the cognitive jungles of our lives,” emphasizing that tertiary and secondary emotions are fundamentally based on the neural activity of primary affects.

It is also possible to view emotions as inherently conative in that they function as motivation for action. However, this perspective is also consistent with both cognitivism about emotions and affect theory, so conative views of emotion simply represent an additional theoretical option. Indeed, conation and question of motivation are standardly subsumed under the umbrellas of cognitive and affective theories of emotion (Izard, 2010).

Cutting across the classical state-of-mind categories, approaching an understanding of emotions presents special theoretical and methodological concerns (Plutchik, 1991: 24): (1) feeling states or introspection, (2) behavior and overt expression, and (3) physiology and neurology. This understanding suggests the usefulness of an interdisciplinary approach. Appropriately, the psychology of emotions seems to be increasingly turning into an interdisciplinary field of research (Northoff, 2011; Panksepp 2007; Prinz, 2007; von Scheve and von Luede, 2005). Therefore, interdisciplinary approaches are also critical for correctly understanding practices that involve emotions, such as decision-making (see Ch. 6).

Although the study of emotions is a relatively recent development within the field of psychology, researchers have proposed a wide range of different theories of emotion (Izard, 1991: 24). Over the past three decades or so, cognitive theories have dominated psychological research (Furtak, 2010: 51; Panksepp, 1998). Many cognitive theories in general have been

put forward, particularly in emotions research (Reisenzein, 2009: 7). Only a relatively small part of cognitive psychology has previously been preoccupied with emotions (LeDoux, 2002), however. Attention appears to have been diverted away from emotion research by the popularity of heuristics and biases literature (Greenberg, 1990: 108), which led many psychologists to investigate suboptimal cognition rather than emotions to explain judgment and behavior. The hegemony of cognitive theories of emotion, however, is increasingly challenged (Panksepp, 2012; Cromwell and Panksepp, 2011; Zellner, 2011; Izard, 2009; LeDoux, 2002; Solms and Turnbull, 2002).

It should be noted that a number of classical terms significantly overlaps with “emotions” and “affects” in the psychology literature. These notions include “drives” (e.g., Hull), “impulses” (e.g. Freud), and “instincts” (Solms and Turnbull, 2002). In this review, however, the focus will be limited to “affects” and “emotions” because these words suffice both in terms of identifying analytically the relevant theoretical contrasts and in providing a platform for the rest of the argumentative structure of the thesis. At the same time, philosophers suggest that it can be useful to treat terms like “emotion” as “cluster concepts” (Goldman, 2009: 26) or “thick concepts” (Williams, 1989) that overlap and amalgam with related concepts and are world-guided in the sense that their appropriate usage “is determined by what the world is like” (Williams, 1989: 29). This suggests that emotions have an empirical basis and yet are only vaguely or tentatively definable.

“Affect” is conventionally construed as broader than just “emotion” (Matsumoto and Hwang, 2012; Panksepp, 2012: 8; MacLean, 1990), where emotions designate a special category of affects, namely “emotional affects” (Panksepp, 1998). Affect can include other homeostatic sensations like pain and pleasure, and hunger and thirst, in addition to emotional affect (Panksepp, 1998). Buck (1999: 301) defines affects as “subjectively experienced feelings and desires” involving “interoceptive perceptual systems.” It is widely accepted that

feelings are subjective and that they cover a wider experiential domain than specific emotions (Frijda, 2007: 199).

In the psychological emotions literature, “affect” and “emotion” are sometimes used interchangeably, and it has been argued that both notions point to the same evaluative domain (Charland, 2005: 96). As previously noted, some emotions theorists (cognitivists) resist the idea that emotions are inherently or necessarily affective and opt instead for a separate definition of emotion. Some of the most influential theories of emotion in psychology are reviewed below.

2.3.1 Arnold’s cognitive appraisal theory of emotion

Arnold (1960) propounds what she calls an “appraisal theory of emotion.” The notion of “appraisal” denotes a *cognitive* as opposed to an *affective* understanding of evaluation (Izard, 1991). Unlike most cognitive theories of emotion, however, Arnold presents a theoretical perspective where the division line between cognition and affect is somewhat blurred.

Arnold (1960: 82) defines emotion as “the felt tendency towards anything intuitively appraised as good (or beneficial), or away from anything appraised as bad (or harmful),” and suggests a corresponding behavioral output in that such “attraction or aversion is accompanied by a pattern of physiological changes organized towards approach or withdraw.” The structure of the emotional process is thus one that is initiated by “perception,” “apprehension,” or “sensation” of “objects,” which are subsequently “appraised” and finally produce feelings and behavior. An interpretation of what “emotion” means critically depends on what is meant by the phrase, “intuitively appraised” and to some extent on what the appropriate “objects” of emotion are.

In most contemporary appraisal theories of emotion, “appraisal” is taken as inherently cognitive and conscious. Arnold, however, characterizes “appraisal as a sense judgment that is

made automatically and unconsciously” (Greenberg, 1990: 112), and sees the appraisal process as direct and immediate (hence intuitive). This viewpoint indicates an affective rather than a cognitive interpretation of her theory of emotion. Arnold resists an insistence on the presence of consciousness and an element of free will for something to qualify for the label “cognitive.” Making no such requirements, Arnold’s cognitive appraisal theory can in this sense be seen as *cognitive minimalism*. As Greenberg (1990: 112) notes, her theory stops short of embracing the higher-level cognition required for rationality and deliberate judgment. Thus, this approach seems to open up a place for raw affective input into decision-making.

Another interesting point about Arnold’s account is that she recognizes that several psychological processes happen practically concurrently: “Normally, the sequence perception-appraisal-emotion is so tightly knit that our everyday experience is never strictly objective knowledge of a thing; it is always a knowing-and-liking, or a knowing-and-disliking” (Arnold, 1960: 177; cf. Greenberg, 1990: 112). This perspective suggests that her rendering of the steps of the emotional process are not to be interpreted as strictly discrete. We can also note that the classification of Arnold’s theory of emotion as “cognitive” is somewhat dubious because the concept of cognition is stretched rather broadly.

2.3.2 Lazarus’ cognitive appraisal theory of emotion

Richard Lazarus has become an influential emotions theorist in psychology (e.g., Mauss and Robinson, 2009; Dunn and Schweitzer, 2005) as well as in the management literature (e.g., Watson and Spence, 2007), including business ethics (e.g., Henik, 2008). By primarily researching psychology of stress and coping strategies, Lazarus (1991; 1984a) follows in the footsteps of Arnold (1960), expanding on her cognitive appraisal theory of emotion.

Lazarus (1993: 244) analyzes psychological stress in terms of “conflict emotions,” which include anger, anxiety, guilt, shame, sadness, envy, jealousy, and disgust (all negative

emotions). It can be noted that, in the terminology of Panksepp (1998), some of the suggested emotions are primary (e.g., “anger”), whereas others are cognitive-affective emotions of a higher order (e.g., “jealousy”). Lazarus’ account of emotions, hence, does not operate on different levels. Lazarus (1993) makes a case for a positive role of negative emotions; negative emotions can be “good” and even critically important for successful coping. Significantly, Lazarus develops appraisal theory by making an explicit account of the nature of the cognitions (i.e., appraisals). He also creates a more refined version of the theory by specifying requisite antecedent conditions of the cognitions.

According to Lazarus (1991; 1984a), emotional reactions result from “cognitive appraisals.”; “Cognitive appraisal can be most readily understood as the process of categorizing an encounter [or event, or situation], and its various facets, with respect to its [perceived] significance for well-being” (Lazarus, 1984a: 31). This definition of appraisal makes it clear that appraisal denotes a *cognitive* process and that the relevant activity that takes place is *categorization* of things that happen, filtered by salience for well-being.

Lazarus (1991) operates with a remarkably wide notion of “cognition.” According to Lazarus (1991: 6), cognitive activity covers “diverse forms of thought, whether conscious or unconscious,” and, he continues, it “enters virtually everything we do; we use it to pilot our lives and to be responsive to feedback from the environment that could be relevant for survival.” Thus, it appears as if he views virtually the entire human mental apparatus as inherently cognitive, including such fundamental components as emotion, motivation, and attention.

According to Lazarus, emotions in adults are inherently dependent on cognition (Lazarus, 1991; 1984a) and also on the motivation of the given person because motivation is implied in emotion (Lazarus, 1991). For Lazarus, emotions not only supposedly *depend on*

cognition but also upon cognition as an inseparable *part of* emotion (Lazarus, 1984a). In some places, Lazarus argues to the point of making emotions cognitive by definition: “Those less sanguine than I about the causal role of cognition in emotion of often point to the startle response, since cognition is obviously absent or negligible in this reaction. I do not consider startle an emotion” (Lazarus, 1982: 1023). Hence, he rejected Zajonc’s idea that cognition is not necessary for emotion.

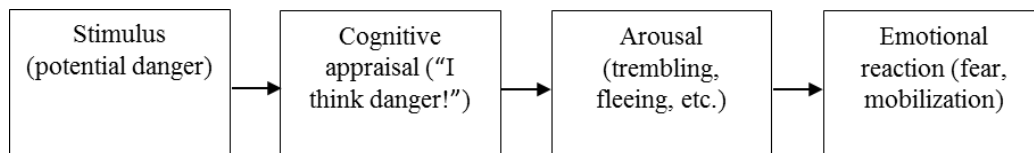
For Lazarus, physiological arousal and emotional experiences happens *after* (and results from) the initial appraisal. The more important the expected outcomes (threats or rewards) are to a person, the more intense the final emotional experience. Lazarus therefore considers *affect* to be *post-cognitive* (Lazarus, 1991; Brewin, 1989; Lazarus, 1984a) as it is elicited only after a certain amount of cognitive processing of information has taken place. In contrast, affective psychological theories of emotion (e.g., Zajonc, 1984; 1980) hold affect to be *pre-cognitive*. To support his position, Lazarus (1991) suggests that “innate reflexes” are separate from affects and cannot be experienced pre-cognitively.

It should be noted, however, that in some places Lazarus departs from the conception of a unidirectional emotional process, suggesting instead that cognition and emotion are mutually interdependent factors with causality pointing both ways (e.g., Lazarus et al., 1982). Nevertheless, Lazarus generally conceives of and has explicitly argued for the occurrence of emotion only as after appraisal and as dependent on the objective of appraisal having personal meaning in light of a person’s goals and values (e.g., Lazarus, 1984a).

Lazarus’s theory of emotions suggests the following emotional process (Fig. 2): (1) stimulus (potential danger), (2) cognitive appraisal (“I think this is dangerous!”), (3) physiological arousal (trembling, fleeing, etc.), and finally, (4) subjective emotional reaction (fear). As noted, however, Lazarus is occasionally ambiguous about the directions’ causality,

and it is also somewhat unclear whether or not Lazarus allows concurrent occurrence or the reverse order for the last two items.

Fig. 2–Lazarus’s cognitive appraisal theory of emotion:



2.3.3 Zajonc's affective theory of emotion

Zajonc argues that not all emotions involve deliberate thinking or labeling. Cognition and conscious awareness of what is happening, therefore, does not offer a complete theory of emotions. By definition, cognitive science is not a science of the whole mind but only its cognitive parts (LeDoux, 2002: 24; Ekman and Davidson, 1994; Zajonc, 1984). Affect, therefore, can be seen as the crucial characteristic that strictly cognitive theories leave out in their theories of emotion.

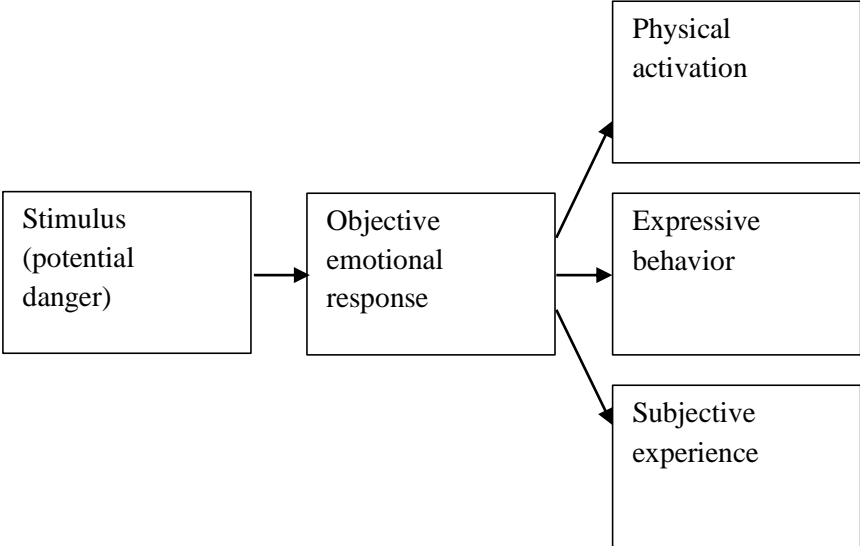
According to Zajonc (1984; 1980), cognitive theories of emotion in some cases have gotten the factual structure of the emotional process exactly back-to-front; the general picture should instead be that *first* we feel and *then* comes the thinking (in case deliberation is involved). In defense of this position, Zajonc cites studies from empirical psychology and presents a number of intuitively compelling common-experience arguments.

Zajonc's arguments include (1) "the irrevocability of affective judgments," which refers to the difficulty of altering one's initial emotional reaction when later presented with information suggesting that it was not justified, (2) "the difficulty of verbalizing affective reactions," referring to our difficulty of verbalizing our reasons for why we dislike or like

something or someone, and (3) “the separation of affective reactions from content,” which refers to the strength of our ability to remember our emotional reactions to something, such as a particular book, even when our memory of the content has faded away.

Zajonc “concludes that affective and cognitive processes are under the control of separate and partially independent systems” (Greenberg, 1990: 120), arguing that the evidence suggests “that affective judgments may be fairly independent of, and precede in time, the sort of perceptual and cognitive operations commonly assumed to be the basis of these affective judgments” (Zajonc, 1980: 151). Fig. 3 shows a rudimentary sketch of Zajonc’s theory:

Fig. 3–Zajonc’s affective theory of emotion:



Zajonc (1980) makes an important theoretical contribution to the emotions literature by formulating what he terms the *affective primacy hypothesis*, which states that positive and negative affective reactions can be evoked with minimal stimulus input and virtually no cognitive processing. This hypothesis has undergone rigorous testing (Murphy and Zajonc,

1993) in which the effects of affective and cognitive priming were compared and on which basis empirical support has been supplied. Sherman et al. (1998) empirically documented the perseverance of induced affective preferences that Zajonc hypothesized. Affect thus appears to arise earlier than cognition and also has longer lasting effects.

However, we could not hope to finally confirm or discredit Zajonc's affective theory until we looked inside what was actually happening in the brain. Examining the neurophysiological and neurochemical processes inside the brain in sophisticated ways has just recently become possible. In hindsight, Zajonc's (1980) position that affect and cognition must be essentially separable processes and that affective experience can arise essentially (and even completely) without cognitively elaborated input have been strengthened and elaborated upon in more detail by evidence (e.g., Decety and Cacioppo, 2012; Jiang and He, 2006; Pasley et al., 2004) and theoretical progress in the field of neuroscience. Similarly in the case of perseverance of emotions, clinical research shows the preservation of emotion and emotion-based memory in patients who suffer profound impairments in terms of cognitive memory (Evans-Roberts and Turnbull, 2011; Turnbull and Evans, 2006).

Recent findings in neuroscience (Panksepp and Biven, 2012; Panksepp, 1998; LeDoux, 1996) regarding key points of Zajonc's theory of emotions may even present decisive support. This body of evidence is discussed in greater detail later on (see Ch. 4), so at this point I will highlight three key points.

Current evidence shows that (1) humans who are born without any functional cortex still appear to express the entire range of documented primary emotions (Solms, 2013: 108; Panksepp, 1998), (2) expression of primary emotions in humans, as well as other mammals, can be induced reliably by simple electrical stimulation of the corresponding emotional centers in the subcortical brain (Panksepp and Biven, 2012; Shapira et al., 2006), and (3)

separate but interacting neurological circuits have been identified, suggesting that cognition and emotions are separate but interacting processes (LeDoux, 2002; Panksepp, 1998; LeDoux, 1996).

2.3.4 Plutchik's evolutionary theory of emotion

Emphasizing the evolutionary perspective, Robert Plutchik represents an updated and sophisticated psychological version of Darwin's position on emotions. Plutchik was one of the emotion theorists who argued early on that we need to look for the foundations of emotion in biology and that there is a correlation between activity in the nervous system and psychodynamic observations (Plutchik, 1986). Plutchik (2003; 1991) proposes eight "basic emotions," and argues that they are handed down through the process of evolution based on their ability to promote survival and reproduction.

As will be discussed later (Ch. 4), "basic" emotions are underpinned by a key finding in neuroscience, although the term "basic" itself challenges us to define what exactly makes something basic. Perhaps "primary emotions" is a better term (Panksepp and Biven, 2012) as it highlights the relevant causal flow in the emotional process, which is required to locate the sources of emotion. In support of his theory of basic emotions, Plutchik (2003) shows empirically that emotions map onto a limited set of dimensions.

For Plutchik there is an intimate relationship between cognition and emotion, and he proposes that cognition evolved in order to treat basic emotions and biological needs so that the efforts they represent are met more optimally (Greenberg, 1990: 117); "Plutchik defines emotions as a patterned bodily reaction corresponding to one of the underlying adaptive biological processes common to all animals" (Izard, 1991: 39).

Based on a review of the theories and available evidence concerning emotions, Plutchik (1991: 41–42) proposes six clear postulates about emotions:

- (1) There is a small number of pure or primary emotions.
- (2) All other emotions are mixed; that is, they can be synthesized by various combinations of the primary emotions.
- (3) Primary emotions differ from each other with regard to both physiology and behavior.
- (4) Primary emotions in their pure form are hypothetical constructs or idealized states whose properties can only be inferred from various kinds of evidence.
- (5) Primary emotions may be conceptualized in pairs of polar opposites.
- (6) Each emotion can exist at different degrees of intensity or levels of arousal.

These postulates, most of which seem reasonable in light of what has been discussed thus far, provide input for a more detailed analysis of emotions. Furthermore, they have the advantage of being empirically testable, at least in principle. I will briefly comment on these postulates and their place within the current review.

The first postulate is in line with many other notable emotion theorists, as seen in the summarizing table (*Table 1*) later in this chapter. The same is true of the second postulate, although or perhaps most theorists will argue that not only affective primary emotions factor in here but also a variety of cognitions (expectations, beliefs, etc.) and cognitive structures (patterns of reasoning, etc.). Postulate 3 seems necessary for making a complete account of emotions, which is then taken as encompassing potentially discernable differences in body, mind, and behavior. Postulate 4 expresses a commitment to evidence as the primary scientific guide to truth.

The fifth postulate is somewhat more dubious. The present review does not grant any particular reason to suppose that emotions come in pairs of “polar opposites.” On the contrary, we are led to suppose that the range of human emotions will be less neatly ordered

than this because it will represent approximately what has been adaptively functional and selected for over the evolution of our species. Postulate 6 seems intuitively plausible as well as consistent with what theorists have inferred from empirical observation.

2.3.5 Other psychology-based theories

There is a substantial number of interesting theories regarding emotion in psychology; inevitably, I shall not be able to do justice to them all. In this section, I mention further theories that have been developed within psychology in recent times and relate them to the theories that have already been discussed.

Nico Frijda (2007; 1986) argues for a cognitive appraisal theory where the concept of action readiness plays an important part in the emotional process. Frijda (2007: 98; 1986) views appraisal as an antecedent to emotion; in this respect, his theory bears close kinship with that of Lazarus (1991). For Frijda (1986), emotions are the tendencies a person has to establish, maintain, or disrupt relations with his or her environment. These “tendencies,” he holds, can equivalently be termed as “action readiness.” Frijda (2007: 1986) proposes that there is a definable set of basic emotions and has made several suggestions as to how many emotions should be included (Frijda, 1987; 1986). Although identifying emotions as forms of readiness seem too reductive as an account of emotions, both the idea that action readiness is central for understanding basic emotions as well as the idea that there are such basic emotions enjoy considerable support from research in neuroscience (Panksepp, 2005).

There is a group of theories that view emotions as socially constructed. Well-known social constructivist theories of emotion include those of Bodor (2004), Harré and Parrott (1996), Harré (1986), and Averill (1980). Averill (1980), for example, holds that emotions are cognitive but at the same time are constructed socially by the continuous social exchanges taking place in society in conformance with cultural norms. According to this view, cultural-

cognitive structures guide the appraisal of stimuli, the organization of responses, and the monitoring of emotional behavior (Greenberg, 1990: 118). Harré and Parrott (1996), meanwhile, argue that emotions, the *displays* of emotions, as well as the way they *feel* (Harré and Parrott, 1996: 1), are (at least to some extent) “socially constructed.” In this regard, therefore, emotions vary from culture to culture, over time, and from one group of people to another.

An integrative, interactionist perspective on emotions was proposed by Buck (1999) in which emotions are seen as cognitive “read-outs,” where affects have to be labelled by cognitive processes to become proper emotions. What are termed “emotions” in these theories are essentially what Panksepp (1998) defines as the “tertiary layer” of emotions.

Russell (2003) has proposed a theory sometimes called “the core affect mode” or “the circumplex model of affect” (Russell, 1980). This dimensional model looks at emotions structurally in terms of two bi-polar dimensions (valence and arousal): “activation” vs. “deactivation” on one axis (1), and “pleasant” vs. “unpleasant” on the other (2). According to Russell (2003: 147), “core affect is that neuropsychological state consciously accessible as the simplest (nonreflective) feelings evident in moods and emotions,” where core affects influence memory through what he calls an “attribution effect.” Russell’s sense of “core affect” is similar to Damasio’s (2010) notion of “core affect,” and Panksepp’s (1998) term “primary affect” but without a finer resolution of discrete affect emotions. In this way, “affect” is similar to “subjective experience.”

Some theories of emotions emphasize their evolutionary origins (e.g., Keltner et al., 2006; Nesse, 1990). While the specific question as to exactly how emotions might have evolved does not need to concern us at this point, such explanations can help us understand why we experience the specific emotions we do. Evolutionary theories of emotion tend to

emphasize specific problems that our hominoid ancestors presumably faced. They are difficult to pin down given the limited empirical data we have about them (Solms and Turnbull, 2002), although the idea that specific emotions are related to their “survival value” seems reasonable and central for understanding how they came to be as they are at present (Panksepp and Biven, 2012; Damasio, 2010; Panksepp, 1998).

Keltner et al. (2006: 121) proposes that emotion “guilt” is an evolutionary response to reining in cheaters and that “envy” promotes action against individuals who enjoy an unjustified favorable status. Each of these emotional response patterns seem to sediment social cohesiveness, which may foster group-level efficiency. Furthermore, Keltner et al. (2006: 119) suggest that “desire” and “love” have been (and still are) evolutionarily important for finding a mate, while “compassion” and love of children (Keltner et al. 2006: 120) are important for protecting offspring.

Similarly, Nesse (1990: 272) suggests that “social anxiety” is a response to the prospect of “social ridicule”; “pride” and “humiliation” regulate the reciprocity of exchanges (Nesse, 1990: 276–277), “anger” is a deterrent against cheating (Nesse, 1990: 277); and “anxiety” disposes one against initiating cheating behavior and failing to meet expectations (Nesse, 1990: 278). Nesse also makes a similar point as Keltner et al. (2006) concerning the social role of the emotion “guilt.” While these theories are interesting, one inherent problem is that it is difficult to test their correctness and accuracy in the absence of a sufficiently detailed and verifiable empirical background. How could we know what social interactions were like a few million years ago with a requisite degree of specificity and certainty?

2.4 Contemporary theories of emotion in neuroscience

As a basis for understanding moral issues and moral decisions, practical ethics frequently draws on discussions in psychology concerning emotions, whereas neuroscience has received

comparatively much less attention. This trend is perhaps understandable since practical ethics and psychology are both social sciences, whereas neuroscience belongs to the natural sciences and is often communicated in more technical terms. As we have seen, psychology only occasionally explicitly builds its theories of emotion on the findings of neuroscientists, which suggests that there may be a pervasive shortfall of evidence from neuroscience being brought to bear on the practical issues of practical ethics. Given that neuroscience offers a second “viewpoint” on all psychological issues (Solms and Turnbull, 2002), this shortfall represents an area of opportunity for further disciplinary integration.

Neuroscientists have investigated the brain to gather a substantial amount of detailed data to understand how the brain works, how it is organized, and what physiological and chemical processes it sets in motion. Still, many details remain unclear and poorly understood. Nevertheless, neuroscience has seen an increasingly stronger platform of common knowledge emerge, including an understanding of the underpinnings of affect and emotion. For example, we know that basic emotional expressions stem from neural circuitry in the subcortical regions of the brain (Panksepp and Biven, 2012; Winkielman and Berridge, 2004; Whalen et al., 1998; Panksepp, 1998: 307; LeDoux, 2002; LeDoux, 1996; MacLean, 1990). While this position is not universally accepted or even always known among neuroscientists, the empirical evidence is quite solid (Merker, 2007; Shewmon et al., 1999; Panksepp, 1998).

Emotional centers have been uncovered in the evolutionarily (phylogenetically) older parts of the brain, and their functioning in producing emotional expressions is reported to be robust and repeatable with high reliability (Panksepp and Biven, 2012). Specifically, seven separate but interacting emotional circuits in the brain have been located (Panksepp, 1998: 52), each associated with distinct patterns of emotional behavior (see Ch. 4 for details). These emotional circuits have been presented in physiological as well as neurochemical detail.

While the scientific map in this area is still being charted, considerable detail is known about which specific chemical neurotransmitters act to bridge the synapses between the nervous system cells and how they do so (LeDoux, 2002; Panksepp, 1998). Humans are not alone in being equipped with emotional circuitry; all other mammals that have been extensively studied exhibit the same main neurotransmitters and an almost identical subcortical brain physiology (Solms and Panksepp, 2012: 148; Panksepp, 1998: 100; MacLean, 1990).

Findings from animal research are strongly indicative of a human emotional registry being generated from subcortical brain regions (Panksepp, 2011). For example, other mammals with which we are closely related in respect to brain anatomy, when surgically decorticated, nevertheless appear to be capable of expressing each of the seven identified types of emotional behavior (Kolb and Tees, 1990). As noted, this ability is also present in humans born essentially without a neocortex (Panksepp, 1998), a condition sometimes resulting from a lesion at a certain fetal stage. In such cases, there are sometimes traces of neocortex left, but since these are completely dysfunctional, the evidence is strong (Solms and Turnbull, 2002).

Traditionally, emotions and their affects have been less emphasized in brain research; “Because of the paucity of relevant discussions in neuroscience, few explicit and testable theoretical proposals concerning the nature of affect have been placed on the intellectual table” (Panksepp, 2005: 160). Affective neuroscience is now becoming more of a hot topic of research as illuminating results and convincing evidence pave the way (Bingman, 2011; Davidson, 2003).

Neuroscience, like psychology, can be divided into cognitive neuroscience, affective neuroscience, and more behaviorally inspired approaches. In contrast to the field of psychology, however, most theories of emotions in neuroscience refrain from proposing that

emotions are cognitively based. It is generally acknowledged that the cortex alone cannot account for how emotions arise or how they play out; indeed, the cortex for some emotional processes need not play any role at all (LeDoux, 2002: 201).

The label “cognitive” is typically used simply to demarcate the study of the *effects of* emotions on cognition (i.e., memory, attention, learning, problem-solving, etc.; Deak, 2011: 73). While the subcortical regions of the brain are mostly special-purpose and genetically predetermined areas, the cognitive regions (cortical regions) appear to be more similar and intended for general-purpose computing unit at birth (Panksepp and Panksepp, 2000: 108). Much of the agreement in the neuroscience concerning emotions surrounds the position and insights of Paul MacLean (LeDoux, 2002; Damasio, 1999; Panksepp, 1998). As LeDoux (2002: 212) points out, “In particular, the notion that emotions involve relatively primitive circuits that are conserved throughout mammalian evolution seems right on target.” Some of the most important and perhaps most well-known theories of emotion in neuroscience are presented below.

2.4.1 LeDoux’s high route and low route of fear

LeDoux (1996) views emotions as essentially thoughtless and reflexive processes, over which we exercise only *post hoc* and partial cognitive control. LeDoux (2002: 206) argues that the amygdalae are the key emotional centers in the human brain for all fear-panic emotions. These centers channel stimulus inputs, either for direct behavioral responses or for a cognitively processed response. Emotions are “the process by which the brain determines or computes the value of a stimulus” (LeDoux, 2002: 206), automatically and thoughtlessly.

The process is as follows: (1) stimulus, (2) emotion, and (3) “feelings emerge as we become aware that our brain has determined that something important is present and we are reacting to it” (2002: 206). This delineation is structurally in agreement with Zajonc’s model

presented above; cognition comes only *after* emotion. However, it should be noted that LeDoux (2000; 1996) is a cognitivist about emotions in the sense that he thinks that feelings occur only after our cognitive awareness of them has taken place. Thus, the affective and feeling aspects of emotions are essentially kept outside the emotional process until the emotions are handled by higher cortical areas. LeDoux (1996: 150) argues that several basic emotional expressions, such as the characteristic freezing posture resulting from fear stimuli (e.g., a snake) in humans, are almost certainly generated by subcortical neural pathways alone. The neurological explanation of the emotional process is described as follows: “Information received by sensory systems activates emotional-processing circuits, which evaluate the meaning of the stimulus input and initiate specific emotional responses by triggering output circuits” (2002: 206). The quickest possible route for input flow from the emotional stimulus to reach the polymodal cortex (where cognition takes place) is first via the sensory thalamus, then the sensory cortex, and after that to the cognitive parts of the brain. By that time, the amygdala and several other emotional centers are busy responding.

LeDoux’s work mainly concerns research on the particular emotion of “fear,” which incidentally is the most extensively studied and mapped out emotion. His research indicates how other emotional systems might work as well, although this question raises certain issues about external validity. LeDoux (2002) found evidence for two different pathways in the processing of fear, which he termed “the high route,” and “the low route.” It should be noted that in both cases, emotion and its associated feelings are the outcome of the emotion process. The two routes can be depicted as follows:

Fig. 4–LeDoux’s fast route:

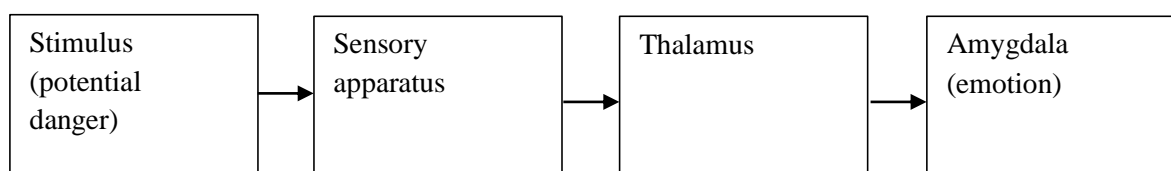
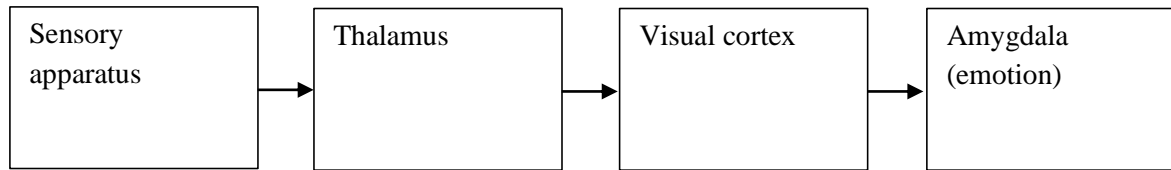


Fig. 5–LeDoux’s slow route:



2.4.2 Panksepp’s psychobiological theory of emotions

While Panksepp first and foremost is a neuroscientist, his extensive works also cover practical applications in psychology, psychotherapy, and psycho-pharmacy, as well as contributions to scientific methodology and philosophy of science (e.g., concerning consciousness). Several decades ago, Panksepp (1982) proposed an evidence-driven psychobiological theory of emotions, which neurologically accounted for four basic emotions. This theory has now been expanded to cover at least seven emotional centers in the brain (Panksepp and Biven, 2012; Panksepp, 1998; 1992). Each of these centers is presented in more detail later on (see Ch. 4).

According to Panksepp (1998), it is vital for progress in emotions research to understand (1) that the brain produces primary emotions, (2) that the brain is layered according to evolutionary development and operates accordingly, and (3) that the subcortical regions of the brain are specifically responsible for generating these “raw affects” or “primary emotions,” which are often called “unconditioned responses” (in classical behavioristic terms).

From this perspective, much of what the psychology literature calls “emotion” is in fact cognitively processed upshots of primary emotions. Essentially, everything that is

cortically based is learned (Panksepp, 1998). Primary affects, in contrast, cannot be learned but instead are built into the brain at an early fetal stage (Solms and Turnbull, 2002). The affective theory of emotions has stood up well to empirical scrutiny and now covers *seven* identifiable emotional systems in humans (Panksepp and Biven, 2012; Panksepp, 2004; Panksepp, 1998). The current state of research does not rule out the possibility of discovering further emotional systems; theoretically, “social dominance” (Toronchuk and Ellis, 2012) and “disgust” (Prinz, 2007) would not be unlikely candidates, but thus far only seven systems are well-documented. In broad strokes, Panksepp structures the emotional processes consistently with LeDoux (2002), as described above.

According to Panksepp (1982: 407), emotions “arise ultimately from hard-wired neural circuits in the visceral-limbic brain that facilitate diverse and adaptive behavioral and physiological responses to major classes of environmental challenges.” Panksepp (1998) further argues that defining emotions in terms of neural circuits is as close as science presently gets to defining what emotions actually are. Again, this approach does not preclude the possibility of more precise scientific definitions of emotions (or overlapping concepts) in the future. Solms (2013: 107) specifies that affect is an intrinsic property of the brain that is expressed in emotions, which is in agreement with the picture presented by Panksepp..

Emotional systems that are of particular relevance to practical ethics, normativity, and decision-making are discussed in more detail later in the thesis (Ch. 4 and Ch. 6), including systems that relate to “moral emotions” and “pro-social emotions.” Since emotional systems interact with one another in complex ways, each of the primary systems is likely to have some general relevance for a given practical domain, such as business ethics or medical ethics. Here I will focus on Panksepp’s general criteria for identifying an emotional system. Panksepp (1998: 48–49) presents the following seven criteria:

1. The underlying circuits are generally predetermined and designed to respond unconditionally to stimuli arising from major life-challenging circumstances.
2. These circuits organize diverse behaviors by activating or inhibiting motor subroutines and concurrent autonomic-hormonal changes that have proved adaptive in the face of such life-challenging circumstances during the evolutionary history of the species.
3. Emotive circuits change the sensitivities of sensory systems that are relevant for the behavioral sequences that have been aroused.
4. Neural activity of emotive systems outlasts the precipitating circumstances.
5. Emotive circuits can come under the conditional control of emotionally neutral environmental stimuli.
6. Emotive circuits have reciprocal interactions with the brain mechanisms that elaborate higher decision-making processes and consciousness.
7. The emotional circuits must be able to generate affective feelings.

Panksepp (1998: 48) points out that “from the perspective of affective neuroscience, it is essential to have neutrally based definitions that can be used equally well in brain research and in the psychological and behavioral studies we conduct on mature humans, infants, and other animals.” This approach offers an open-minded and scientifically based theory of emotions that looks promising as a basis for gaining an inter-disciplinary foundation for emotions.

2.4.3 Rolls’s informational theory of emotions

Edmund Rolls (2005) presents a neuroscientific theory of emotional affects, which is cognitive in the sense that while affective experiences are not themselves cognitive, they are *caused* by higher cortical cognitive processes.

Rolls (2005) explains the emotional process as follows: Sensory *information* (e.g., taste, texture), as opposed to affect, is first sent from the old mammalian brain structures to (evolutionary) old cortical regions (the basal ganglia). This old cortical area issues certain behavioral commands (e.g., continue eating) and additionally passes information up to the higher cortical structures of the (relatively newer) neocortex. The neocortex engages in conceptual and linguistic transformations of the informational input, which finally results in affective experiences.

The informational depiction of the emotional process leads Rolls to draw the conclusion “that ‘unintelligent’ species have no emotional experiences” (Panksepp and Biven, 2012: 72). The various emotional expressions of other less cortically furnished animals are in this view totally distinct from human emotional expressions in that theirs are not accompanied by affective experiences while ours are; for example, we feel playful joy, while lower animals such as dogs do not (they just appear as if they do).

2.4.4 Damasio on emotions

Damasio (1999: 79) defines emotions functionally as chemical and neural patterns that are in place for the purpose of maintaining life and prompting adaptive behavior. As Damasio (1994) points out, emotions are necessary in order to make good decisions in a social world; rationality alone is blind and unable to correctly identify what matters to us. While Damasio (1994) ascribes a central role to cognitive regions of the brain in the emotional process, he has since taken the view, similar to Panksepp, that emotions stem from subcortical affective structures (Damasio, 2010: Ch. 5).

Damasio (2010) views emotions and values as fundamentally biologically based. The reason we have emotional systems, he explains, is that they have had an evolutionary survival value. He views the subconscious homeostatic regulation of tissue as important in the

generation of emotions; “the physiological state of a living organism's tissues, within an optimal homeostatic range, is the deepest origin of biological value and valuations” (Damasio, 2010: 48). In addition, “when tissues operate in the best part of the homeostatic range, the brain mapping of the related states is experienced with a quality we eventually named pleasure and reward” (Damasio, 2010: 52).

2.4.5 Other neuroscientific theories of affect and emotion

There are a number of theories of affect and emotion in the neuroscience literature, and I can only include some of them. While those mentioned previously appear to be the most influential theoretical perspectives, excellent work has also been done by other researchers.

Based on an extensive body of research on epileptic patients as well as various animal experiments spanning many decades, MacLean (1990: 422) proposes that affect represents “subjective forms of psychological information” and surmises that there are three types of affect (MacLean, 1990: 427), which he calls “basic affect,” “specific affect,” and “general affect.” Basic affects arise from to basic homeostatic needs, such as the need for “air,” “food,” “water,” and “sleep.” Specific affects refer to pain and pleasure in localized areas of the body. Finally, general affects refer to emotions, specifically “fear,” “anger,” “dejection,” “desire,” “gratulence,” and “affection.” MacLean views these general affects as causally linked to central moral aspects of our lives, such as our ability to have conviction and conscience. Many recent neuroscientists, including the aforementioned Joseph LeDoux, Jaak Panksepp, and Mark Solms, in many ways represent modern updated versions of MacLean’s work.

2.5 Contemporary theories of emotion in philosophy

Emotion has been a richly discussed topic in philosophy through the ages, and it still is. The philosophical literature will only be briefly reviewed, since a more focused presentation will be presented in the chapter on emotivism (Ch. 9). Philosophers, like psychologists, can roughly be divided into cognitivists about emotions and non-cognitivists, who for the most part view emotions as affective. The line between cognitive and non-cognitive views is drawn in slightly different ways (how wide, for example, is the scope of cognition?), and some theories do not neatly classify. Still, the bifurcation of the theory of emotion into two main camps is conspicuous. Notable contemporary cognitivist philosophers who have discussed the theory of emotion include Goldie (2012), Solomon (2007; 1988), Green (1992), Nussbaum (1990), Gordon, (1987), and Kenny (1963), but a number of philosophers could be added to this list. Notable contemporary non-cognitivist philosophers include Blackburn (2010; 1998), Goldman (2009), Robinson (2005; 1995); Griffiths (1997), Gibbard (1992), Williams (1989), Ricoeur (1986; 1966) in a partial sense, and Ayer (1984).

2.5.1 Solomon's cognitive theory of emotions

The thoughts of Robert Solomon have been influential in psychology (e.g., Frijda, 2007) and philosophy, as well as in business ethics, to which he also has contributed. Solomon (2007; 1993; 1988) proposes a strong version of cognitivism about emotions. In this view, emotions are (cognitive) *judgments* about our self and our place in the world (Solomon, 2007: 203; Solomon, 1988). The emotion “anger” thus *corresponds to* a judgment: for example, the judgment that one has been wronged. Emotions are, in effect, conscious acts.

Solomon (2007) presents an *externalist* view of emotions, claiming that emotions are not inside our mind but outside in the external social world instead. He takes a bleak view of disciplines that focus on studying the internal foundations of emotions, such as neuroscience, claiming that it is reductive and that “[n]europhysiological reductionism is but the most

current and most exotic of the various theories with which we distance ourselves from our own emotions” (2007: 122).

Apart from developing his own “judgmental” theory of emotions, Solomon (2007) sets out to debunk a series of what he claims are “myths” about emotions: (1) the myth that emotions are ineffable, (2) emotions are feelings, (3) emotions pass through us (i.e., “the hydraulic model”), (4) emotions are in the mind, (5) emotions have no or require no intelligence, (6) emotions can be seen in terms of positive and negative valence, and (7) emotions are irrational.

Solomon’s theory has a number of implications (Solomon, 2007). One of the most important moral implications is that we are fully *responsible* for our emotions. Thus, for instance, we may have a moral duty to grieve at someone’s funeral (Solomon, 2007: 75). Anger is not something that happens to us and takes control over us but instead is “a matter of conscious choice.” We act as purposeful agents and use emotions, such as anger, as a strategic tool for coping in the world. Again, it follows that we are fully responsible for our emotions. This same concept also applies to love because falling in love “is a process of willful escalation” (Solomon, 2007: 194).

Finally, Solomon (2007) sees these conscious and agent-controlled emotions as central for having “a life worth living,” to “ethical integrity,” and to the “pursuit of wisdom.” Robert Solomon presents a set of views that purports to challenge scientifically based theories of emotion on a number of accounts, centrally including the proposed “myths” highlighted above.

2.5.2 Ricoeur's theory of emotions and feelings

Philosopher Paul Ricoeur makes an important contribution to conceptualizing emotions and feelings mainly from a linguistic angle. Ricoeur proposes a dualism involving the mind and brain that, while resembling Descartes's, is more modest and less committing (dualism is discussed in more detail in Ch. 3). Earlier works by Ricoeur view emotions as a source of visceral disturbance and a threat to self-possession (1966: 252-253), whereas later works (Changeux and Ricoeur, 2002) take a more positive view in the sense that emotions are central *for* the construction of self and morality. Emotions are also important because they sometimes summon the will (Ricoeur. 1966: 260).

Ricoeur suggests that the accounts of emotion we find in cognitive psychology frequently suffer from a deficiency of clear conceptual distinctions. One of the problems, he maintains, is that emotions and feelings are often used interchangeably. A clear distinction between the two notions is crucial because they refer to different phenomena. This idea resonates with Buck's (1999: 301) observation that "it is sometimes unclear how what is labeled *emotion* at one level is related to what is termed emotion at another."

According to Ricoeur (1986; 1966: 250–280), emotion is a *first-order*, non-cognitive, experience. Feeling, in contrast, is a *second-order*, intentional structure, in which cognition plays a central role. Ricoeur recognizes that emotions are physical and suggests that they look inward towards the self and its needs. Feelings, according to Ricoeur (1986: 83), are viewed as simultaneously outward looking and inward directed. When we react to something, we project something outwards *from* ourselves, as well as something inward to ourselves, bringing about specific inner states.

Ricoeur views emotion as "an involuntary which *sustains* voluntary action, which *serves* it in preceding it and limiting it" (1966: 251). The phenomena of emotion are, in this

view, best seen as instances of a “dialogue of the voluntary and the involuntary” (ibid: 280). In this way, he follows the structure of William James (whom he refers to), proposing that cognition comes *before* emotion, but (unlike cognitivists about emotion) *without* the emotions thereby being cognitive themselves. This view makes it possible to see emotions as fundamentally affective and subconscious processes of the brain, with which we are biologically equipped as a result of evolution. Thus, Ricoeur’s view of emotions appears to be cognitive in that emotion is born out of cognition but non-cognitive when it comes to the content of the emotions.

2.5.3 Broad’s quasi-cognitivist theory of emotions

An early example of (apparently) explicit cognitivism regarding emotions can be seen in Broad (1954: 206): “Every emotion is an epistemologically objective or intentional experience, i.e., it is always a cognition, either veridical or wholly or partly delusive.” However, he opens the door for something *more* than mere cognition by adding a further point about “emotional tone”: “To be fearing a snake, e.g., is to be cognizing something—correctly or incorrectly—as a snake, and for that cognition to be toned with fearfulness” (Broad, 1954: 206). Despite his cognitivist rhetoric, this approach seems to lend itself to a non-cognitivist interpretation, for what is the nature of this fearfulness that cognitions are toned with? Broad himself answers this question by explaining that emotions have a “cognitive aspect” as well as an “affective aspect” (Broad, 1954: 209). This view, then, appears to suggest that emotions are partly affective but that the affective part of the nature of emotions is somehow intentionally experienced.

Broad (1954) brings up several interesting distinctions that potentially can serve as tools for analyzing or reflecting on emotions and are based essentially on the careful systematization of input from personal introspection. According to Broad (1954: 206),

emotions can be motivated or unmotivated, misplaced, appropriate or inappropriate, first-hand or second-hand, and finally pure or mixed. The first distinction between motivated and unmotivated emotions separates general likes (approvals) and dislikes (disapprovals), such as the feelings that all animals (of a certain sophistication) might have towards specific objects, from the more sophisticated emotions that are particular to human beings. In addition to general affective dislikes, these emotions also cognitively connect these attributed feelings with specific *attributes* of the object in question.

The second distinction, which is between emotions that are misplaced and those that are not, refers simply to the question of whether the objects to which emotions are attributed actually exist or not and then whether or not their attributes exist. If the objects of emotions do not actually exist, the emotions are said to be misplaced. If key attributes attributed to the objects do not exist, then it is said to be a case of emotions being partially misplaced.

The third distinction between appropriate and inappropriate emotions concerns the link between affect and cognition. Broad (1954: 209) explains that emotions can be fitting or unfitting; for example, it would be fitting to respond, or “cognize” a frightening object with some degree of fear. From this perspective, epistemological objects call for or require specific types of affective responses. This leads to central questions of what are *morally* appropriate and inappropriate responses in practical situations and decision-making tasks. While Broad does not inform us, at this point, about whether he sees appropriateness as objective or subjective, he defends the latter position elsewhere (Broad, 1971); defending normative ethics based on “self-referential desires” against reliance on theoretically constructed objective moral principles.

The fourth distinction between first-hand and second-hand emotions is associated with the directness of the evaluation of the object of an emotion. While first-hand emotions

concern the perceived object with its perceived attributes directly, second-hand emotions represent cases in which broader “emotional moods” and associations are brought into the evaluative process. Second-hand emotions seem likely to be especially important in practical terms when we look at how emotions might bias decisions; as Broad (1971) notes, second-hand emotions can sometimes be so strong as to replace (would-be) first-hand emotions altogether.

Finally, the fifth distinction between *pure* and *mixed* emotions denotes a key distinction between primary biologically based emotions (the “pure” ones) and secondary emotions (the ones that are “mixed”). According to Broad (1954: 210), “we may fairly assume that there is a certain fairly small number of primacy species of emotional tone, just as there is a limited number of primary colors, and that the vast majority of human beings are born with dispositions corresponding to each of them.” He refers to these dispositions as “primary emotional dispositions.” As the summary of the literature review shows, this is a position on which contemporary psychologists and neuroscientists converge with increasingly empirical backing.

2.5.4 Prinz’s constructionist theory of emotions

From a constructionist point of view, Prinz (2004: 204) holds that “emotion is a form of perception,” more specifically, a *bodily* form of perception, and claims that: “Having an emotion is literally perceiving our relationship to the world” (Prinz, 2004: 204). How does Prinz arrive at this conclusion?

He recognizes that not all emotions can involve cognition since the evidence clearly shows that some emotions arise without the intervention of the neocortex. However, Prinz argues that emotions are meaningful. From this inference, he concludes that meaningfulness cannot depend on cognition. However, meaning is commonly thought of as cognitive. Thus,

he views this conclusion as *paradoxical*; a paradox he calls the “emotion problem” (Prinz, 2003: 78), adding: “Noncognitive theories give us too little, and cognitive theories give us too much.”

Prinz’s solution to the “emotion problem” is to propose that there must be something like sub-cortical emotional “inner concepts” that are active in perception. Morally, he thinks that emotions can be either justified or unjustified. He takes one of the central functions of emotions to be to provide us with information. This information can be critically important for survival because it allows a rapid response (e.g., to a threatening bear), but since the accuracy of this information varies, it can also lead to poor results.

The two central ideas of “emotions as perception” and of pre-cognitive “inner concepts” are challenging to defend. Generally, emotion theorists tend to think of concepts as cognitive by definition, and it needs to be shown how the conceptualization of non-cognitive concepts can be meaningful in light of empirical evidence. MacLean (1990: 424) argues that “compulsions, emotions, and thoughts can be subjectively distinguished from sensations and perceptions by their capacity to occur and persist ‘after-the-fact,’” which brings the theory of “emotions as perception” into question. The thesis that there are emotions *in* perception, in contrast, is strongly corroborated (Droit-Volet and Meck, 2007; Panksepp, 1998; Arnold, 1960), recognizing that there are deeper brain structures from which emotions emanate (Panksepp and Biven, 2012; Panksepp, 1998).

2.5.5 Other philosophical theory of emotion

There is great diversity among philosophical theories of emotion. Several theories could be insightfully discussed, including works such as Goldie (2012), Robinson (2005), Nussbaum (2003), and de Sousa (1987). Here I highlight only two further philosophical theories

(Griffiths, 1997, and Green, 1992) since they have attracted particularly significant attention. For space considerations, I will only briefly present their main ideas.

Green (1992) views emotions as intentional states of mind formed by a combination of desires and beliefs. This perspective gives emotions a cognitive dimension as well as a motivational one (Green, 1992: 18) because desires aim at satisfaction (a motivational function), whereas beliefs aspire for truth and are directed as providing accurate information (a cognitive function). At the same time, emotions are different from desires and beliefs because emotions involve feelings, whereas desires and beliefs do not (Green, 1992: 26). He disagrees, however, with Hume's view that emotions are affective (Green, 1992: 68-69).

Griffiths (1997) argues for an *eliminativist* position regarding the prospect of making a holistic theory emotion. The reason he doubts that it is possible to present a coherent account of emotions is because emotions refer to two different things. On the one hand, there are "basic emotions," which he considers to be adequately captured by Ekman's "affect program research" and the neuroscientific "somatic marker theory" of Damasio. On the other hand, there are also "higher cortical emotions," which he holds are best explained by evolutionary psychology. The main point in his argument is that these two different types of "emotions" are lumped together by mere analogy, which means that these phenomena are only superficially similar. Hence, they should be studied separately.

2.6 Historically influential theories of emotion

This section provides a brief look at the major historical events and resulting shifts in perspectives over the past few centuries that have developed into current theories of emotion. Many of the major historical views are still alive in the research community today. These views need critical assessment because science ought to aim to establish an empirically satisfactory and theoretically plausible understanding of emotions instead of preserving historical lineage. In this regard, it needs to be emphasized that much of what is currently known about emotions by science today was in the dark only *decades* ago (Panksepp, 2012).

While paradigm shifts in understanding have taken place over the history of emotion theory, the paradigm shifts in psychology appear to have placed emerging paradigms side-by-side with the old paradigms rather than replacing them (Panksepp, 1998: Ch. 1). Hence, over this century and the past one, the Kuhnian processes of scientific revolution seem to have been largely incomplete in emotions theory. As Churchland (1987: 546) notes, this is perhaps typical of paradigms: “Paradigms rarely fall with decisive refutations; rather they become enfeebled and slowly lose adherents.”

In neuroscience—a discipline that has rather recently emerged—there is at present a somewhat more theoretical convergence compared to the psychology literature. Several previously esteemed models have been eschewed on empirical grounds. At the same time, even here, much of the intellectual baggage from behaviorism has been preserved (Panksepp, 2012); for example, behaviorism often poses needless difficulties by an entrenched disbelief in the idea that other animals can have emotions and experience feelings (Panksepp, 2012; 2011).

In his seminal work, *The Expression of Emotions in Man and Animals*, Charles Darwin (1872) was one of the first to make an explicit scientific account of emotions. The

biologically inclined philosopher, Nietzsche, guided by his philosophical intuition and rudimentary understanding of the life sciences, in his polemic but perceptive manner, highlighted the intimate connection between biological affects and morality as follows: “moralities are also merely a sign language of the affects” (Nietzsche, 1885: 100).

The psychology of “passions,” which (as a cluster concept) closely intersect with the concepts of “emotions” and “affects,” has a much longer history (Rorty, 1978: 141). It is for example present in the writings of David Hume and has precursors in Western philosophy at least as far back as Aristotle. It should be noted that insightful philosophical discussions of emotions or affect also are found in Eastern philosophy at an impressively early stage. Consider the following passage:

“When I say that all men have a mind which cannot bear to see the suffering of others, my meaning may be illustrated thus: even nowadays, if men suddenly see a child about to fall into a well, they will without exception experience a feeling of alarm and distress. They will feel so, not as a ground on which they may gain the favor of the child’s parents, nor as a ground on which they may seek the praise of their neighbors and friends, nor from the dislike to the reputation of having been unmoved by such a thing. From this case we may perceive that the feeling of commiseration is essential to man.” (Mencius (*Mèngzǐ*, 372-289 BC): 78).

This statement shows that sophisticated philosophical accounts of emotion have been around for a long time in human intellectual history. Scientifically, however, Darwin’s *affective* and *evolutionary* theory is a milestone. Darwin argued that emotions have evolved as a result of adaptation. He further contended that emotions express underlying bodily emotional states (Hess and Thibault, 2009: 122). Darwin suggested that the expression of emotions was a widespread adaptive phenomenon: “Even insects express anger, terror, jealousy and love by

their stridulation” (Darwin, 1872: 349). While Darwin may have somewhat overestimated the sophistication of the emotional lives of fruit flies, the general emotional expressivist position he outlined appears to stand up remarkably well in light of modern neuroscience (Panksepp, 1998: 11).

A number of important theories of emotion, including those of Tomkins (1963; 1962), Plutchik (1991), and Panksepp (1998), have been inspired by Darwin. Darwin’s broader view of the role of emotions is structurally captured by the Cannon-Bard account of emotions (described below). Soon after Darwin published his views, a conceptually and structurally different theory, which was to become influential, surfaced: the James-Lange theory of emotion. This view remains popular in contemporary theory of emotion, perhaps more so than the position Darwin advocated.

2.5.1 The James-Lange theory of emotion

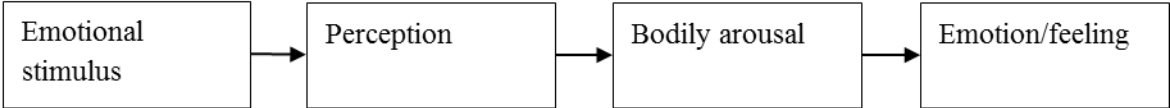
Different versions of the James-Lange theory of emotion (after William James and Carl Lange) have been and remain influential within philosophy (e.g., Russell, 1921: 146–150) as well as psychology (e.g., Schachter, 1964). James and Lange presented their *cognitive* theories of emotion independently around the same time in the early 20th century. The central claims of the James-Lange theory are as follows: (1) emotions *are* feelings, which (2) are *caused by* physiological changes in the body, and (3) these physiological changes in turn are *caused by* cognition.

According to James (1892: 449), “we feel sorry because we cry, angry because we strike, afraid because we tremble, and not that we cry, strike, or tremble, because we are sorry, angry, or fearful, as the case may be.” In the emotional process, the following sample scenario may occur: (1) stimulus (danger), (2) perception (internalizing stimulus), (3) physiological arousal (trembling, fleeing, etc.), and finally (4) the subjective experience of

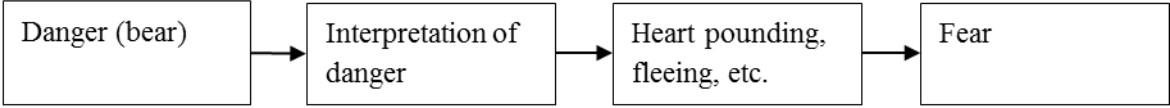
emotion (being afraid). Thus, the James-Lange theory would have us running from the bear, not because we were afraid but being afraid because we were running.

Fig. 6–The James-Lange theory of emotion:

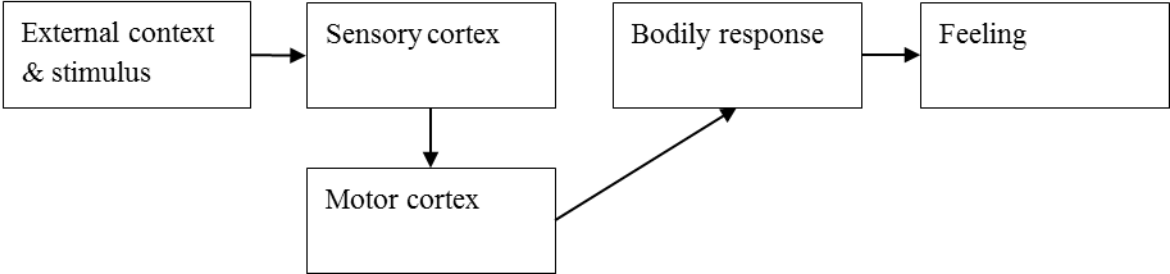
Emotional process:



Example event:



Brain and physiology:



It should be noted that the idea that our feelings of bodily changes *are* the emotion makes it questionable whether this really is a theory of emotion at all. That is, this may rather be a *theory of feeling* inappropriately categorized as a *theory of emotion*.

It has been suggested that the perspective of this theory as a theory of emotion “was quickly refuted by research showing that complete removal of the neocortex failed to disrupt the expression of emotional responses elicited by sensory stimuli, thus suggesting that the sensory and motor cortex could not be the key” (LeDoux, 2002: 201; see also MacLean, 1990; 1949; Cannon, 1929).

With the cortex surgically cut off, animals will still respond with “sham rage” to emotional stimuli. By performing various surgical cuts on animals, Cannon and his assistant

Bard managed to isolate the subcortical area and the hypothalamus as necessary components in the production of emotional responses (MacLean, 1990: 30). This particular area of the brain plays no role in the James-Lange theory. Thus, early evidence suggests that the James-Lange theory is both *incomplete* as well as *incorrect* as an approach to structuring the emotional process. It also suffers from other shortcomings, such as its inability to account for unconscious emotions (i.e., “repressed emotions”), and its inability to make sense of the distinction between persistent emotions and ones that are acute (Plutchik, 1991: 26).

It is perhaps only due to a streak of irony that the views of William James provided fertile ground for the development of a range of cognitive theories of emotion, many of which essentially retained the same overall structure. Apparently, he intended to avoid precisely this type of theory. According to James (1892: 449): “Without the bodily states following on the perception, the latter would be purely cognitive in form, pale, colourless, destitute of emotional warmth.” Inasmuch as he concedes this point, he also appears to reveal that emotion is in fact already involved, at least in some primitive form, in the process of perception (presumably a point Lange would object to). This concept would strengthen the interpretation that the way “emotions” is used by James—as a placeholder for feelings—is a bit of a misnomer. If this analysis is correct, it would suggest that what William James intended to put forth was in fact an *affective* theory of emotions (what Lange might have intended is a different matter).

2.5.2 The Cannon-Bard theory of emotion

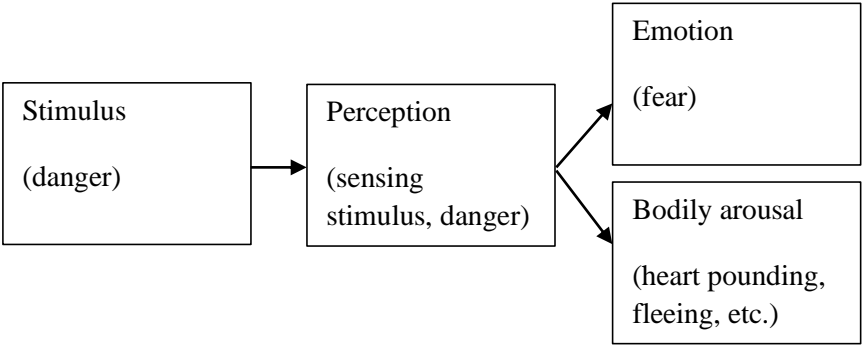
As noted, the James-Lange theory was soon criticized by competing views, most notably what became known as the Cannon-Bard theory of emotion, developed by Walter Cannon and Philip Bard. They argue that it is indeterminate which comes first, an emotional reaction or

bodily arousal (depending on different situational cues and individual personal differences), because these processes are triggered more or less simultaneously.

The Cannon-Bard theory of emotion advances a physiological view. Based upon experiments involving different surgical cuts, Cannon (1929) bases emotions in the brain, located in the emotional centers subcortically in the *thalamus* and *hypothalamus*. While research now shows that this locus is too narrow, since additional affective regions also need to be included, the thalamus and hypothalamus *do* appear to play a crucial role in emotional processes (LeDoux, 2002: 122; Panksepp, 1998: 57).

Walter Cannon’s recognition that emotional feelings reflect activity in distinct brain networks, which he correctly (Panksepp and Biven, 2012) pins down to the *subcortical regions* of the brain, is an important precursor to modern brain sciences. Early studies, which demonstrated that animals without a neocortex can still express behavior associated with distinct emotions—positive emotions such as pleasure as well as negative emotions like fear—(Bard and Rioch, 1937), represent valid scientific evidence that theories of emotion need to incorporate (Panksepp, 1998). The Cannon-Bard theory of emotion suggests the following generic scenario: (1) stimulus (danger), (2) perception (internalizing stimulus), (3a) physiological arousal (trembling, fleeing, etc.), (3b) the subjective experience of being emotional (being afraid). Now we are “allowed” to run from the bear while being afraid.

Fig. 7–The Cannon-Bard theory of emotion:



2.5.3 The Schachter-Singer theory of emotion

Another historically influential perspective is the Schachter-Singer theory of emotion, which is a product of the “cognitive revolution.” Holding cognitive factors to be major determinants of emotional states, Schachter and Singer (1962) propose a two-factor psychological process involving first a stage of physiological arousal and subsequently *cognitive labeling* of this arousal in order to produce an emotion. Their conjecture is that the *situation* a person is in will essentially determine the labeling, and hence the resulting emotion.

For example, if a person happens to be at a party, she might label a given arousal as “happy.” Schachter claims that a certain physiological response can be interpreted in many different ways, so that potentially it can produce a wide range of different emotions. We have the following generic structure: (1) stimulus (e.g., danger), (2) physiological arousal (e.g., trembling, fleeing), (3) cognitive labeling of arousal (e.g., “I am trembling, fleeing, etc.” combined with “my trembling, fleeing, etc., warrants fear”), and (4) conscious emotion (e.g., fear). The physiological arousal combined with the cognitive labeling produces the emotional reaction.

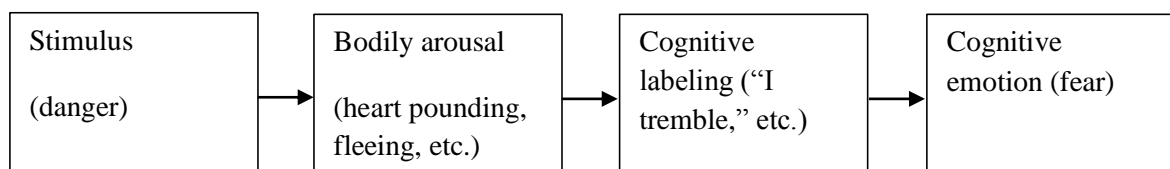
This model retains and perhaps even compounds the weakness mentioned above relating to the James-Lange model, namely insisting that emotions always come as the final response of this internal emotional chain of cause and effect. Additionally, it must be shown that particular linguistic regions in the neocortex involved in assigning labels are active in all emotions. It can be doubted whether linguistics is part of any basic emotion; for example, while it is generally understood that infants do not have any linguistic vocabulary for labeling, it is also widely agreed that infants experience and express basic emotions (Stern, 1985;

Emde, 1983). This suggests that the Schachter-Singer theory can only be true of certain types of emotions that are specific to linguistically capable human adults.

Another serious criticism for the Schachter-Singer theory is based on the well-established phenomena of *confabulation* (Solms and Turnbull, 2002; Griffith, 1997: 81–83). Confabulation refers to the fact that when subjects are unable to account for their behavior and physiological reactions, they have a propensity to invent or fabricate an explanation. Thus, what is reported by subjects in the famous Schachter and Singer (1962) study cannot be taken at face value as reporting true explanations of their emotions. Confabulation therefore represents an empirically supported and credible alternative theoretical explanation to the Schachter-Singer theory.

The model has also been criticized for proposing a sort of labeling that the persons who experience the emotional episodes generally do not themselves recognize (Arnold, 1994), which is striking insofar as the labeling process is supposed to be conscious and overtly cognitive. (in between “stimulus” and “bodily arousal” in Figure 3, there should be an additional box labeled “perception,” which has been omitted for the brevity of presentation).

Fig. 8–The Schachter-Singer theory of emotion:



2.5.4 Watson-Skinnerian behaviorism

Behaviorism, as advocated by John Watson and Burrhus Skinner, attempted to restrict the field of psychology to make it more like a “hard science.” As Frith (2007: 6) puts it, “psychologists pretended to be real scientists by studying only behavior.” Meaningful

emotion research became difficult under this paradigm since behaviorism basically outlawed theorizing about emotion. Initially, this banishment was only half-hearted; Watson himself recognized three emotions: “fear,” “love,” and “rage” (Ortony and Turner, 1990: 315). Soon, however, behaviorism came to view the mind, and hence the brain, as a “black box” (Skinner’s term) that could not be scientifically studied. Everything mental was ruled out, including emotions. Skinner viewed emotions as mere fictions of folk psychology (Izard, 2010: 363). Since this perspective was unexciting for anyone interested in emotions, it marks a time-gap in emotion research (Panksepp, 1998). Remarkably, behaviorism aimed to establish a paradigm for studying the mind without the mind; in other words, “psyche-ology” *without* the psyche.

Thus, subjective states of mind were completely ignored. As LeDoux (2002: 175) puts it, “John Watson and fellow behaviorists replaced this focus on subjective states with a mindless psychology of objectively measurable events (stimuli and responses).” Because the behaviorist paradigm claimed that “emotions” were “fictive” (Watson, 1924; Skinner, 1938)—indeed going beyond the mere ignorance about content implied by the black box metaphor—this paradigm became especially hostile to emotions research (Panksepp, 1998: 9). Hence, in the heydays of behaviorism, it was common to deny that emotions (affect) and thoughts (cognition) could have any influence on human behavior. Instead, psychologists looked to factors in the environment in order to explain behavior. It can be argued that even the use of notion of “behavior” by itself is problematic. As MacIntyre (1981: 194) notes: “There is no such thing as ‘behaviour’, to be identified prior to and independently of intentions, beliefs and settings. Hence the project of a science of behaviour takes on a mysterious and somewhat outré character.”

Starting in the 1920s, behaviorism dominated psychology for several decades, until it finally gave way to the “cognitive revolution.” According to Izard (1991: 1), emotions as a

research topic was more or less ignored in psychology until the 1980s. However, not every part of psychology was restored as behaviorism receded: “When the cognitive revolution later made the mind fair game again, it did not do so by reviving subjective psychology. The thinking process itself, rather than the conscious content that results from thinking, became and largely remains the subject matter of cognitive science” (LeDoux, 2002: 175).

Behavior is still valuable for emotion research, however. Behavioral experiments and inferences made from behavior are crucial for understanding what goes on inside the head, but while behaviorism has been, and still is, an important part of understanding the emotional process, in isolation it proves to offer only an overly limiting view (Panksepp, 1998).

2.7 Towards a theoretical platform

2.7.1 Positive and negative emotions

Valence is important for any coarse categorization of emotions, allowing their classification into negative and positive categories. This point is widely accepted by emotion researchers (Izard, 2010). Given the weight of scientific evidence from neuroscience cited in the review, the proposition that the nature of negative and positive valence is *affective* seems reasonably well substantiated by evidence (see Ch. 5 for further detail). Animal studies suggest that emotions are *dichotomous* such that they are either experienced as *positive*, in which case mammals are willing to work to have additional stimulation, or *negative*, in which case mammals strive to make the emotional stimuli disappear or flee (Panksepp, 1998). Valence is also evident in affective-cognitive conglomerate emotions. For example, many have personal experience with grief as an unmistakably negative emotional feeling, one that is typically accompanied by a desire for emotional closure (Goldie, 2012: 70–71).

Valence is an important characteristic (or dimensional aspect) of discrete emotions. However, specifying discrete emotions in more detail is both possible and informative (Izard,

2010). In relation to decision-making, explicitly specifying the emotions allows a finer resolution of the conceptualization of decision goals; value what one wants to achieve or maximize, and disvalue what one wishes to avoid or minimize. Decision-makers and stakeholders involved in a decision will often have *mixed* emotions about a given alternative. This is typically the case when one faces what one perceives to be a moral dilemma. In such cases, it will often be helpful to understand what discrete emotions this emotional mix is composed of.

By specifying these discrete emotions, more explicit communication concerning what is at stake—what one’s concerns are—is facilitated, moving those involved beyond merely expressing agreement or disagreement. For example, while fear and anger are both negative emotions, each entail markedly different implications for decision-making (Lerner and Keltner, 2001); fear tends to reduce risk-taking, whereas anger tends to promote risk-taking.

Discrete emotions appear to be fairly well mapped by emotions scientists. A broad picture emerges, which centers on a set of “basic” or primary emotion categories. Each category corresponds to a particular neurological emotional system, for which abundant neurological evidence is available (Panksepp and Biven, 2012; Panksepp, 2012; Watt and Panksepp, 2011; Solms and Turnbull, 2002; LeDoux, 2000; Panksepp, 1998; MacLean, 1990: Ch. 16-27). These emotional systems are presented in more detail later (in Ch. 4).

According to Panksepp (1998), neuroscience offers the only viable scientific way to define emotions (i.e., in terms of neural circuits), because it is the only way to pin emotional terms to concrete physical processes. A further scientific advantage of defining emotions this way is that facts about emotions can be empirically demonstrated and tested. For example, as noted, the characteristic emotional expressions corresponding to each of the primary emotion categories hypothesized by affective neuroscience can be reliably produced by means of

electrical stimulation of specific areas of the mammalian brain (Shapira et al., 2006; Panksepp, 1998). Current neuroscience moreover suggests that all emotional systems are located in the subcortical regions of the brain (Panksepp and Biven, 2012; Panksepp, 1998). This idea and related points are elaborated on in more detail further on (again, see Ch. 4).

While affective neuroscience (e.g., Panksepp, 1998) provides an important scientific account of the neural correlates of emotion, it is important not to see the contribution of this field as reductive but rather as establishing a much-needed biological foundation for emotion. It can be argued that “emotion refuses to be reduced to any simple one or two elements but remains stubbornly multifactorial, involving facial and other motor system changes and motor primings, a subjective feeling state or ‘valence,’ and physiological/anatomical changes, along with frequent (but not invariant) ‘top-down’ cognitive drives or appraisals” (Watt, 2003: 85).

In relation to normativity, it is also important to gain an understanding of higher-order emotions (secondary and tertiary), especially characteristically social emotions, such as shame, guilt, and pride (Buck, 1999). A crucial part of this understanding, however, involves illuminating how higher-order emotions build on primary emotions. As Plutchik (1991) points out, there is considerable variability in the manifestations of human emotions, although the basic affective components probably boil down to just a few distinct types. The dynamic interaction between affective and cognitive processes is crucial for understanding the variety of emotions and their manifestations.

A number of theories of emotion propose a set of basic or primary emotions. As can be seen below (*Table 1*), they form a relatively coherent picture, with only small variations in terminology and constructs. In psychology, there seems to be a historical line running from Charles Darwin through the Cannon-Bard theory of emotion via Tomkins (1962; 1963) to more recent affective theories of emotion, including Izard (2011; 2009; 1991), Panksepp

(1998; 1992), Plutchik (1991), and Zajonc (1984; 1980). Among these theories, there is approximate agreement on which emotional categories to include as well as on the causes and nature of emotions. While numerous other competing perspectives on emotion stand in contradiction with this position, recent evidence appears to corroborate the body of affective theories of emotion (see Ch. 4 and Ch. 6).

In light of considerable hard evidence, it seems that the fundamental nature of emotions is affective (Damasio, 2010; Panksepp, 1998). The basic theoretical assumption necessary for the explanatory strength of the cited evidence (detailed concerning hard evidence can be found in Ch. 4.2.1) is that a similar subjective affective experiences arise from almost identical brain structures in humans as other mammals (i.e., from the *reticular activating system* in the brain stem; Solms and Turnbull, 2002; see Ch. 4.2.3).

Still, however, cognitive theories of emotions appear to make up the center of gravity in contemporary philosophy as well as in psychology. Concerning human emotions, there seems to be a discrepancy between the availability of empirical evidence, which suggests that emotions are fundamentally affective, and the theories that are commonly adopted or assumed, which suggests that they are cognitive. Hence, there is a clarification need in relation to applied areas of research, such as practical business ethics, insofar as they have to cope with and understand the role of human emotions.

2.7.2 Primary emotion categories

An image of seven or eight *primary emotion categories* emerges that is consistent with current evidence in neuroscience (Panksepp and Biven, 2012; Davis and Panksepp, 2011; Watt and Panksepp, 2011; Vytal and Hamann, 2010; Panksepp, 1998) and supported by cross-disciplinary coherence between a number of theories in neuroscience, psychology (e.g., Izard, 2011, 1991; Levenson, 2011; Ekman and Cordaro, 2011; Ekman, 2003; 1992; Plutchik,

1991), and philosophy (e.g., Robinson, 2005) and corroborated by research on other primates (e.g., de Waal, 1996).

This image, however, is not congruent with all, or even most, of the theories of emotion represented in current psychology and philosophy. The platform should therefore be viewed as a minority position (Watt, 2005). However, it is a robust scientific position because it conforms to the current corpus of physiological and neurological evidence about the human body and brain (see Ch. 4).

The classical affect theory of Tomkins (1991; 1963; 1962) proposes 9 primary emotion or affect categories. Six of them he considered basic and evolutionarily older: (1) interest-excitement, (2) enjoyment-joy, (3) surprise-startle, (4) distress-anguish, (5) anger-rage, and (6) fear-terror. In addition, he proposed three more: (7) shame-humiliation, (8) disgust, and (9) “dissmell.” In the table below (*Table 1*), these primary affects are entered in an abbreviated form, while the ninth affect is omitted from the summary since it is not well substantiated by other emotion theorists (and in any case seems subsumed under disgust).

Plutchik (1991) suggests that there are 8 basic emotions: (1) trust (acceptance), (2) anger, (3) anticipation (interest), (4) disgust, (5) joy, (6) fear, (7) sadness, and (8) surprise. “Surprise,” which Plutchik shares with Tomkins and Ekman, seems less well established as a distinct category of affect or emotions. According to Ortony and Turner (1990: 317), “[when] a person is surprised by something, nothing is entailed about the affective state of a person,” leading them to conclude that surprise is not itself an emotion. Most emotion theorists do not include “surprise” as an emotion.

Panksepp and Biven (2012) and Panksepp (1998) present detailed evidence for seven distinct emotional brain circuits, each causally associated with a distinct set of emotional expressions and behavioral patterns. On this basis, Panksepp advances a theory of emotions

that holds that there are at least seven “primary affects” or “primary-process emotions”: (1) play, (2) panic/grief, (3) fear, (4) rage, (5) seeking, (6) lust, and (7) care. These emotions are considered *inherently* affective and subcortical. In addition to primary-process emotions, Panksepp suggests that there are several higher-order emotions that stem from cognitive-affective interactions. Primary emotions or affects are thus taken to be “building blocks” for more complex, conglomerate emotions, such as guilt, shame, loneliness, and hatred (Panksepp and Watt, 2011: 389; Levenson, 2011).

Similarly, several other emotion theorists have recently suggested that there is a limited set of primary or basic human emotions, including Ekman and Cordaro (2011), Frijda (2007), Izard (2011), and Levenson (2011). Most of these theorists share the view of primary emotions as affective, and their views, together with those that are already discussed, are presented in the *Table 1* for comparison.

Table 1—Theoretically and empirically supported basic emotions:

Table 1: Theoretically and empirically supported basic emotions

Tomkins (1962; 1963; 1991)	Plutchik (1991)	Levenson (2011)	Ekman & Cordaro (2011)	Izard (2011; 1991)	Panksepp & Biven (2012)
Joy (positive)	Joy (positive)	Enjoyment (positive)	Happiness (positive)	Happiness (positive)	Play (positive)
Distress (negative)	Sadness (negative)	Sadness (negative)	Sadness (negative)	Sadness (negative)	Panic/grief (negative)
Fear	Fear	Fear	Fear	Fear	Fear

(negative)	(negative)	(negative)	(negative)	(negative)	(negative)
Anger	Anger	Anger	Anger	Anger	Rage
(negative)	(negative)	(negative)	(negative)	(negative)	(negative)
Disgust	Disgust	Disgust	Disgust	Disgust	
(negative)	(negative)	(negative)	(negative)	(negative)	
Interest	Anticipation/interest	Interest*		Interest	Seeking
(positive)	(positive)	(positive)		(positive)	(positive)
Shame	Surprise	Love*	Surprise		Lust
(negative)	(neutral)	(positive)	(neutral)		(positive)
	Trust/acceptance	Relief*			Care
	(positive)	(positive)			(positive)
Contempt			Contempt	Contempt*	
(negative)			(negative)	(negative)	

Table 1 is based on the overviews given by Tracy and Randles (2011) and Ortony and Turner (1990). It provides an overview of some of the most central theories about basic emotions in psychology and neuroscience. In addition to proposing that there are “primary,” “fundamental,” or “basic” emotions, these theories also view emotions as affective rather than cognitive. Furthermore, each theory builds on its own detailed empirical material. In this sense, the perspectives included in the overview have essentially arrived at their conclusions

independently. The starred items indicate where the respective researchers consider the empirical evidence insufficient for drawing definite claims.

In addition to these affectively based basic emotions, most of the enlisted accounts suggest that there are further, non-basic and more complex emotional conglomerations of affect and cognition. “Jealousy,” “guilt,” and “shame,” for instance, are examples of such conglomerate emotions (Watt, 2005). The level of agreement between the affective theories of basic emotion is evident. If we compare them with recent cognitivist accounts of basic emotions, however, the level of agreement drops markedly. Consider for example the classical cognitive appraisal perspective on basic emotions (Arnold, 1960): anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, and sadness. While Arnold (1960) suggests that 11 are basic, Scherer (2005) proposes 16 basic emotions. These are just two examples, but they are illustrative.

One might expect that adopting an outlook in which cognition is foundational to the idea basic emotions would open a potentially *wider range* of basic emotions compared to an outlook where one sees the nature of basic emotions as affective. There are multiple ways in which affect and cognition can interact to form complex emotional conglomerations (Scherer, 2005: 614–615; Panksepp, 1998; Tomkins, 1962). In the cognitive accounts of emotion, it is also more difficult to pin down the exact criteria for what to count as “basic,” especially when the affective foundations of emotions are rejected. Many of the cognitive views see all the neuronal processes of the brain in terms “information processing,” but it should be kept in mind that “affective feelings are to a substantial degree, distinct neurobiological processes in terms of anatomical, neurochemical, and various functional criteria” (Panksepp, 2003: 6). A comparative advantage of the affective view of emotions and feelings, then, seems to be that it can hold core affective emotions as basic building blocks in order to understand more complex affectively infused psychological structures.

Certain cognitive accounts of emotions reject the very idea of basic emotions, citing arguments like the low level of agreement (Frijda, 1999; Ortony and Turner, 1990), often in favor of conceptually slimmer dimensional accounts of emotion that map emotions onto a two-dimensional plane of valence and arousal. Theories of basic emotions need not reject these dimensions (Panksepp, 2007) but clearly have the empirical backing to hold that there is much more to say about emotions than this (Panksepp and Biven, 2012). Looking at the evidence and the affective accounts of emotions, it appears that a broad, scientifically based agreement is in place, which includes at least the six first emotions listed in *Table 1*. Not only is there agreement that specific key emotions are basic, but there is also agreement on the valence of these emotions.

The most empirically robust basis we have for proposing basic emotions and for grounding affect arise from research in affective neuroscientific processes (Ch. 4). There is therefore scientific pressure for philosophically and psychologically based accounts of emotions and affect to conform to the best current evidence provided by neuroscience. While the intellectual exchange between fields of research travels by two-way lanes, there should be traffic rules where “softer” evidence must yield to “harder” scientific evidence.

2.7.3 Affect and cognition

Both “affective” and “cognitive” processes are likely to play an important role in moral deliberation and ethical decision-making (Gaudine and Thorne, 2001), but this importance can only be appreciated if we are clear about the meaning of these terms. Distinguishing the concepts of “cognition” and “affect” is essential for several inter-related reasons. One reason is that the ensuing discussion cuts into *physiological processes* of the brain to which these key terms correspond. A second reason is that the *experiences* that attach to the affective and cognitive brain processes are likely to be different (Panksepp, 1998). A third reason is that the

experiential difference is potentially important in relation to understanding normativity. As stated in the previous chapter, I hold as a working hypothesis that normativity, as a psychological quality, can be captured in terms of affect. Determining whether emotions are affective or cognitive is in this perspective crucial for explaining the nature of normativity (Ch. 1.3).

The central notions of “affect” and “cognition,” as we have seen, are sometimes used in markedly different ways in the emotions literature; the exchange between Lazarus (1984b; 1982; 1981) and Zajonc (1984; 1980) is illustrative. Most strikingly, specific cognitive theories of emotion appear to downplay the importance of affect by presenting a *deflated* version of the term. In the more extreme cases, “cognition” is interpreted (or re-interpreted) so that it becomes wide enough that it in effect includes what is conventionally understood as “affect” (Levenson, 2011: 384). As we saw in the review, Lazarus (1991) presents such a view. The problem with this approach, however, is that it is not helpful (Cromwell and Panksepp, 2011); it glosses over important and useful distinctions and thereby adds to the problems of understanding rather than diminishing them. I propose that what is needed is neither a deflated nor an inflated concept of “affect.”

The science of emotions makes it clear that both affective and cognitive processes can do a variety of quite different things in our minds and brains. In this sense, “affect” and “cognition” are placeholders for a host of more detailed empirical specificities, such as the variety of effects different neurotransmitters, neuronal types, and neurochemical modulators might have. Discussion at this level of detail, however, would be too technical and unwieldy. Meanwhile, the analytical separation between “affect” and “cognition,” as generally recognized, imposes practical and informative distinctions for discussing the emotions and practical deliberation.

As general types of brain processes, cognition and affect are “deeply inter-penetrating, hierarchically positioned activities in which emotion [affect] provides the motivational foundations for all directed cognitive activity, while cognition allows for modulation, blending, and especially adaptive inhibition of prototype affective states” (Watt, 2003: 94). Based on knowledge derived from affective neuroscience, we “now know that there is very early, low-level processing of incoming sensory information for its emotional relevance in brain centers such as the amygdala, that this processing is fast and automatic, that it can occur in the absence of conscious awareness, and that it has the capacity to influence subsequent behaviors profoundly” (Levenson, 2011: 384). Panksepp (1998) argues that such processes are best characterized as “affective” rather than “cognitive.”

Philosophers sometimes suppose that the internalist position of affect as motivational implies that there is no necessary connection between cognition and affect (e.g., Swan, 2004: 375), but this is ultimately an empirical question that can only be answered by investigating the actual workings of the human brain. Current neurobiological evidence suggest that there are massive synaptic interconnection between affective and cognitive processes and mutual influences, but also that these are hierarchically structured so that motivation works in a bottom-up fashion, where cognition inhibits rather than motivational tendencies (Panksepp, 1998).

For the purpose at hand, Cromwell and Panksepp (2011: 2029) present a gratifyingly clear overview of the relevant distinction between affective processes and cognitive processes (*Table 2*). It is as follows:

Table 2—Affect vs. cognition

Table 2: Affect vs. cognition

Affective processes (values)	Cognitive processes (information)
More subcortical	More neocortical
Less computational (analog)	More computational (digital)
Intentions-in-action	Intentions-to-act
Action to perception controls	Perception to action controls
Neuromodulator codes	Neurotransmitter codes

Source: Cromwell and Panksepp (2011)

This overview presents central psychological distinctions pertaining to the mind as well as neurological distinctions that are related with the physical brain. It associates affective processes with the concept of “value,” and cognitive processes with “information.” Anatomically, it indicates a rough placement of the different categories of activity in the brain. Metaphorically, it places the conceptual categories in opposite camps. Centering on the important notion of “intention,” it distinguishes between intentions at two levels: the affective and the cognitive. It highlights mutual mechanisms of regulation and control. Finally, it says something about relevant differences in neurochemistry, a topic I shall not be concerned with.

Overall, the overview offers an inter-disciplinarily informed and evidence-based distinction between affect and cognition. At the same time, it does not prematurely rule out the possibility that the separation between affect and cognition may be one of degree. A common mistake, which the overview thankfully avoids, is to view cognition as necessarily

conscious. As Berlin (2011: 5) reminds us, a “great deal of complex cognitive processing occurs at the unconscious level and affects how humans behave, think, and feel.”

The overview Cromwell and Panksepp presents can be kept in mind as a tool for conceptual navigation in the discussions throughout the thesis. It nicely sums up the essences of the differences between affect and cognition roughly that these emotion researchers in the review who are jointly informed by psychology and neurosciences appear to converge on.

2.7.4 Emotions as inherently affective

Some of the most robust pieces of evidence about the brain point toward a view of emotions as inherently *affective*. In the review, we have seen what one of the strongest alternative views, namely Roll’s (2005) informational theory of emotions, is like. This view has been shown to have certain implications that many take to be implausible (e.g., Panksepp and Biven, 2012). One of the implausible implications is that other animals (with the possible exception of the apes most closely related to humans) have to be seen roughly as robots made out of flesh, totally deprived of affective experiences. For example, it may seem implausible to hold that the inner life of a purring cat is affectively empty. While we cannot know exactly what it is like to be a cat or a bat from the outside, we cannot help ourselves from drawing the conclusion that it is not in fact like anything at all. Moreover, current scientific evidence suggests the opposite conclusion (Panksepp, 2012).

One of the scientific reasons it seems implausible to suppose that other animals do not experience affects is, as we saw in the review of Pankepp’s psychobiological view, the simple laboratory observation that animals routinely, although non-verbally, *inform* us about their positive and negative experiences by responding positively to “rewards” (e.g., they strive for more) and negatively to “punishments” (e.g., they work towards the extinction of the stimuli)

on their own initiative (Panksepp and Biven, 2012). More detailed neurochemical and neuroanatomical evidence to corroborate this position will be presented later (see Ch. 4).

In contrast with Roll's position, there are also psychological observations that suggest that emotional experiences in humans do not require cognition and linguistic processing; "Discrete emotion experiences emerge in ontogeny well before children acquire language or the conceptual structures that adequately frame the qualia we know as discrete emotion feeling" (Izard, 2009: 5).

2.7.5 Emotional feelings

The distinction between conscious emotional feelings and unconscious emotions is important for understanding decision-making. There are different ways of conceptualizing emotional feelings in the emotions literature. From a psychological angle, Izard (2009: 5) argues that "emotion feelings can be activated and influenced by perceptual, appraisal, conceptual, and noncognitive processes [...], but cannot be created by them." This suggests that emotional feelings stem from some more basic biological structures to which it is natural to turn to the brain sciences in search of answers.

In contemporary neuroscience, feelings are conventionally taken to appear when an emotion enters into conscious experience (Damasio, 1999). An important question here is what should count as conscious experiences. Traditionally, consciousness has been considered to arise from and essentially involve *cognitive* processes. However, this view has been challenged by scientists who suggest that conscious experience has a wider scope. The revolutionary idea is that there is a more fundamental layer of human consciousness that is *affective* (Panksepp, 2005; 2004; Solms and Turnbull, 2002).

In the picture that emerges, feelings can be experienced affectively without also requiring that the experiencing subject is aware of them; "Our emotional feelings reflect our

ability to subjectively experience certain states of the nervous system” (Panksepp, 1998: 9). This ability is centered on more fundamental *affective experiences* rather than on our conscious awareness of them. According to Solms and Nersessian (1999), current evidence suggests that consciousness should be seen as quintessentially affective. In this view, feelings are affectively conscious at a pre-cognitive level and layered so that we can be more or less aware of them. I return to the discussion of affective consciousness later (Ch. 5.2.3). This idea does not change the basic distinction between feelings and emotional expressions, namely that “emotions are what an outside observer can see; feelings are what the individual subjectively experiences” (Bechara and Naqvi, 2004: 102). What it means is that affective feelings are conscious in a deeper and more stratified way than previously thought.

2.7.6 Do emotions meet the criteria?

In the previous chapter (Ch. 1.3.3) I presented seven criteria by which to determine whether emotions could reasonably serve as explanatory of normativity. In light of the extensive review of emotions that is made in this chapter, it should be possible to determine whether the concept of emotions can serve this explanatory role. It should be noted that the list of criteria that were set is not necessarily definite and exhaustive. However, this “test”, if successful, can be taken as rendering the normativity-as-affect hypothesis more plausible insofar as several attempts to falsify it have been unsuccessful.

Here are the results as determined by what we have arrived at through the review of the emotion literature across philosophy, psychology, and neuroscience (the criteria are stated first, then the assessment): (1) Emotions must, at least in principle, be describable at the level of *neurobiology*. Emotional expressions must correlate with connective activity between nerve cells. This point seems beyond doubt. Such a description is already made available by

neuroscientists and psychologists, most notably by the extensive research of Panksepp and associates.

(2) In the face of moral issues, subjective experiences and perspectives of “ought”, “good”, “right” (etc.) must be accompanied emotional activity. *Subjective normative experience must correlate with neurobiological activity* associated with emotions. A wide range of imaging studies provided by neuroscientists and increasingly experimental psychologists confirm that morality correlate with specific patterns of neural activity. However, since many processes in the brain are also inter-correlated, such imaging studies must be interpreted with some caution.

(3) Emotions must be *evaluative*, meaning that they are subjectively non-neutral. They must render objects to which they attach attractive or unattractive in some way. This point is almost unanimously accepted by emotions researches and philosophers alike.

(4) Emotions must cover positive as well as negative *valence*, such that they are capable of pointing either in the direction of “good”, “right”, and “ought” or “bad”, “wrong”, and “ought not”. However it should be allowed for ambivalence, such as in the case of mild fear combined with satisfaction derived from controlling this fear. The emotional systems that are currently mapped cover negatively valenced (“fear”, “anger”, “grief”) emotions as well as positively valenced emotions (“playfulness”, “lust”, “care”, “enthusiasm”). There are further candidates for primary emotions that may come to be added to this list, and some of these are less easy to classify in terms of valence. It is for example less clear whether “social dominance” should be classified as positive or negative, although I am inclined to assess this as something that would correspond to a positive experience for a given person.

(5) There must be emotions that are about the various features of moral issues as they are conceptualized. There must be at least some emotions that are capable of being directed at

objects. Conversely, it should be possible that emotions are more fleeting with unspecified targets. It seems clear that higher level emotions – what Panksepp refers to as “tertiary emotions” – are capable of taking objects. “Shame”, for example, represents oneself as an object, and this seems a necessary aspect of this distinctive emotion. All of the primary emotions (Panksepp, 1998), with the exception of “seeking” or general “enthusiasm”, appear to be able to take explicit objects although doing so is not a necessity. For example, you can plainly be angry *at* someone; then this someone is your object. Conversely, you may simply find yourself in an angry mood one day, but without having any clue as to why.

(6) Emotions must be *motivational*. Emotions must be capable of being experienced as motivating action (whether or not behaviorally efficacious). For every practical moral issue there must be emotions that exert psychological pressure for or against action. Emotion researchers almost unanimously agree that one of the central functions of emotions is to motivate behavior.

(7) No cases of normative experience or normative perspective holding can happen without at least a minimal involvement of emotions. It seems to be case that subjective experiences in general involve some activity in the seeking system, and moreover necessarily build on activity in affect-inducing centers in the brain stem. Thus, it seems that any normative experience should involve at least *some* emotional activity. Beyond this, it is clear that many normative experience and ruminations are highly emotional.

In sum, emotions meet each of the criteria that were set in order to determine whether they could serve as explanatory of normativity. This lends some tentative plausibility to the normativity-as-affect hypothesis. As a result of not having being able to reject this hypothesis at this stage, I will proceed to explore and develop this theoretical idea further.

2.8 Conclusion

Contemporary neuroscience corroborates theories that propose that there are discrete primary emotions, and that the nature of these is affective rather than cognitive. Affect-based theories of emotion are therefore empirically adequate (in the constructive empiricist sense; see van Fraassen, 1980), as far as their idea of what ground emotions are concerned. There seems to be broad agreement among affective emotion theorists not only regarding the affective nature of emotions but also in regards to which specific primary human emotions there are. If we place a primacy on scientific robustness and empirical data, the most plausible way to construe affective emotions seems to be in terms of neural circuits. Causality in the emotional process can be tested, for example, by associating brain stimulation input with emotional expression and behavioral output, often with high reliability. Moreover, the science of emotions appears to have become increasingly more informative as more evidence becomes available. It may not be able to provide all that practical decision-makers need to know concerning how to handle emotions for decision-making, but at the same time it can certainly offer much more than they need to know. Undoubtedly, affective neuroscience and psychological affect theory provides an important empirical footing for understanding the role of emotions in the decision-making process. Emotions appear to meet the most rudimentary requirements necessary to serve as explanatory of normativity.

Chapter 2 References

- Arnold, M. B. (1994) Cognitive theories of emotion. In R. J. Corsini, *Encyclopedia of psychology*. New York: Wiley.
- Arnold, M. B. (1960) *Emotion and personality*. New York: Columbia University Press.
- Averill, J. R. (1980) A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 305–339). New York: Academic Press.
- Ayer, A. J. (1984) *Freedom and morality and other essays*. Oxford: Clarendon Press.
- Bagnoli, C. (2011) Emotions and the categorical authority of moral reason. In C. Bagnoli (Ed.) *Morality and the emotions*. Oxford: Oxford University Press.
- Bandyopadhyay, D., Pammi, V. S., and Srinivasan, N. (2013) The role of affect in decision making. *Progress in Brain Research* **202**: 37–53.
- Bard, P. (1928) A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. *American Journal of Physiology* **84**: 490–515.
- Bard P. A. and Rioch, D. M. (1937) A study of 4 cats deprived of neocortex and additional portions of forebrain. *Bull Johns Hopkins Hospital* **60**: 201–250.
- Bechara, A. and Naqvi, N. (2004) Listening to your heart: Interoceptive awareness as a gateway to feeling. *Nature Neuroscience* **7**: 102–103.
- Berlin, H. (2011) The neural basis of the dynamic unconscious. *Neuropsychanalysis* **13** (1): 5–31.

- Berridge, K. C. and Kringelbach, M. L. (2008) Affective neuroscience of pleasure: Reward in humans and animals. *Psychopharmacology*.
- Bingman, V. P. (2011) The maturation of affective neuroscience: Celebrating the research career of Jaak Panksepp. *Neuroscience & Biobehavioral Reviews* **35** (9): 1789–1790.
- Blackburn, S. (2010) *Practical tortoise raising and other philosophical essays*. New York: Oxford University Press.
- Bodor, P. (2004) *On emotion: A developmental social constructionist account*. Budapest: L'Harmattan.
- Brewin, C. R. (1989), Cognitive change processes in psychotherapy. *Psychological Review* **96**: 379–394.
- Broad, C. D. (1971) Egoism as a theory of human motives, in *Broad's critical essays in moral philosophy*, D. R. Cheney (Ed.). London: Allen & Unwin.
- Broad, C. D. (1954) Emotion and sentiment. *The Journal of Aesthetics and Art Criticism* **13** (2): 203–214.
- Broome, J. (2013) *Rationality through reasoning*. Oxford: Wiley-Blackwell.
- Buck, R. (1999) The biological affects: A typology. *Psychological Review* **106**: 301–336.
- Cannon, W. B. (1929) *Bodily changes in pain, hunger, fear, and rage*. New York: Appleton.
- Changeux J. and Ricoeur, P. (2002) *What makes us think? A neuroscientist and a philosopher argue about ethics, human nature, and the brain*. Princeton: Princeton University Press.

- Charland, L. C. (2005) The heat of emotion: Valence and the demarcation problem. *Journal of Consciousness Studies* **12** (8–10): 82–102.
- Churchland, P. S. (2011) *Brain trust: What neuroscience tells us about morality*. Princeton: Princeton University Press.
- Churchland, P. S. (1987) Epistemology in the age of neuroscience. *The Journal of Philosophy* **84**: 544–553.
- Cromwell, H. C. and Panksepp, J. (2011) Rethinking the cognitive revolution from a neural perspective: How overuse/misuse of the term “cognition” and the neglect of affective controls in behavioral neuroscience could be delaying progress in understanding the BrainMind. *Neuroscience and Biobehavioral Reviews* **35**: 2026–2035.
- Damasio, A. R. (2010) *Self comes to mind: Constructing the conscious brain*. New York: Random House.
- Damasio, A. R. (1999) *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.
- Damasio, A. R. (1994) *Descartes’s error: Emotion, reason, and the human brain*. New York: Putnam.
- Darwin, C. R. (1872/1963) *The expression of the emotions in man and animals*. London: John Murray.
- Davidson, R. J. (2003) Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology* **40** (5): 655–665.

- Davis, K. L. and Panksepp, J. (2011) The brain's emotional foundations of human personality and the affective neuroscience personality scales. *Neuroscience and Biobehavioral Reviews* **35**: 1946–1958.
- de Sousa, R. (1987) *The rationality of emotion*. Cambridge: MIT Press.
- de Waal, F. B. M. (1996) *Good natured: The origins of right and wrong in humans and other animals*. Cambridge: Harvard University Press.
- Deak, A. (2011) Brain and emotion: Cognitive neuroscience of emotions. *Review of Psychology* **18**: 71–80.
- Decety, J. and Cacioppo, S. (2012) The speed of morality: A high-density electrical neuroimaging study. *Journal of Neurophysiology* **108** (11): 3068–3072.
- Droit-Volet, S. and Meck, W. H. (2007) How emotions colour our perception of time. *Trends in Cognitive Science* **11** (12): 504–513.
- Dunn, J. R. and Schweitzer, M. E. (2005) Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology* **88** (5): 736–748.
- Edelman, G. M., & Tononi, G. (2000) *A universe of consciousness: How matter becomes imagination*. New York: Basic Books.
- Ekman, P. and Cordaro, D. (2011) What is meant by calling emotions basic. *Emotion Review* **3** (4): 364–370.
- Ekman, P. (2003) *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York: Owl Books.
- Ekman, P. and Davidson, R. J. (1994) *The nature of emotion*. New York: Oxford University Press.

- Ekman, P. (1992) An argument for basic emotions. *Cognition and Emotion* **6**: 169–200.
- Emde, R. N. (1983) The prerepresentational self and its affective core. *Psychoanalytic Study of the Child* **38**: 165–192.
- Etzioni, A. (1988) *The moral dimension: Toward a new economics*. New York: The Free Press.
- Evans-Roberts, C. E. Y. and Turnbull, O. H. (2011) Remembering relationships: Preserved emotion-based learning in Alzheimer’s disease. *Experimental Aging Research* **37** (1): 1–16.
- Fineman, S. (2000) *Emotions in organizations*. Thousand Oaks: Sage Publications.
- Frijda, N. H. (2007) *The law of emotions*. Mahwah: Lawrence Erlbaum Associates.
- Frijda, N. H. (1999) Emotions and hedonic experience. In D. Kahneman and E. Deiner (Eds.) *Well being: The foundations of hedonic psychology* (pp. 190–210). New York: Russell Sage.
- Frijda, N. H. (1994) Emotions require cognitions, even if simple ones. In P. Ekman and R. J. Davidson (Eds.). *The nature of emotions: Fundamental questions* (pp. 197–202). New York: Oxford University Press.
- Frijda, N. H. (1986) *The emotions*. Cambridge: Cambridge University Press
- Frijda, N. H. (1987) Comments on Oatley and Johnson-Laird’s “Towards a cognitive theory of emotions.” *Cognition & Emotion* **1**: 51–58.
- Frith, C. (2007) *Making up the mind: How the brain creates our mental world*. John Wiley & Sons.

- Furtak, R. A. (2010) Emotion, the bodily, and the cognitive. *Philosophical Explorations* **13** (1): 51–64.
- Gaudine, A. and Thorne, L. (2001) Emotion and ethical decision-making in organizations. *Journal of Business Ethics* **31** (2): 175–187.
- Gibbard, A. (1992) *Wise choices, apt feelings: A theory of normative judgment*. Oxford: Oxford University Press.
- Goldie, P. (2012) *The mess inside: Narrative, emotion, and the mind*. Oxford: Oxford University Press.
- Goldman, A. H. (2009) *Reasons from within: Desires and values*. Oxford: Oxford University Press.
- Gordon, R. M. (1987) *The structure of emotions: Investigations in cognitive philosophy*. Cambridge: Cambridge University Press.
- Greco, M. and Stenner, P. (2008) *Emotions: A social science reader*. Routledge.
- Green, O. H. (1992) *The emotions: A philosophical theory*. London: Kluwer Academic Publishers.
- Greenberg, L. S. (1990) *Emotions in psychotherapy*. New York: The Guilford Press.
- Griffiths, P. E. (1997) *What emotions really are: The problem of psychological categories*. Chicago: University of Chicago Press.
- Harré, R. and Parrott, W. G. (Eds.) (1996) *The emotions: Social, cultural and biological dimensions*. Thousand Oaks: Sage Publications.
- Harré, R. (1986) *The social construction of emotions*. Oxford: Blackwell Publishers.

- Hess, U. and Thibault, P. (2009) Darwin and emotion expression. *American Psychologist* **64** (2): 120–128.
- Henik, E. (2008) Mad as hell or scared stiff? The effects of value conflict and emotions on potential whistle-blowers. *Journal of Business Ethics* **80** (1): 111–119.
- Hilgard, E. R. (1980) The trilogy of the mind: Cognition, affection, and conation. *Journal of the History of Behavioral Sciences* **16**: 107–117.
- Isen, A. M. (1993) Positive affect and decision making. In M. Lewis and J. M. Haviland (Eds.). *Handbook of emotions* (pp. 261–277). New York: Guilford Press.
- Izard, C. E. (2011) Forms and functions of emotions: Matters of emotion-cognition interactions. *Emotion Review* **3**: 371–378.
- Izard, C. E. (2010) The many meaning/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review* **2** (4): 363–370.
- Izard, C. E. (2009) Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology* **60**: 1–25.
- Izard, C. E. (1991) *The psychology of emotions*. New York: Plenum Press.
- James, W. (1884) What is an emotion? *Mind* **9**: 188–205.
- James, W. (1892) *Psychology: Briefer course*. Paris: University of Notre Dame Press.
- Jiang, Y. and He, S. (2006) Cortical responses to invisible faces: Dissociating subsystems for facial-information processing. *Current Biology* **16**: 2023–2029.

- Keltner, D., Haidt, J., and Shiota, M. N. (2006) Social functionalism and the evolution of emotions. In M. Schaller, J. A. Simpson, and D. T. Kenrick (Eds.). *Evolution and social psychology* (pp. 115–142). New York: Psychology Press.
- Kenny, A. (1963) *Action, emotion, and will*. London: Routledge and Kegan Paul.
- Koestler A. (1978) *Janus: A summing up*. New York: Random House.
- Koestler A. (1967) *The ghost in the machine*. London: Penguin.
- Kolb, B. and Tees, R. C. (Eds.) (1990) *The cerebral cortex of a rat*. Cambridge: MIT Press.
- Lazarus, R. S. (1993) Coping theory and research: Past, present, and future. *Psychosomatic Medicine* **55**: 234–247.
- Lazarus, R. S. (1991) *Emotion and adaptation*. New York: Oxford University Press.
- Lazarus, R. S. (1984a) *Stress, appraisal, and coping*. New York: Springer.
- Lazarus, R. S. (1984b) On the primacy of cognition. *American Psychologist* **9**: 124–129.
- Lazarus, R. S. (1982) Thoughts on the relations between emotion and cognition. *American Psychologist* **37**: 1019–1024.
- Lazarus, R. S., Coyne, J. C., and Folkman, S. (1982) Cognition, emotion, and motivation: The doctoring of Humpety-Dumpety. In R. W. J. Neufeld (Ed.), *Psychological stress and psychopathology*. New York: McGraw-Hill.
- Lazarus, R. S. (1981) A cognitivist's reply to Zajonc on emotion and cognition. *American Psychologist* **36**: 222–223.
- LeDoux, J. (2002) *Synaptic self: How our brains become who we are*. Middlesex: Penguin Books.

- LeDoux, J. (2000) Emotion circuits in the brain. *Annual Review of Neuroscience* **23**:155–184.
- LeDoux, J. (1996) *The emotional brain: The mysterious underpinnings of emotional life*.
Simon and Schuster.
- Lerner, J. S. and Keltner, D. (2001) Fear, anger, and risk. *Journal of Personality and Social Psychology* **81**: 146–159.
- Levenson, R. W. (2011) Basic emotion questions. *Emotion Review* **3**: 379–386.
- Lines, R. (2005) The structure and function of attitudes toward organizational change. *Human Resource Development Review* **4**: 8–31.
- Loewenstein, G. and Lerner, J. S. (2003) The role of affect in decision making. In R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (Eds.), *Handbook of emotions*. Oxford: Oxford University Press.
- MacIntyre, A. (1981) *After virtue: A study in moral theory*. Paris: University of Notre Dame Press.
- MacLean, P. D. (1990) *The triune brain in evolution: Role in paleocerebral functions*. New York: Plenum Press.
- MacLean, P. D. (1949) Psychosomatic disease and the “visceral brain”: Recent developments bearing on the Papez theory of emotion. *Psychosomatic Medicine* **11**: 338–353.
- Matsumoto, D. and Hwang, H. S. (2012) Culture and emotion: The integration of biological and cultural contributions. *Journal of Cross-Cultural Psychology* **43** (1): 91–118.
- Mauss, I. B. and Robinson, M. D. (2009) Measures of emotion: A review. *Cognition and Emotion* **23** (2): 209–237.

- McDougall, W. (1908) *An introduction to social psychology*. London: Methuen & Co.
- McNaughton, N. (1989) *Biology and emotion*. Cambridge: Cambridge University Press.
- Mencius. (n.d.) *The works of Mencius*. (G. Lu, Trans.). Shanghai: Shangwu.
- Merker, B. (2007) Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral Brain Science* **30**: 63–134.
- Murphy, S. T. and Zajonc, R. B. (1993) Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology* **64** (5): 723–739.
- Nesse, R. (1990) Evolutionary explanations of emotions. *Human Nature* **1**: 261–289.
- Nietzsche, F. (1885) *Beyond good and evil*. New York: Vintage.
- Northoff, G. (2011) *Neuropsychanalysis in practice: Brain, self and objects*. Oxford: Oxford University Press.
- Nussbaum, M. (2003) *Upheavals of thought: The intelligence of emotions*. Cambridge: Cambridge University Press.
- Nussbaum, M. (1990) *Love's knowledge*, Oxford: Oxford University Press.
- Ortony, A. and Turner, T. J. (1990) What's basic about basic emotions? *Philosophical Review* **97**: 315–331.
- Panksepp, J. and Biven, L. (2012) *The archeology of mind: Neuroevolutionary origins of human emotions*. New York: Norton.
- Panksepp, J. (2012) What is an emotional feeling? Lessons about affective origins from cross-species neuroscience. *Motivation and Emotion* **36**: 4–15.

- Panksepp, J. (2011) Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals. *PloSOne* **6** (9): e21236.
- Panksepp, J. (2008) Cognitive conceptualism—Where have all the affects gone? *Perspectives on Psychological Science* **3**: 305–308.
- Panksepp, J. (2007) Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives on Psychological Science* **2**: 281–296.
- Panksepp, J. (2005) On the embodied neural nature of core emotional affects. *Journal of Consciousness Studies* **12**: 158–184.
- Panksepp, J. (2004) Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition* **14**: 30–80.
- Panksepp, J. (2003) At the interface of the affective, behavioral, and cognitive neurosciences: Decoding the emotional feelings of the brain. *Brain and Cognition* **52**: 4–14.
- Panksepp, J. and Panksepp, J. B. (2000) The seven sins of evolutionary psychology. *Evolution & Cognition* **6**: 108–131.
- Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.
- Panksepp, J. (1992) A critical role of “affective neuroscience” in resolving what is basic about basic emotions. *Psychological Review* **99**: 554–560.
- Panksepp, J. (1982) Toward a general psychobiological theory of emotions. *Behavioral and Brain Sciences* **5**: 407–422.

- Pasley, B. N., Mayes, L. C., and Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron* **42**: 163–172.
- Plutchik, R. (2003) *Emotions and life: Perspectives from psychology, biology, and evolution*. Washington: American Psychological Association.
- Plutchik, R. (1991) *The emotions*. Lanham: University Press of America.
- Plutchik, R. (1986) *Biological foundations of emotions*. University Press of America.
- Porges, S. W. (2009) Reciprocal influences between body and brain in the perception and expression of affect: A polyvagal perspective. In D. Fosha, D. Siegel, & M. Solomon (Eds.). *The healing power of emotion: Affective neuroscience, development, and clinical practice* (pp. 27-54). New York: W. W. Norton & Co.
- Posner, J., Russell, A., and Peterson, B. S. (2005) The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology* **17**: 715–734.
- Prinz, J. J. (2007) *The emotional construction of morals*. Oxford: Oxford University Press.
- Prinz, J. J. (2004) *Gut reactions: A perceptual theory of emotions*. Oxford: Oxford University Press.
- Prinz, J. J. (2003) Emotion, psychosemantic, and embodied appraisal in A. Hatzimoysis (Ed.). *Philosophy and the Emotions*. Cambridge: Cambridge University Press.
- Reisenzein, R. (2009) Emotions as metarepresentational states of mind: Naturalizing the belief-desire theory of emotion. *Cognitive Systems Research* **10** (1): 6–20.
- Ricoeur, P. (1986) *Fallible man*. (C. A. Kelbley, Trans.). New York: Fordham University Press. (Original work published 1960)

- Ricoeur, P. (1966) *Freedom and nature: The voluntary and the involuntary*. (E. Kohak, Trans.). Evanston: Northwestern University Press. (Original work published 1950)
- Robinson, J. (2005) *Deeper than reason: Emotion and its role in literature, music, and art*. Oxford: Oxford University Press.
- Robinson, J. (1995) Startle. *The Journal of Philosophy* **92**: 53–74.
- Rolls, E. T. (2005) *Emotion explained*. Oxford: Oxford University Press.
- Rorty, A. (1978) Explaining emotions. *Journal of Philosophy* **75**: 139–161.
- Roseman, I. J. (1984) Cognitive determinants of emotions: A structural theory. In P. Shaver (Ed.), *Review of personality and social psychology: Vol. 5. Emotions, relationships, and health* (pp. 11–36), Beverly Hills: Sage.
- Russell, B. (1921). *The analysis of mind*. Peking: A public domain book.
- Russell, J. A. (2003) Core affect and the psychological construct of emotion. *Psychological Review* **110**: 145–172.
- Russell, J. A. (1980) A circumplex model of affect. *Journal of Personality and Social Psychology* **39**: 1161–1178.
- Scherer, K. R. (2005) What are emotions? And how can they be measured? *Social Science Information* **44**: 695–729.
- Schachter, S. (1964) The interaction of cognitive and physiological determinants of emotional state. In Berkowitz, L. (Ed.). *Advances in experimental social philosophy* (pp. 49–79). New York: Academic Press.

- Schachter, S. and Singer, J. (1962) Cognitive, social, and physiological determinants of emotional state. *Psychological Review* **69**: 379–399.
- Shapira, N. A., Okun, M. S., Wint, D., Foote, K. D., Byars, J. A., and Bowers, D. (2006) Panic and fear induced by deep brain stimulation. *Journal of Neurology, Neurosurgery, and Psychiatry* **77**: 410–412.
- Sherman, D. A., Kim, H., and Zajonc, R. B. (1998) Affective perseverance: Cognitions change but preferences stay the same. Annual meeting of the *American Psychological Society*.
- Shewmon, D. A., Holmes, G. L., and Byrne, P. A. (1999) Consciousness in congenitally decorticate children: Developmental vegetative state as self-fulfilling prophecy. *Developmental Medicine and Child Neurology* **41**: 364–374.
- Siegel, D. J. (2009) Emotion as integration: A possible answer to the question, what is emotion?. In D. Fosha, D. Siegel, & M. Solomon (Eds.). *The healing power of emotion: Affective neuroscience, development, and clinical practice* (pp. 145-171). New York: W. W. Norton & Co.
- Skinner, B. F. (1938) *The behavior of organisms: An experimental analysis*. Oxford: Appleton-Century.
- Solomon, R. C. (2007) *True to our feelings*. Oxford: Oxford University Press.
- Solomon, R. C. (1993) *The passions* (2nd ed.). Indianapolis: Hackett.
- Solomon, R. C. (1988) On emotions as judgments. *American Philosophical Quarterly* **25**: 183–191.

- Solms, M. (2013) "The unconscious" in psychoanalysis and neuropsychology. In S. Akhtar and M. K. O'Neil (Eds.), *On Freud's "The Unconscious."* London: Karnac Books.
- Solms, M. and Panksepp, J. (2012) The "id" knows more than the "ego" admits: Neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Sciences* **2**: 147–175.
- Solms, M. and Turnbull, O. (2002) *The brain and the inner world*. New York: Other Press.
- Solms, M. and Nersessian, E. (1999) Freud's theory of affect: Questions for neuroscience. *Neuro-psychoanalysis* **1** (1): 5–14.
- Stern, D. (1985) *The interpersonal world of the infant*. New York: Basic Books.
- Suchy, Y. (2011) *Clinical neuropsychology of emotion*. New York: Guilford Press.
- Swan, K. (2004) Moral judgment and emotions. *The Journal of Value Inquiry* **38**: 375-381.
- Tomkins, S. S. (1962) *Affect imagery consciousness: Volume I: The positive affects*. London: Tavistock.
- Tomkins, S. S. (1963) *Affect imagery consciousness: Volume II: The negative affects*. London: Tavistock.
- Tomkins, S. S. (1991) *Affect imagery consciousness: Volume III: The negative affects*. New York: Springer.
- Toronchuk, J. A. and Ellis, G. F. R. (2012) Affective neuronal selection: The nature of the primordial emotion systems. *Frontiers in Psychology* **3**: 1–16.
- Turnbull, O. H. and Evans, C. E. Y. (2006) Preserved complex emotion-based learning in amnesia. *Neuropsychologia* **44** (2): 300–306.

- van Fraassen, B. C. (1980) *The scientific image*. Oxford: Oxford University Press.
- von Scheve, C., and von Luede, R. (2005) Emotion and social structures: Towards an interdisciplinary approach. *Journal for the Theory of Social Behaviour* **35** (3): 303–328.
- Vytal, K. and Hamann, S. (2010) Neuroimaging support for discrete neural correlates of basic emotions: A voxel-based meta-analysis. *Journal of Cognitive Neuroscience* **22** (12): 2864–2885.
- Watson, J. B. (1924) *Psychology from the standpoint of a behaviorist*. Pennsylvania : J. B. Lippincott Company.
- Watson, L. and Spence, M. T. (2007) Causes and consequences of emotions on consumer behaviour: A review and integrative cognitive appraisal theory. *European Journal of Marketing* **41**: 487–511.
- Watt, D. F. and Panksepp, J. (2011) What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review* **3** (4): 387–396.
- Watt, D. F. (2005) Panksepp's common sense view of affective neuroscience is not the commonsense view in large areas of neuroscience. *Consciousness and Cognition* **14** (1): 81–89.
- Watt, D. F. (2003) Psychotherapy in an age of neuroscience. In *Revolutionary connections: Psychotherapy and neuroscience*, J. Corrigan and H. Wilkinson (Eds.). London: Karnac Books: 79–116.

- Weiss, H. M. and Cropanzano, R. (1996) Affective events theory: A theoretical discussion of the structure, causes and consequences of affective experiences at work. *Research in Organizational Behavior* **18**: 1–74.
- Whalen, P. J., Rauch, S. L., Ectoff, N. L., McInerney, S. C., Lee, M. B., and Jenike M. A. (1998) Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience* **18** (1): 411–418.
- Williams, B. (1989) *Ethics and the limits of philosophy* (2nd. ed.). Harvard: Harvard University Press.
- Winkielman, P. and Berridge, C. K. (2004) Unconscious emotion. *Current Directions in Psychological Science* **13** (3): 120–123.
- Zachar, P. (2010) Has there been conceptual progress in the science of emotion? *Emotion Review* **2** (4): 381–382.
- Zajonc, R. B. (1984) On the primacy of affect. *American Psychologist* **39**: 117–123.
- Zajonc, R. B. (1980) Feeling and thinking: Preferences need no inferences. *American Psychologist* **35**: 151–175.
- Zellner, M. (2011) The cognitive unconscious seems related to the dynamic unconscious: But is not the whole story. *Neuropsychoanalysis* **13** (1): 59–63.

Chapter 3: The mind-brain argument: The mind is created by and depends on a socially responsive brain³

3.1 Introduction to the mind-brain argument

Two tentative propositions that emerge from the review of the emotions literature (Ch. 2) are that (1) the nature of human emotions is affective, and (2) emotions are biologically based. While this information provides us with a coarse answer to the two research questions, it also raises new philosophical questions. Are these two conclusions mutually consistent? Can they form a coherent picture? To appreciate the difficulties involved, I wish to draw attention to the *duality* these two positions represent. On the one hand, there are *subjective* affective experiences; on the other hand, there are the *objective* biological facts of human physiology. Many scientists and philosophers have argued that these two dimensions are difficult to conceptualize in a unified manner. The difficulties become especially apparent when we contemplate consciousness.

Several interconnected philosophical-scientific problems are entailed by the subjective-objective duality (as presented below). In this chapter I argue that these problems are not satisfactorily resolved scientifically or philosophically. Therefore, I propose that the best that can be achieved based on current knowledge is to identify a plausible theoretical position that fits both the scientifically derived empirical data (i.e., van Fraassen's requirement of "empirical adequacy") and the qualitative "data" provided by subjective experience. A plausible theoretical position acknowledges the problems and specifies an empirically informed manner of treating them holistically. Although the precise solution to the problems is not known, something is known about what a solution must achieve. In

³ I wish to give special thanks to Kai Leitemo for reviewing and commenting on this chapter.

essence, it must provide a way of bridging the gap between the *subjective mind* and the *objective brain*.

Thus, the focus of this chapter is on bridging the concept of the psychological *mind* with the biological *brain*, with the specific aim of elucidating normativity. At the outset, these two perspectives appear to designate fundamentally different approaches to understanding what goes on inside the head. If we explore the possibility of construing normativity in *psychological* terms and at the same time base normativity on *biological* emotions, we need an empirically adequate and plausible theory that allows unification of the mind and brain that at the same time does not ignore or prematurely dismiss the problems involved.

The mind-brain argument proposes that the *mind* is *brain-dependent*, in the sense that the mind originates in the brain (the dependence claim) and is created by it (the causal claim). However, I refrain from (but do not conclusively reject) identifying the mind with the brain (the identity claim). Such conflation would involve of conflating experience with the physical reality that underpins this experience, which some argue is a conceptual error (e.g., Hacker, 2011). The conflation of consciousness with brain activity implied by identity theory has also been argued against by Kripke (1980), who holds that brain activity fails to “rigidly designate” consciousness, which would be required for establishing identity. Equating the mind and brain is furthermore called into question by the inability of science to directly access the domain of subjective experience, as demonstrated by the so-called “hard problem” (Chalmers, 1995).

In this chapter, I conclude by endorsing *dual-aspect monism*, as advocated by Solms and Turnbull (2002) as a means of explicating the relationship between the mind and the brain. This position seems reasonable given available scientific evidence as well as the conjecture that every emotionally feeling human being has internal subjective experience.

This is a conception of a mind that is *created by* the brain and in a fundamental sense *dependent on* the proper functioning of its basic emotional infrastructure. Moreover, since a healthy brain interacts extensively with its social and physical surroundings, sociocultural factors also shape the subjective mind.

3.2 Problems of the mind and brain

Chalmers (1995) coined the term the “hard problem,” which is now commonly used to describe the difficulty in making a unified theoretical account of the subjective mind and the objective brain. It is called the “hard problem” because there are many easier problems for the philosophy of the mind and neuroscience to deal with. Philosophers have proposed a number of different formulations of the hard problem. Some of these invoke other notions than the “mind” that they consider equivalent in a roundabout way, such as the “soul” or “consciousness.” However, a word of caution is appropriate here; it seems plausible that the mind and our experiences of it are much more extensive than what is captured by what we ordinarily think of as “conscious awareness” (Panksepp, 2005).

One way to characterize the hard problem is to present it as a problem of *translating* objectivity into subjectivity, or the reverse. Chalmers (2003) offers two parallel formulations: (1) “How does experience (qualia) arise from a physical basis?” and (2) “Why does experience (qualia) arise from a physical basis?” The first of these seem directed more at a scientific answer, while the second formulation is likely more aimed at a philosophical answer.

There are other specific problems as well, which intimately relate to the “hard problem.” These include the “binding problem” and the “grain problem” (Feinberg, 2000), as well as the problem of how to square “determinism” with “free will.” The “binding problem” concerns how the brain brings scattered (visual, auditory, etc.) information located in different

parts of the brain to bind them together (Whitney, 2009) into psychological wholes (i.e., impressions, concepts, experiences). While the “binding problem” mainly appears to be a problem of neuroscience and psychophysics, it is not isolated from the philosophical problem related to giving shape to subjective experience.

The “grain problem” concerns the problem of how the “grainy” information of the brain becomes the “smooth” experiences of the mind; this is essentially the problem that occupied Bergson (1896) and prompted his seminal distinction between “time” (corresponding to the graininess) and “duration” (the smoothness of experience). The idea that the mind is experienced as a seamlessly united whole is widely accepted; as we experience life, each of us is always at “one mind.”

The “normativity problem” is a problem whose connection with the “hard problem” has gone less noticed but which is also implicated. The *normativity problem* can be formulated (paraphrasing Hume) as: How can the normative “ought” of the mind arise from the “is” of the brain? A similar formulation has been proposed by Railton (2009: 291–292). As a psychological term, normativity is inherently subjective and experiential. Normative outlooks make us *perceive* goodness in the external world (the cognitive view) or *ascribe* goodness to it (the affective view).

Dennett (1991) argues for a cognitive solution to the normativity problem in which he defends a strong version of material reductivism, according to which he claims that normative intentionality is built into the physical reality of the brain. From this perspective, sufficiently sophisticated robots would be able to become genuinely intentional creatures. This view implies that humans (and other animals) are not substantively different from robots. Achieving genuine consciousness reduces to a matter of sophistication of design. The causal

mechanism by which objective design could translate into subjective experience, however, is not known.

Meanwhile, Cash (2009: 134) argues that the reduction of normative content of the mind to natural facts cannot be accomplished, although he still maintains that normative perspectives remain characterizable in terms of naturalistic science. The normativity problem has often been discussed in connection with the questions of “free will,” “moral responsibility,” and “determinism”. It is beyond the scope of the present inquiry to make a detailed exposition of each of these notoriously difficult issues; it suffices for now to note that they appear to be *interconnected* and that they remain *unresolved*. The hard problem, the normativity problem, and the related problems of mind cited above suggest how philosophical questions and questions facing neuroscience dovetail and therefore need to be addressed concurrently.

3.3 Social science and natural science

The relation between the mind and the brain marks a junction between social science (e.g., psychology) and natural science (e.g., neuroscience), which respectively focus on the workings of the mind and on the workings of the brain. As already noted (Ch. 1), these two lines of inquiry are closely interdependent. These interdependencies may become clearer once philosophical quandaries about the mind-brain relation are sorted out. The mind-brain relation invokes the distinction between the *objectivity* of the physical brain and the *subjectivity* of the mind. In the previous section, I showed how this relationship forces us to face the “hard problem.”

However, the mind-brain relation also seems to involve *causality*, as indicated by Chalmers’s (2003) first formulation of the hard problem above. There are a number of competing theories of what shapes the mind. Some suggest that the mind is literally socially

constructed through social interactions with the external world (parallel positions for the social construction of emotions were presented in Ch. 2). However, most researchers appear to find the *causal claim* compelling. After all, if the brain matter does not influence the brain, then what *does*? And if the brain did *not* shape the mind, could the mind be sustained and live on *without* the brain? Concrete evidence to back up an alternative story to the causal claim seems virtually non-existent.

It seems fairly trivial that even social interaction presupposes the causal involvement of the brains of people who interact. Social interaction stories therefore do not represent genuine alternatives to the causal claim *per se*. Thus, it seems difficult to escape some formulation of a causality claim as explanatory of the mind. In other words, the physical brain *explains* the mind and *produces* it (Damasio, 1994: 90). In support of the dependency claim, Edelman and Tononi (2000: 15) make the following comment about subjective qualia: “We can analyze them and give perceptions of how they emerge, but obviously we cannot give rise to them without first giving rise to the appropriate brain structures and their dynamics within the body of an individual organism.”

The *causal dependence* of the mind on the underlying physical reality of the brain has a number of important implications. Notably, we cannot take just any given concept of the mind and hook it up with the brain; rather, we need an appropriately *refined* concept of the mind that *conforms to* physiological and biochemical facts about the brain. This is important because it means, among other things, that ordinary concepts of the mind cannot necessarily be taken at face value.

Several prominent researchers (e.g., Panksepp and Biven, 2012; Northoff, 2011; Bennett and Hacker, 2003; Changeux and Ricoeur, 2002; Solms and Turnbull, 2002) have realized the necessity of approaching the questions concerning the mind-brain relation in an

inter-disciplinary manner. Investigating the relation between the mind and the brain requires the integration of different perspectives, research tools, and approaches.

While it is beyond the scope of this thesis to explicate, much less settle, the age-old philosophical and scientific quandaries of “free will” and “determinism,” it will be useful to intersect with these questions because they concern the mind-brain relation and the role of affective emotions in moral deliberation. In what sense are our decisions *our* decisions? In what sense, if any, can we *freely* decide? Below, I analyze two different strands of dualism concerning the mind-brain relation.

3.4 Dualism about the brain and mind

Dualism theories regarding the mind-brain relationship can be found in many different forms. In contrast to reductive monism or material reductivism, dualism aims to keep both parts of the mind-brain relation within a single conceptual framework. Due to space considerations, I will focus only on two characteristically different dualist positions; the first position is *Cartesian dualism*, while the second is *dual-aspect monism*. The most striking difference between these positions is that whereas the former view holds that the mind and the brain are essentially independent, the latter view emphasizes dependence and material unity. Dual-aspect monism presents the mind and brain as the same “thing” (monism) while at the same time being separate in that the mind and brain represent fundamentally different “aspects,” “angles,” or “dimensions” (dualism) that resist reductive conflation. Below I present these two forms of dualism as well as some general arguments for dualism (i.e., against full-fledged naturalism and material reduction).

3.4.1 Cartesian dualism

Classical Cartesian dualism holds that the mind and brain are of different substances, which is a doctrine credited to the philosopher René Descartes. Sometimes this position is called “substance dualism,” or “spiritual dualism” (Changeux and Ricoeur, 2002: 14). Descartes thought the mind was an extra-corporeal entity that interacts with the human brain and expresses itself through a specific area of the brain called the pineal gland (Panksepp, 1998: 336; Smith, 1998). Descartes held that the mind was independent of the brain. This allowed him to argue that the mind or “soul” could be immortal. This idea was distilled in Gilbert Ryle’s (1949) famous metaphor “the ghost in the machine,” suggesting how the mind (the “ghost”) stands in relation to the brain (the “machine”).

Through Western philosophical, social, and scientific thought, Cartesian dualism has enjoyed immense influence (Adriaens and De Block, 2011: 290; Hayes, 2000: 2; Gabbard, 2000) and continues to do so in many disciplines and domains of practical affairs. For instance, in the law profession it underpins the idea of *mens rea*, which is concerned with what legally can be classified as a “guilty mind” (Curran and Gabriel, 2012; Greene and Cohen, 2004). Guilt can be ascribed because the mind is a self-contained unit of responsibility and seen as separate from the brain. The literature review (Ch. 2) showed that Solomon (2007) grounds moral responsibility along these lines, a conceptual treatment he also applies to business ethics (e.g., Solomon, 1992). In the contemporary sciences of the mind, Cartesian dualism has grown increasingly less popular (Modell, 2003), although there are still a few vocal supporters (e.g., Eccles, 1994).

The Cartesian type of dualism has come under attack from several competing theories. Against the Cartesian version of dualism there is the trivial point that the pineal gland does not have the functions Descartes ascribed to it (Smith, 1998); it is mostly occupied with such

tasks as regulating sleep patterns, not rationality or deliberative action. It would be facile to take this mistake as a decisive blow to Cartesian dualism. In principle it is not inconceivable that a specific location in the brain will be identified, although empirically it does not look that way at present. More crucial points of criticism focus on the idea of separating the mind and the brain (Panksepp and Biven, 2012; Smart, 2004) and the failure of Cartesian dualism to appreciate the neural correlates of mental activity.

There is also an argument against Cartesian dualism's treatment of emotions. Descartes's dualism excludes emotions completely from his notion of an independent faculty of rationality on the grounds that emotions pertain to the body whereas rationality characterizes the "soul" or "mind" (Damasio, 1994). Clinical studies provide disconfirmatory evidence, rejecting Cartesian dualism (Damasio, 2011; Damasio, 1994). The idea that rationality is emotionally cold and that rational thought controls everything at the top both appear suspect in light of current knowledge.

Most of the processes in the brain happen entirely without conscious awareness (Berlin, 2011), including an array of control mechanisms. Moreover, it is well-documented how automatic low-energy processes in the brain control the various high-energy processes in the human body (Panksepp, 1998: 336), and many current theories conceptualize the mind as a "representation" or "product of" the brain. This idea suggests that the mind is something akin to "the brain in action" rather than something independent of it. Hence, instead of the brain being the "tool" of the mind, it is the mind that is the "tool" of the brain; some current theories (epiphenomenalism) even suggest that the mind is merely a byproduct of the functions of the brain.

Dualist theory of the Cartesian type makes a speculative leap from the lack of an intimate experiential identification with our own brains to its alienation. The fact that we

cannot *feel* our own brain in any direct sense does not permit us to infer that whatever we do feel is separate from and independent of the physical matter of the brain. Another point that makes this leap unwarranted is that it implies the mistaken assumption that a given person *is* her conscious aware mind, whereas her brain *is not*. The lack of sensory stimuli from the brain itself does not warrant a detached conceptualization of what a person's "self" is.

The Cartesian idea of the "soul as conscious self" is confused in several ways. Above, I have already pointed out that the mind is more extensive than the relatively narrow scope afforded by "conscious awareness." Similarly, it is a conceptual mistake to think of the "self" as a unit because different senses of self can be disentangled, as suggested by contemporary theory in neuroscience. The "self" can more appropriately be seen as a *layered* concept (Panksepp and Biven, 2012; Damasio, 2010; Panksepp, 2005). A person's *sense of self* covers more than what she is consciously aware of (Solms and Panksepp, 2012; Damasio, 2010; Panksepp, 2005). The evidence suggests that a fundamental affective sense of self *precedes* cognitive forms of consciousness altogether (Solms and Panksepp, 2012).

Neuroscience currently suggests that an integrated cortical-subcortical (limbic) midline system is a neural substrate for the human "self" (Northoff et al., 2006), mediating between affective feelings of self and the self-referential self. The "ghost in the machine" metaphor is, as Ryle (1949) indicated, therefore inadequate as a model of the relation between the mind and the brain. The "self," although experienced as seamless and unified (Bergson, 1896), is sustained by hierarchically structured and shifting brain processes that draw on the concerted dynamics between multiple brain areas (Solms and Panksepp, 2012; Damasio, 2010).

3.4.2 Hard problems for naturalism

Conventional (general) naturalism (e.g., Dennett, 1991) reasons that since human beings are part of physical nature, our minds must be part of nature as well, and that hence nothing more can be said about the mind than what could be said on the basis of science. As I have already indicated, this view is too dismissive. In particular, conventional naturalism is too quick to disavow what can be said based on subjective experience, and the qualia of experience are not directly accessible to science (Horgan, 1999: Ch. 8; 1996; Chalmers, 1995).

Hence, I argue that it is possible and preferable to *accept* that facts about the mind are causally dependent on facts about nature while at the same time *resisting* the conclusion that this is all that can be said about the mind. The resulting position has specific advantages over conventional naturalism, as explained below. This picture of the mind-brain relation departs from conventional naturalism but also from Cartesian dualism. For instance, it will no longer be supposed that the mind or “soul” can exist independently of the brain. At the same time, a dualistic aspect is retained. I present this form of dualism further below. First, I will describe the problems that face naturalism. These problems are related to the “hard problem,” as mentioned above.

The so-called “knowledge argument” exposes one of the particular problems generated by conventional naturalism (or physicalism). Consider the following argument presented by Jackson (1982):

“Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specializes in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and uses terms like ‘red,’ ‘blue,’ and so on. She discovers, for example, just

which wavelength combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal cords and expulsion of air from the lungs that results in the uttering of the sentence ‘The sky is blue.’ [...] What will happen when Mary is released from her black and white room or is given a color television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then is it inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false” (Jackson, 1982: 130).

The “knowledge argument” suggests that physical and functional scientific knowledge are insufficient to account for experience. Experience is inherently subjective and perspectival, and these *internal* aspects seem inaccessible using knowledge of physical reality alone. In particular, the *qualities* of experience (“secondary qualities”, as John Locke called them) cannot be captured in scientific terms. Redness as an experience, for example, cannot readily be reduced to physical facts about redness. Redness is inherently a mental phenomenon, as Jackson’s example points out, and cannot be captured merely by scientifically clarifying the relevant empirical facts.

A second, related problem for naturalism concerns *agency*. Qualities are experienced *by* an “agent” or a “self” in the world. The knowledge *that* there is an agent and all the attributes of this agent seem insufficient as a means for establishing *what it is like* to be an agent. This can be called the “agency problem,” which in turn brings the discussion back to the “normative problem” (mentioned earlier). A human agent has a *normative outlook*. We can describe what this normative outlook does, how it functions, and perhaps even what its evolutionary causes were, but the *affective feelings* or sentiments as constitutive of a normative outlook seem to escape us regardless of how fine-grained and accurate the

description is. While affective feelings (at least in principle) can be *classified*, their specific subjective qualities remain elusive.

The strength of Jackson's argument above may not be so much that experience provides a different type of knowledge as that it provides an affective attachment to or appreciation of knowledge. In contrast with Jackson's perspective (1982), this interpretation may indicate that the central point is not that there is a different *kind* of physical knowledge that experience represents but rather that there is *something else* than knowledge. Hence, calling the argument the "knowledge argument" can be seen as a misnomer, and we could instead think of it as an "affect argument."

It should be noted that this reading blunts the frequent criticism against the "knowledge argument" that Jackson is inconsistent by simply helping himself to a conclusion that is ruled out by his own premises, notably including Jackson's (2003: 251) self-criticism. It can be argued that, in essence, cognitive information differs from affect, in line with the classical Humean (e.g., Stevenson, 1963) distinction between "facts" and "values" (i.e. positive and normative).

Similarly, cognitive beliefs can be said to track the positive according to standards of truth (i.e. a mind-to-world direction fit) whereas affect consists of urges and desires that seek satisfaction in the external world (i.e. a world-to-mind direction of fit) (Humberstone, 1992; Platts, 1979: 257). While cognitive information in principle can be fully described in terms of an external account, affect cannot because it is inherently internal and subjectively experiential. Again, this suggests that it is specifically the *affective* mental component that makes reduction of mind to brain problematic. Hence, Jackson's intuition stands but is backed by a different argument than the one he proposes.

Another seminal argument against the naturalist reduction of the mind to the brain is presented by Saul Kripke (1980). Kripke argues that although some forms of knowledge result from empirical discovery, they generally apply universally to all “possible worlds.” For example, scientists discovered that water is H₂O, and this knowledge extends to all conceivable worlds that are ontologically possible. It must hold for any possible world because, as Kripke puts it, H₂O is a “rigid designator” of water. Water is *necessarily* H₂O. Anything which is *truly* water must be H₂O, regardless of what other worlds are like. If there is water on a planet, we *know* that there is H₂O on it. Now, the question remains regarding how this reasoning works in the case of the relation between the mind and the brain.

Kripke argues that quite *unlike* the case with water and H₂O, the mind does *not* rigidly designate the brain. It is entirely possible, Kripke maintains, to construe possible worlds in which conscious minds are produced by some other substrate than a brain. This suggests that the concept of a mind cannot be successfully reduced to the brain in a straightforward manner as conventional naturalism presupposes.

It is important to notice that the arguments against reduction do not in any way prevent us from seeing the mind as *derived from* or *caused by* the brain. The causal claim that the mental is derived from the biological is not equivalent to, and does not imply, the claim that the mental (mind) can be equated with the biology of the brain (Pally, 2000: 2).

3.4.3 Dual-aspect monism

Dualism about the mind-brain relation is not limited to the rather restrictive Cartesian view described above. Departing from Descartes, philosopher Ricoeur (Changeux and Ricour, 2002: 15) accepts that mental experiences imply the corporeal but holds that experience nevertheless cannot be reduced to the objective subject matter studied in the natural sciences. Because this position directly connects the mind in a dependent relationship with the brain but

resists reducing the mind to the natural, it is in opposition to general naturalism, including its materialist versions (e.g., Dennett, 1991).

This less ambitious type of dualism, I surmise, equips us with a reasonable explanation of normativity in terms of affective experience (see Ch. 5). It allows for a conception of the mind as something that in principle can be described naturalistically according to relevant facts at the same time as its nature gives rise to an internal perspective that resists complete naturalistic description. It allows for a subjective dimension of experience and meaning that has to be *felt* in order to be captured.

According to this view, there are legitimate scientific questions as to how a normative perspective comes about, whereas it is doubtful whether science can say anything concerning the qualitative feel of how things look from a particular affective-normative perspective. The qualitative and experiential aspect of psychology, therefore, appears to be outside the scope of science. While naturalism, by naturalizing the mind, effectively ignores or denies the problem that subjective experience poses, empirically oriented dualists can argue that the “hard problem” needs to be taken seriously and that we need a way to conceptualize it that recognizes both horns of this scientific dilemma.

Some neuroscientists have recently arrived at the same type of dualism but without explicitly considering the connection to the phenomenon of normativity. Pondering the brain-mind relation, Solms and Turnbull (2002: 53) advocate the scientific theory they call *dual-aspect monism*: a position that has its intellectual origins in the writings of Spinoza. This position appears to be accepted by a growing number of neuroscientists (e.g., Zellner et al., 2011; Panksepp, 2005). It is similar or identical to the philosophical theory referred to as “neutral monism” (e.g., Chalmers, 2003; Carnap, 1928; also espoused at some point by

Bertrand Russell) and the position some refer to as “naturalistic dualism” (e.g., Chalmers, 1996).

Dual-aspect monism argues “the brain is made of stuff that *appears* physical when viewed from outside (as an object) and mental when viewed from the inside (as a subject)” (Solms and Turnbull, 2002: 56). This viewpoint (which is explored further in Ch. 5) suggests that an internal, normative, subjective perspective is distinct from the objective, scientific perspective that looks at the brain externally. Thus, while for example pain can be characterized objectively as the body’s protective mechanism identified as a complex neurological phenomenon, pain also corresponds to an internal experience whose qualitative aspect cannot be captured in neurological terms. This qualitative painfulness, moreover, is an *essential* part of the notion of pain (Chalmers, 1996: 147). Correspondingly, this relationship suggests the normative *badness* of pain is not an objective scientific fact but rather something dependent on the affectively painful experience.

The realization that subjective experience “is singular and unique, and is only observable by the subject himself or herself” is a key point (Solms and Turnbull, 2002: 295). Moreover, it can be argued that the “inner world of subjective experience, *as we experience it*, is as real as are apples and tables” (Solms and Turnbull, 2002: 297). Subjective experience therefore has to be taken considered within a holistic theoretical framework alongside the physical and chemical facts of reality. While proposing this dualism regarding perspectives (internal and external), Solms and Turnbull (2002) retain the idea that the mind and the brain refer to the same “thing” (hence, “monism”). Thus, one way of scientifically approaching what is going on in the mind is to investigate the electro-chemical processes that take place in the brain.

This theory presents a solution to the mind-body problem or equivalently to the mind-brain problem that Descartes tried to solve. According to Solms and Turnbull (2002: 57), the mind-brain problem (mind-body problem) “boils down to a *problem of observational viewpoints*, and the distinction between your brain and your body (between mind and matter) is therefore merely an artifact of perception.” On this view, as a *subject* you are your mind, whereas as an *object* you are your body. The precise mechanisms by which the brain produces the subjective mind, however, are still unknown to science. Nevertheless, the fact that they are unknown does not preclude progress from being made in this area.

From a philosophical vantage point, Bennett and Hacker (2003) corroborate the position of dual-aspect monism with the following argument (in the spirit of Wittgenstein):

“It is not that as a matter of fact brains do not think, hypothesise and decide, see and hear, ask and answer questions; rather, it makes no sense to ascribe such predicates or their negations to the brain. The brain neither sees, nor is it blind—just as sticks and stones are not awake, but they are not asleep either. The brain does not hear, but it is not deaf, any more than trees are deaf. The brain makes no decisions, but nor is it indecisive. Only what can decide can be indecisive. So, too, the brain cannot be conscious; only the living creature whose brain it is can be conscious—or unconscious. The brain is not a logically appropriate subject for psychological predicates” (Bennett and Hacker, 2003: 72).

In effect, this argument pins consciousness to internal subjective experience and objects to investing the brain with psychological powers only a sentient being can wield. By so doing, some neuroscientists commit “the mereological fallacy” (Bennett and Hacker, 2003: Ch. 3). Specific facts regarding how the brain works complicates this picture, however. For instance, the phenomenon of “blindsight” (Celesia, 2010; Frith, 2007; Wieskrantz, 1986) shows that

vision is not one single process in the brain but involves at least two separate processes. The brains of persons who have damaged their visual cortex can still detect or “see” stimuli, even though the person does not experience this vision, at least not consciously. When asked to point to the visual stimuli, a blind person (in controlled experiments) is able to do so far better than would be expected by chance; some patients can make inferences based on information and even detect letters and words subliminally (Berlin, 2011).

With this backdrop, Bennett and Hacker’s (2003) position can be disputed as there appears to ways in which the brain can see independently of the experience of the subject. However, in agreement with Bennett and Hacker (2003), there are *also* ways in which the subject can see, hear, and experience that are *not* fully captured by physical science. Moreover, a point will expand on later (see Ch. 5), is that we are not consciously aware of all that we experience.

It should be pointed out that this is a different interpretation of “experience” than is sometimes used by psychologists. For example, Frijda (2007: 71) makes the following claims: “Pleasure is only *felt* as pleasure after having withdrawn from full engagement”, and “Pleasure usually is not experienced until the work is finished, and one looks back on it, and sees that it is good.” This is an instance of the appraisal views of emotion that were discussed in the preceding chapter. In contrast, current neuroscience (Panksepp, 1998) suggests that pleasure also materializes *during* activity, but that our limited conscious awareness is often too busy to take notice of it.

This opens for an inherent duality of the mind. On the one hand there are the things that enter conscious awareness, on the other there are the things that do not enter awareness but that we nevertheless experience. William James was one of the first psychologists to explicate this type of dual of the mind. According to James (1890), we should distinguish

between “self as subject” (the “I” or Freudian “Id”, as a continuous subjective experience) and “self as object” (“me” or Freudian “Ego”, as an object of reference in thought). While recent theoretical development in affective neuroscience and psychoanalysis (Solms and Panksepp, 2012; Legrand, 2007) suggests that this is a crucial distinction for explaining how the human mind operates, cognitive neuroscience and cognitive psychology tends to ignore or be ignorant of this topic (Legrand, 2007: 588-592).

Following the scientific positions of dual-aspect monism, we have a reasonable way of separating between the science of decision-making and moral deliberation (the external perspective), on the one side, and the active *mode* of decision-making and moral deliberation on the other (the internal perspective).

To shed light on the mind-brain dualism, it can be helpful theorize about why we have such a duality of self as a subject and self as an object. What could the function and advantage of this built-in capacity for dualism be? Northoff (2011: 9) suggest that this type of distinction is necessary for higher order thinking. In order to be able to think about our self and objects in the external world, the brain must first be able to distinguish the brain from a “self”, and also the brain from objects. “The concepts of brain-self differentiation and brain-object differentiation refer to and describe those processes that enable and predispose to the constitution of both self and objects as distinct from the brain” (Northoff, 2011: 9).

3.5 Curtailment of the freedom of will

“Free will” is often, perhaps normally, taken as a prerequisite for an agent to be “morally responsible” or “accountable” for her actions. If decision-makers are not genuinely free, at least to some non-trivial extent, how can they be legitimately blamed for their decisions? In applied ethics, the possibility of free will is typically tacitly assumed, perhaps because the sense of ourselves as freely acting agents is so ingrained in human psychology. More explicit

commitment to free will is sometimes also seen in applied ethics. In business ethics, for example, Solomon (1992) explicitly assumes free will as a premise for moral agency, and Drascek and Maticic (2008) present “free will” as the core of normativity.

Psychologically, each of us apparently *feels* as though we have a conscious free will of our own (Wegner, 2002: 342). The feeling of free will is a feature of our subjective perspectives in its advanced self-reflexive form and includes not only actual and potential courses of action but also evaluative projections ascribing normative qualities (e.g., “good,” “bad,” “nice,” “nasty”) to perceived and imagined objects. However, simply noting that individuals have this *feeling* of free will does not answer the underlying question of whether we have *genuine* free will (Wegner, 2002: Ch. 1).

There is a trivial sense in which our will is not free. In our social decision contexts, our will can never be *completely* free because it is always conditional on our physical and social surroundings. As Ricoeur (Changeux and Ricoeur, 2002: 136) elegantly put it: “Men make their own history, but they do not make it just as they please; they do not make it under circumstances chosen by themselves, but under circumstances directly encountered, given and transmitted from the past.” Under dire conditions, we may not be able to satisfy our basic needs because we are unable to given those circumstances and our power to change them. However, evidence from neuroscience suggests that there is a more fundamental sense in which our will is not free, to which I will now turn.

3.5.1 Illusions and representations

Based on evidence about the how the brain works, it has been argued that our feeling of free will turns out to be a mere “illusion” (Frith, 2007: Ch. 2; Wegner, 2002); our brain does most of the thinking and deliberation about what to do *for us*, automatically, without our knowing. Our conscious brain has limited capacity and most of our thinking and behavior result from

the unconscious brain. Bargh and Chartrand (1999) estimated that roughly 95% of our behavior is unconsciously determined.

In this picture, the conscious mental world of our minds may only account for a tiny fraction of what the brain “does,” and consciousness is somewhat like icebergs floating along in a sea of unconscious brain processes. Wegner (2002: Ch. 2) cites evidence suggesting that conscious will arises from neurological processes that are anatomically distinct from the processes by which the mind generates action. Although this argument hinges on exactly how the brain is interconnected, it seems to suggest that our ordinary sense of ourselves as freely acting agents is exaggerated.

According to Wegner (2002), the illusion of free will is an integral part of a set of illusions that the brain creates (resulting from its evolutionary advantages), including the illusions of having a *self*, of having a *mind*, and of being *agents*. Because these illusions are so ingrained in our psychology, he argues, they are impossible to escape from. This concept is analogous with the established facts about vision (Frith, 2007); what we see represented in our minds is a result of heavily sifted information coupled with a range of specific constructions (i.e., fictive elements) that the visual cortex adds to the sensory information. The brain generates these illusions because they are generally helpful in coping with our external environment; we see not what is there but instead what helps us. Specific functional illusions may therefore be benign. At the same time, they may have implications for how we should think about morality, deliberation, and moral decision-making.

One of the specific points that evidence shows (Soon et al., 2008; Libet, 1985) is that, at the time we make decisions and seem to be in control of their making, the decisions have *already* been made. Hence, decisions seem to be made *before* cognitive consciousness even enters the stage. This is an important piece of evidence given its wide-ranging repercussions

in relation to human agency, responsibility, and free will. Massumi (2002: 29) describes Libet's studies as follows: "during the mysterious half second, what we think of as 'free,' 'higher' functions, such as volition, are apparently being performed by autonomic, bodily reactions occurring in the brain outside consciousness, and between brain and finger but prior to action and expression." While the evidence has been subject to much debate over the years, it has also increasingly gained support by evidence (Berlin, 2011).

This has important implications for views that place rationality and conscious awareness at the center of moral decisions and moral development (e.g., Kohlberg, 1984). MacIntyre (1999: 53-54), for instance, claims that "mature human rationality" requires the ability to consider exactly why we are doing what we are doing. If we can never reliably know precisely why we do the things we do, however, this rational capacity may be unattainable in practice. Self-reflective thought may still be central in our normative deliberation, but it will neither be infallible nor represent some normative "stamp of approval" form rationality.

Similar effects as those shown by Libet have also been detected in the domain of moral judgment, where our brains for example are able to tell whether a harmful action is intended or not *pre-consciously* (Decety and Cacioppo, 2012). The idea that the cognitive processes of the higher cortical brain structures form a genuinely independent unit of deliberation is illusory (Watt, 2003). This can be concluded from the fact that in life-threatening situations, great pain, or in states of extreme hunger or thirst, "homeostatic mandates dominate any and all cognitive processing and conscious focus" (Watt, 2003: 92).

It is moreover clear that we are always under the chemical influence of a cocktail of active neurotransmitters at any given moment, which except for special cases (e.g., taking medication) is not subject to choice. For instance, it is argued on the basis of experiments that

the hormonal neurotransmitter oxytocin promotes empathetic concern for victims and a tendency for pro-social approach behavior (Krueger et al., 2013); the lack of oxytocin appears to have an opposite effect (Carter, 2005). Similarly, serotonin, another neurotransmitter, influences normative reasoning and deductive thinking (Stollstroff et al., 2013). This evidence suggests that decision-makers are firmly in the grip of subconscious, non-volitional forces, in which affective emotional systems are central participants (see Ch. 4 and Ch. 6).

The obvious threat of the scientific position that has been sketched (i.e., Soon et al., 2008; Frith, 2007; Wegner, 2002; Libet, 1985) is *determinism*. Determinism is the view that everything that happens is causally predetermined and predicated in the physical state of the world. Determinism should be distinguished from fatalism, which is the view that the outcome (i.e., what will happen) is predetermined regardless of the actions of the agent. Determinism does not imply this because if an agent does something different, then a different result will occur. What determinism implies is rather that there is a fundamental sense in which the agent is not genuinely free to *choose* to do something different than what she ends up doing. This may appear to threaten our standard and widely shared views of “free will” and “moral responsibility.”

A related threat, which I will discuss later on (Ch. 7), is a form of relativism: (meta-ethical and normative) *nihilism*; appropriately or not, determinism is sometimes used as an argument for nihilism. For example, it is sometimes thought that if we are part of the natural physical flow of the world, all talk of genuine agency and value must be *illusory*. In facing such arguments, it is important to keep in mind that there is a difference between the factual data and what they tell us. The data themselves cannot be ignored, but at the same time they do not literally speak for themselves; data need interpretation. Scientific evidence also requires formulation in terms of philosophical and scientific theory (an argument of theory-ladenness, in the Popperian sense), and here there is arguably room for variety of opinion.

One of the key terms (e.g., Frith, 2007 and Wegner, 2002) that seems to warrant scrutiny is the term “illusion.” What they describe as “mere illusions” are not in fact *mere* illusions. Human beings are not randomly deluded, for example, and when “we” do things in the external world, we do so with a striking rate of success; we make appointments on time, for example, and meet the people we expected to meet at the mutually agreed upon location. It may be more appropriate to talk of “representations” (van Fraassen, 2008) than “illusions.”

The external world, thus, can be seen as *represented to us* by our brains, and our brains represent *us* to ourselves as agents acting with some degree of autonomy and authority. Most of these representations are useful, moreover, because they tend to help us to seek the things we (or our brains) see as desirable and avoid those which are not. In van Fraassen’s (2008: 20) terminology, our brain provides us with “predicates” (perspectives; secondary qualities) in place of “referents” (reality; primary qualities).

Representations may perhaps ultimately be all we have, but they stand in relation to reality. For example, when considering pictures, Fraassen (2008: 16) notes “it is hard to accept that a picture could fail to convey anything correct or true about something and still be a picture of that thing.” Thus, when we picture something in our minds, it is reasonable to assume that there is, however inaccurate, something truthful about the picture. From our internal, represented perspective, *representations* are all we have; therefore, it follows that *these* are what we have to take seriously in moral deliberations and decision-making.

Perhaps Frith (2007) and Wegner (2002) go too far in eliminating free will (and “free won’t”) altogether, as Midgley (2014: Ch. 9) suggests. It seems difficult to prove beyond reasonable doubt that there is never any involvement of anything like a will that is at least partially free. It may well be that an element of free will is always present in practical deliberation and decision-making. There is ongoing debate concerning free will and its

implications. Heisenberg (2009: 164) argues “[o]ur influence on the future is something we take for granted as much as breathing.” Roskies (2006), meanwhile, holds that the implications of many of the studies detailing the freedom of the will are overstated. Nevertheless, it is a sobering fact that the will is probably not free in the fullest sense the term, which is often the sense used or assumed in practical ethics, philosophy, and elsewhere. The human mind appears to be controlled by the automaticity of subcortical forces, of which we are completely oblivious, to a much greater extent than is generally appreciated.

Even if free will turned out to be, in part, an illusion, it could be argued (Midgley, 2014) that the *feeling* of free will, control, agency, a self, and responsibility nevertheless constitutes a vital part of our normative moral outlooks. The feeling that we are the *agents* amidst a sea of “cold” physical causality seems central for conferring *personal meaning* to our own actions and decisions and indeed to recognize them as our own. Moreover, we are inclined to *take moral responsibility* for our actions and decisions; we do not, for instance, tend to blame our brains for what we do.

While objective science usefully informs us about how the brain works, it seems it is unable to show us either what it is like to *have* a brain or how we should *use* it. This information is fundamentally speaking what affective subjective experience gives us (Solms and Panksepp, 2012). While science is not in a position to fully explain the subjective mind, we can make scientific theories that provide room for this explanatory gap.

3.5.2 Putting the free will debate on hold

In this section, I will clarify a tentative position concerning free will and my arguments for it. Undoubtedly, the outcome of the free will debate will be extremely important for the concept that is the central topic of this study: *normativity*. Moral notions, such as “moral responsibility” and “blameworthiness,” can be seen as more specific topics within

normativity, and these action-oriented normative concepts are conventionally considered to take free will as a fundamental premise.

In moral philosophy and epistemology, there is extensive literature regarding questions of free will. I will only briefly examine a few general themes from this literature; my reason for doing so is to place some philosophical worries on hold. The most central debates in the literature on free will revolve around the following questions: (1) whether or not *determinism* and *free will* are compatible, (2) whether or not *free will* and *moral responsibility* are compatible, and (3) whether or not there is a causal relationship from *determinism* to *freedom* or *absence of freedom* of the will.

Thus, if, for example, it could be shown that determinism is true, and also that moral responsibility is incompatible with determinism, there would be no basis for justifying moral responsibility. However, there would nevertheless still be psychological, albeit objectively unjustified, normative perspectives and experiences of moral responsibility. While the conception of an objectively justified normative ethics would have to be abandoned, it does not follow that there is no meaningful way of thinking in terms of justice and justifiability in moral deliberation.

A problem that haunts the debates of determinism, free will, and moral responsibility is that the basic philosophical positions that are taken on the issues are *unfalsifiable*. More precisely, while it is possible that they are falsifiable in theory, they are not falsifiable in practice. Centrally, it seems we have no way of determining whether or not full determinism is a true description of the state of affairs. How can it be proved *ex post* that an agent as a matter of fact *could* or *could not* have acted differently than how she did? Invariably, we shall only know for certain that she could do what she did, not what she alternatively could have done but did not do.

Are there no genuine choices or decisions? Are subjectively based normative authority and agency just illusions? There seems to be no practical way of obtaining reliable answers to these questions. We do not even know what would settle the debate on determinism and free will because we do not know the mechanisms by which objectivity is converted into subjectivity, so we have the problem of understanding subjectivity scientifically. We do know that it is there, however, subjectively.

As a consequence, I think it would be premature to insist that there is a genuine justified sense in which agents are morally responsible or blameworthy, but also to insist that there is no such sense. The current state of knowledge seems to favor an *agnostic* view. This agnostic view does not, however, make our practical decision problems go away as we experience them, nor does it imply any particular normative view with regards to solving or resolving them (see Ch. 7). In particular, we have no reason not to take our own normative perspectives, as well as those of others, *seriously* in our practical moral deliberations.

Hence, I think our subjectivity gives us *some* justification (see Ch. 9) for seeing ourselves and others as moral agents. In the same vein, we seem justified to continue our subjective practices of holding ourselves and others morally responsible for actions we see as moral and to provide and request reasons underpinning these practices and judgments. For example, we may reason that we can blame an agent for doing something we consider morally wrong (e.g., stealing) when the action of the agent did not result from ignorance (e.g., she knew she was stealing), while knowing that we are merely projecting merely our evaluative moral attitudes based on our affective emotional responses (i.e., towards stealing) that we are projecting. In case determinism *is* true, the agent performed an act we projected disapproving attitudes towards; in case determinism is *not* true, this is still also where we end up (but we and the agent got us there). Determinism may also be only partially true. In either case, nothing changes in practical terms. Philosophy provides analytical resources to reflect

systematically on such practical issues but seems to lack the resources necessary to resolve them in a definite manner.

3.6 The brain and the body: Two representations

The brain appears to be *exceedingly social* and it is not isolated but *embodied*, and the body is *embedded* in the environment (Solms and Turnbull, 2002: 18; Edelman and Tononi, 2000). The brain is connected both with the inner world of other bodily organs and, through our sensory apparatus, the external world (Solms and Turnbull, 2002: 18–19). While the internal world is the source of our various needs, the external environment is where we have to cope and find the resources to satisfy our needs (Solms and Turnbull, 2002: 19). Social relations, norms, identities, culture, etc., all influence the mind invariably by engaging the processes of the brain.

The nature of the mind may also give us cues about the brain. If our mind is a product of the brain, and the mind is highly social, we seem to be in a position to infer that the brain must be correspondingly social as well. It is broadly agreed that the mind is socially oriented and socially interactive. The social nature of the mind thus indicates that, given the tight brain-mind dependency, the nature of the brain is also social. However, evidence suggests that the brain serves a dual function; it monitors the inner world as well as the external one.

Our brains receive input from our internal bodily world as well as our external social and physical environment. This process happens in complex multi-process ways and is almost entirely without our conscious cognitive involvement. A rudimentary and adequately accurate understanding of this interaction is necessary to correctly understand how the brain comes to create the mind, as it is experienced in our mental lives, and as it is used in making practical decisions and pursuing their goals.

The body is represented in our brains in two distinctly different ways. *One* representation might be called the somato-topic body image (Solms, 2013; Solms and Turnbull, 2002). This is essentially a *map* of the body that is projected onto the surface of the brain. This map connects the relevant *sensory inputs* collected from the surface of the body, with specific designated areas of the brain. The brain zones of this map, however, do not only map the body itself but all the objects encountered in the external world through our senses. As classical empiricist philosophers George Berkeley and David Hume reminded us, our knowledge of the external world, in one way or another, ultimately is obtained through our senses. The map involves two types of projections with different designated brain zones: one for sensory input and one for muscular activity (motor input). However, it should be noted that all movement necessarily gives rise to sensations, and that these two zones, although separate, are also intimately interconnected.

Between the extra-sensory zones of the brain is something neuroscientists call “the convergence cortex.” This area of the cortical brain puts together the input from the extra-sensory zones into coherent sensory representations.

The other representation of the body arrives not at the top of the brain but from below, into the sub-cortical regions of the brain. The designated limbic areas monitor and regulate the autonomic state of the body (the autonomic nervous system). These areas operate essentially automatically outside of conscious control and direct cognitive awareness. Their role is not passive and subservient, however, for they also affectively *arouse* the higher cortical structures of the brain. They do so to meet the “needs” of our body, or, we might say, “our needs,” in the outside world.

3.7 Conclusion

Current scientific theory suggests that the entire mental apparatus of human beings is created by the brain in the process of fusing external somatosensory input with commands from internal representation and regulation processes generated by affectively based needs (Solms and Turnbull, 2002; Panksepp, 1998), which are possibly in turn based on homeostatic needs for the regulation of various tissues of our bodily organs within appropriate ranges (Damasio, 2010)—for example, not “too cold,” not “too warm.” The brain itself is a highly social organ (Edelman and Tononi, 2000; Brickner, 1944), which directs its (electrical) energies at coping in the external world in order to satisfy the needs of the organism (Damasio, 2010; Solms and Turnbull, 2002).

Most of the processes that take place in the brain happen outside the relatively limited domain of conscious awareness, experience, and volitional control (Damasio, 2010; Wegner, 2002; Panksepp, 1998). In deliberation and human morality, however, subjective experience, consciousness, and the brain-generated sense of free will is all we have available at our disposal in our feelings and perceptions of ourselves as deliberating agents.

Chapter 3 References

- Adriaens, P. R. and De Block, A. (2011) *Maladapting minds: Philosophy, psychiatry, and evolutionary theory*. Oxford: Oxford University Press.
- Bargh, J. A. and Chartrand, T. L. (1999) The unbearable automaticity of being. *American Psychologist* **54** (7): 462-479.
- Bennett, M. R. and Hacker, P. M. S. (2003) *Philosophical foundations of neuroscience*. Oxford: Blackwell Publishers.
- Bergson, H. (1911) *Matter and memory*. (N. M. Paul and W. S. Palmer, Trans.). London: George Allen and Unwin. (Original work published 1896).
- Berlin, H. (2011) The neural basis of the dynamic unconscious. *Neuropsychanalysis* **13** (1): 5–31.
- Brickner, R. M. (1944) Man and his values considered neurologically. *Journal of Philosophy* **41** (9): 225–243.
- Cash, M. (2009) “Normativity is the mother of intention: Wittgenstein, normative practices, and neurological representations. *New Ideas in Psychology* **27**: 133–147.
- Chalmers, D. (2003) Consciousness and its place in nature. In S. P. Stich and T. A. Warfield (Eds.), *Blackwell guide to the philosophy of mind* (pp. 102–142). Malden: Blackwell.
- Chalmers, D. (1996) *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.
- Chalmers, D. (1995) Facing up to the problem of consciousness. *Journal of Consciousness Studies* **2** (3): 200–219.

- Carnap, R. (1928/1968) *The logical structure of the world/Pseudoproblems in philosophy*. Berkeley: University of California Press.
- Carter, C. S. (2005) The chemistry of child neglect: Do oxytocin and vasopressin mediate the effects of early experience? *Proceedings to the National Academy of Sciences* **102** (51): 18247–18248.
- Celesia, G. G. (2010) Visual perception and awareness: A modular system. *Journal of Psychophysiology* **24** (2): 62–67.
- Changeux J. and Ricoeur, P. (2002) *What makes us think? A neuroscientist and a philosopher argue about ethics, human nature, and the brain*. Princeton: Princeton University Press.
- Curran, G. and Gabriel, R. (2012) Affective neuroscience and law. *Journal of Consciousness Studies* **19** (3–4): 36-48.
- Damasio, A. (2010) *Self comes to mind: Constructing the conscious brain*. Random House.
- Damasio, A. R. (1994) *Descartes's error: Emotion, reason, and the human brain*. New York: Putnam.
- Decety, J. and Cacioppo, S. (2012) The speed of morality: A high-density electrical neuroimaging study. *Journal of Neurophysiology* **108** (11): 3068–3072.
- Dennett, D. (1991) *Consciousness explained*. Boston: Little Brown.
- Drascek, M. and Maticic, S. (2008) What managers could see in the philosophical block of “free will.” *Journal of Business Ethics* **81**: 1–14.
- Eccles, J. C. (1994) *How the self controls its brain*. Berlin: Springer-Verlag.

- Edelman, G. M. and Tononi, G. (2000) *A universe of consciousness: How matter becomes imagination*. New York: Basic Books.
- Feinberg, T. (2000) The nested hierarchy of consciousness: A neurobiological solution to the problem of mental unity. *Neurocase* **6**: 75–81.
- Frijda, N. H. (2007) *The law of emotions*. Mahwah: Lawrence Erlbaum Associates.
- Frith, C. (2007) *Making up the mind: How the brain creates our mental world*. Oxford: Blackwell Publishers.
- Gabbard, G. O. (2000) A neurobiologically informed perspective on psychotherapy. *The British Journal of Psychiatry* **177**: 117–122.
- Greene, J. and Cohen, J. (2004) For the law, neuroscience changes nothing and everything. *Philosophical Transactions: Biological Sciences* **359**: 1775–1785.
- Hacker, P. M. S. (2011) *Human nature: The categorical framework*. New York: John Wiley & Sons.
- Hayes, N. (2000) *Foundations of psychology*. Cengage Learning EMEA.
- Heisenberg, M. (2009) Is free will an illusion?. *Nature* **459**: 164–165.
- Horgan, J. (1999) *The undiscovered mind: How the human brain defies replication, medication, and explanation*. New York: Touchstone.
- Horgan, J. (1996) *The end of science: Facing the limits of knowledge in the twilight of the scientific age*. Reading: Addison-Wesley.
- Humberstone, I. L. (1992) Direction of fit. *Mind* **101** (401): 59-83.

- Jackson, F. (2003) Mind and illusion. In A O'Hear (Ed.), *Minds and persons*. Cambridge: Cambridge University Press: 251–271.
- Jackson, F. (1982) Epiphenomenal qualia. *Philosophical Quarterly* **32**: 127–136.
- James, W. (1890) *The principles of psychology*. New York: Henry Holt.
- Kim, J. (1993) *Supervenience and mind: Selected philosophical essays*. New York: Cambridge University Press.
- Kohlberg, L. (1984) *The psychology of moral development: The nature and validity of moral stages*. New York: Harper & Row.
- Kripke, S. (1980) *Naming and necessity*. Cambridge: Harvard University Press.
- Kruger, F., Parasuraman, R., Moody, L., Twieg, P., de Visser, E., McCabe, K., O'Hara, M., and Lee, M. R. (2013) Oxytocin selectively increases perceptions of harm for victims but not the desire to punish offenders of criminal offenses. *Social Cognitive and Affective Neuroscience* **8** (5): 494–498.
- Legrand, D. (2007) Pre-reflective self-as-subject from an experiential and empirical perspective. *Consciousness and Cognition* **16**: 583–599.
- Libet, B. (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral & Brain Sciences* **8**: 529–566.
- MacIntyre, A. C. (1999) *Dependent rational animals: Why human beings need the virtues*. Chicago: Open Court.
- Massumi, B. (2002) *Parables for the virtual: Movement, affect, sensation*. Durham, NC: Duke University Press.

- Midgley, M. (2014) *Are you an illusion? (heretics)*. Stocksfield: Acumen Publishing.
- Modell, A. H. (2003) *Imagination and the meaningful brain*. Cambridge: The MIT Press.
- Moore, G. E. (1903) *Principia ethica*. Cambridge: Cambridge University Press.
- Northoff, G. (2011) *Neuropsychanalysis in practice: Brain, self and objects*. Oxford: Oxford University Press.
- Northoff, G., Heinzl, A., de Greck, M., Bermpohl, F., Dobrowolny, H., and Panksepp, J. (2006) Self-referential processing in our brain: A meta-analysis of imaging studies on the self. *NeuroImage* **31** (1): 440–457.
- Pally, R. (2000) How brain development is shaped. In Pally, R. and Olds, D. D., *The mind-brain relationship* (pp. 1-18). New York: Other Press.
- Panksepp, J. and Biven, L. (2012) *The archeology of mind: Neuroevolutionary origins of human emotions*. New York: Norton.
- Panksepp, J. (2005) On the embodied neural nature of core emotional affects. *Journal of Consciousness Studies* **12**: 158–184.
- Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.
- Platts, M. (1979) *Ways of meaning*. London: Routledge and Kegan Paul.
- Railton, P. (2009) *Facts, values, and norms: Essays toward a morality of consequence*. Cambridge: Cambridge University Press.
- Roskies, A. (2006) Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences* **10** (9): 419–423.

- Ryle, G. (1949/2002) *The concept of mind*. Chicago: University of Chicago Press.
- Skinner, B. F. (1953) *Science and human behavior*. New York: Free Press.
- Smart, J. J. C. (2004) The identity theory of mind. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. (Fall 2004 Edition). Stanford: Stanford University Press.
- Smart, J. J. C. (1959) Sensations and brain processes. *Philosophical Review*, **67**.
- Smith, C. U. M. (1998) Descartes's pineal neuropsychology. *Brain Cognition* **36**: 57–72.
- Solms, M. (2013) "The unconscious" in psychoanalysis and neuropsychology. In S. Akhtar and M. K. O'Neil (Eds.), *On Freud's "The Unconscious."* London: Karnac Books.
- Solms, M. and Panksepp, J. (2012) The "id" knows more than the "ego" admits: Neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Sciences* **2**: 147–175.
- Solms, M. and Turnbull, O. (2002) *The brain and the inner world: An introduction to the neuroscience of subjective experience*. New York: Other Press.
- Solomon, R. C. (2007) *True to our feelings*. Oxford: Oxford University Press.
- Solomon, R. C. (1992) Corporate roles, personal virtues: An Aristotelian approach to business ethics. *Business Ethics Quarterly* **2** (3): 317–339.
- Soon, C. S., Brass, M., Heinze, H. J., and Haynes, J. D. (2008) Unconscious determinants of free decisions in the human brain. *Nature Neuroscience* **11**: 543–545.
- Stevenson, C. L. (1963) *Facts and values*. London: Yale University Press.

- Stollstroff, M., Bean, S. E., Anderson, L. M., Devaney, J. M., and Vaidya, C. J. (2013) Rationality and emotionality: Serotonin transporter genotype influences reasoning bias. *Social Cognitive and Affective Neuroscience* **8** (4): 404–409.
- van Fraassen, B. C. (2008) *Scientific representation: Paradoxes of perspective*. Oxford: Oxford University Press.
- van Fraassen, B. C. (2002) *The empirical stance*. London: Yale University Press.
- Watt, D. F. (2003) Psychotherapy in an age of neuroscience. In J. Corrigan and H. Wilkinson (Eds.), *Revolutionary connections: Psychotherapy and neuroscience* (pp. 79–116). London: London: Karnac Books.
- Wegner, D. M. (2002) *The illusion of conscious will*. Cambridge: MIT Press.
- Weiskrantz, L. (1986) *Blindsight: A case study and implications*. Oxford: Oxford University Press.
- Whitney, D. (2009) Neuroscience: Toward unbinding the binding problem. *Current Biology* **19** (6): R251–R253.
- Zellner, M. R., Watt, D. F., Solms, M., and Panksepp, J. (2011) Affective neuroscientific and neuropsychanalytic approaches to two intractable psychiatric problems: Why depression feels so bad and what addicts really want. *Neuroscience and Biobehavioral Reviews* **35**: 2000–2008.

Chapter 4: The emotion location argument: Emotions originate in the subcortical affective brain

4.1 Introduction to the emotion location argument

This chapter gives a non-technical presentation of the biological infrastructure of emotions illuminated by key findings in affective neuroscience. This exposition especially concerns the *location* and *origin* of human emotions in the brain. The output of this chapter is a scientifically based platform that lays the foundation for the remaining chapters. By understanding the rough anatomy and neural circuitry of emotions and the causal effects of emotional systems, it is possible to better understand the nature of emotions and their role in consciousness and normative psychology (discussed in Ch. 5). As stipulated in prior chapters, this chapter indicates how normativity can be *co-located* with affective emotions. Normative decision-making and moral judgment (discussed in Ch. 6) requires the involvement of emotions (Lawton et al., 2009; Blackburn, 1998; Damasio, 1994; Zajonc, 1980), and the current chapter is important for understanding what sort of involvement this is.

The emotion location argument presents empirical evidence and current scientific theory that locates human emotions in the brain, more specifically in the *affective, subcortical* layer of the brain. The field known as affective neuroscience is the most central contributor of evidence-based theory on this particular topic. The suggestion that emotions have a basis in the subcortical brain and that they are inherently affective prior to cognitive processing differs from mainstream psychological theories of emotions, which tends to endorse a view of emotion as grounded in cognition (Ch. 2). Insofar as emotions are relevant for understanding morality and normativity, this suggests that affective neuroscience represents an important contribution to understanding practical moral deliberation and links applied ethics to our subcortical emotional brain structures.

4.2 Science of emotion

As we have seen (Ch. 2), philosophers have a long history of discussing emotions and the affective aspects of human nature, but at the same time remarkably strong traditions exist that neglect or reject this arena of human psychology. As previously noted (Ch. 1), the center of gravity in contemporary moral philosophy appears to lean toward conceptual analysis based on *a priori* assumptions. As argued before (Ch. 3), intuition-driven conceptual analysis is by itself incapable of establishing empirical facts. If we wish to understand the psychology of normativity, however, we have to appreciate facts established by psychology as well as facts regarding its biological underpinnings as established by biology and neuroscience (Ch. 1).

Historically, we find a number of empirically oriented philosophers who anticipated recent scientific development remarkably well, suggesting that emotions are affective (e.g., Mencius, Hume, Darwin, Nietzsche, Bergson, Russell). Other philosophers have proposed views that have exerted considerable influence, although they lead in other directions concerning the nature and relevance of emotions (e.g., Plato, Descartes, James). Such intellectual fault lines remain (Prinz, 2007: 13). Philosophy still shows considerable resistance to establishing a biologically based account of the origin of emotions (Charland, 2005), especially one that locates emotion in the deeper parts of the brain over which we have essentially no direct control (e.g., Solomon, 2007; Green, 1992). However, there are indications that philosophical perspectives are changing in these respects (e.g., Churchland, 2011; 2002; Northoff, 2004).

Before the onset of contemporary brain science, brain processes essentially represented a “black box.” As previously noted (Ch. 2), behaviorists like Skinner, Watson, and Ryle argued that the black box was the appropriate way to think well into the 20th century. Comparatively less was known about the brain at that time. While the rudimentary anatomy of

the brain could be charted, almost nothing was known with confidence about how its parts worked in concert. Instead of making reasonable conjectures about what is *inside* the “black box,” many philosophers and psychologists focused on what could be observed outside the box (i.e., behavior). This trend led to a tradition of summarily denying the existence of emotions (behaviorism and material eliminativism) and later, for conceptualizing emotions in terms of what is *outside* the “black box,” such as language and culture (e.g., social constructivist cognitivism).

The historical legacy has left a theoretical imprint of ideas such as: (1) Emotional states of mind are cognitive, or at least necessarily include an important cognitive component. (2) Emotions must arise from the cortical structures of the brain. (3) Emotions must be cognitively conscious and parts of explicit thought. (4) Affective states of mind can only arise from foregoing cognition. (5) The Cartesian idea suggests that states of mind are causally unrelated to states of the brain.

The evidence presented in this chapter indicates that each of these points, which have been advocated by adherents of the cortico-centric tradition left by the “cognitive revolution” (presented in Ch. 2), is untenable. If that is the case, *affective neuroscience* represents a paradigm shift regarding our understanding of emotions; “the cognitive revolution is gradually giving way to an emotion revolution” (Panksepp, 2003: 5).

4.2.1 Emotions without a cortex

A particularly powerful empirical argument for the subcortical origin of emotions (briefly mentioned in Ch. 2) is that human beings *without a cortical brain can still express emotions*. This can be readily observed—directly as well as through sophisticated brain imaging—in children who are born without a cortex (Damasio, 2010; Merker, 2007). These seminal observations (Bower, 2007) are about as hard scientific evidence as one can hope to obtain

with science (Panksepp and Biven, 2012). They strongly support the conclusion that “primary-process affective states do not require neocortical reflective capacities” (Panksepp et al., 2012: 12). Corroborating results have been shown elsewhere; for instance, Damasio et al. (2013: 842) found that a patient “whose insular cortices were entirely destroyed, experienced body feelings as well as emotional feelings.” These concepts also suggest that our sense of self is not cortically generated (Damasio, 2010) and that the “primary self” (Panksepp, 1998), or ‘proto-self’ (Damasio, 2010), is fundamentally affective (Solms, 2013; Solms and Turnbull, 2002).

Many children who are born without any functional cortex whatsoever (cases of *anencephaly*) appear fully capable of expressing basic emotions (Merker, 2007; Shewmon et al., 1999). Numerous such cases have been reviewed scientifically, and on this basis the presence of rudimentary consciousness (Shewmon et al., 1999) as well as affective emotions is documented beyond reasonable doubt (Merker, 2007). Many of these children are capable of expressing all stereotypical emotional patterns associated with primary emotions (Solms, 2013: 108; Panksepp and Biven, 2012; Panksepp, 1998). This evidence seems to rule out the view that the involvement of cognition and the associated cortical regions of the brain are necessary requirements *for* emotions or necessary components *of* emotions.

Based on firsthand observation, Merker (2007: 30) notes that these children “express pleasure by smiling and laughter, and aversion by ‘fussing,’ arching of the back and crying (in many gradations), their faces being animated by these emotional states.” In some cases, there are even clear indications of “coherent interaction between environmental stimuli, motivational-emotional mechanisms, and bodily actions” (Merker, 2007: 30). In addition, “The fact that the cortex is absent in these cases proves that core consciousness is both generated *and felt* subcortically” (Solms, 2013: 111). Empirically based suspicions that this

might be the case date back at least two centuries to Huxley's (1870) experiments with decorticate frogs.

4.2.2 Emotions and evolution

Why do we have emotions? It is instructive to see emotions, as Charles Darwin did, as a product of evolution (Panksepp and Biven, 2012; Ekman, 1992; Nesse, 1990). According to Panksepp and Biven (2012), emotions arise from parts of the brain that are evolutionary old, notably the mid-brain and sub-cortical structures of the brain (the extended limbic system), and not from the relatively more recent neocortex (Panksepp, 1998; MacLean, 1990). These older emotional areas of the brain we share with all other mammals (Panksepp, 1998), and the oldest of these parts also with all vertebrates.

The sub-neocortical brain has highly specialized functions. In contrast to sub-cortical structures, current empirical evidence suggests “the neocortex was not modularized by evolution but rather becomes specialized for diverse cognitive activities through developmental landscapes” (Panksepp, et al., 2012: 11). This information indicates that socio-cultural impacts have a considerable role in shaping the content of the human mind—greater than for example supposed by Chomsky's (1986) “nativist theory.” During “infancy, evolved subcortical functions conspire with life experiences to guide the development of mental functions” (Panksepp, 2015: 111). Thus, “nature” might have selected for considerable responsiveness to “nurture,” and the evolutionary perspective on the anatomy of the brain elucidates how this event might have happened. Hence, “notions of dichotomy between mind versus brain, nature versus nurture, have been supplanted by a rich web of synergistic relation between mind and brain, nature and nurture” (Pally, 2000: 1).

The evolutionary perspective is also helpful because it suggests that emotions have a built-in “purpose” in that they perform tasks that are generally important for survival (Zellner

et al., 2011: 2000–2001). Evolution can explain how such a built-in purpose or function has come about; it favors traits conducive to the survival of the species. It is logical to assume that the primary emotional systems are essential for the homeostasis and thus survival of primitive animals. A seeking system is necessary for exploring the surroundings; a fear system is essential for avoiding danger, while a lust system is required for reproduction, etc. Through evolution, the emotional infrastructure is then used in new and different ways, partly responding to needs that differ from those functions that were selected for in the ancient times of our pre-human ancestry.

Concerning human morality, it is crucial to point out that normative conclusions cannot be validly inferred from an understanding of the evolutionarily ordained emotional infrastructure alone (Farber, 1994). Preceding Darwin, Hume cannot be presumed to have thought in evolutionary terms, but he was adamant that an understanding of human nature by itself cannot logically provide us with normative conclusions. Evolution, biology, and neuroscience can help us understand how normativity *arises in us*, but it does *not lend direct objective support* for what normative feelings tell us to do. As Singer (1982) alerts us, even a prominent biologist of the stature of E. O. Wilson stands guilty of this type of invalid inference. This is, as previously noted (Ch. 1), not to be confused with the claim that we *can* reach normative conclusions whose legitimacy stand in relation to our factual normative feelings; giving us *internally* based normative justification.

It is sometimes suggested that human beings are special and very different from other animals. This distinction is often attributed to our distinctively human cognitive abilities, including such remarkable features as our sophisticated linguistic capabilities and our complex moral reasoning (Elm and Radin, 2012). This does not contradict, however, that we share our deep emotional infrastructure with other mammals (Panksepp, 1998: 17–18). As Darwin (1906: 193) noted, the difference between humans and our closer mammalian

relatives “is one in degree and not in kind.” One important difference is decision-making abilities, which we will discuss later (Ch. 6). First, however, we should appreciate the extent of our similarity.

We may not even be unique in having an evolved sense of morality. Chimpanzees, for example, have a sense of retributive justice, habitually punishing those who fail to return favors (de Waal, 1996). Even bats (Wilkinson, 1984) appear to engage in altruistic behavior. As Midgley (1978) aptly points out, to say that we are “social animals” is to indicate more than that we are *like* animals. Therefore, as MacIntyre (1999: 5) admits, human *morality* should not be seen as independent of our *animal* nature, which can throw light on what motivates us, what drives us, our limitations, and our morality. This concept resonates with early emotivist philosophy, including Hägerström (1917/1953: 152), who argues that experiencing emotional feelings is necessary for conceiving of an action as morally “right” or “wrong,” closely connecting an understanding of morality with a naturalistic grasp of emotions. Affective neuroscience sheds light on this area of inquiry from the perspective of the science of human emotions.

4.2.3 The triune brain

As was briefly mentioned (Ch. 3), the human brain can be divided into roughly three main parts, which has been captured by the concept of *the triune brain* (MacLean, 1990). This concept is important “for understanding the biological roots of human social behavior and communication” (Ploog, 2003: 487), and how human morality is shaped by the affective processes of the brain (Narvaez, 2008). The three parts of the triune brain are “the reptilian complex” (brain stem), “the paleo-mammalian complex” (the limbic system), and “the neo-mammalian complex” (neocortex) (Panksepp, 1998: 43; MacLean, 1990: 10). These parts are arranged according to their emergence in the history of human evolution (Panksepp and Biven, 2012: Ch. 2; MacLean, 1990).

The emotional centers of the brain are all parts of the “extended limbic system” (Panksepp, 1998: 341). At the bottom of the brain we find the evolutionarily older parts, which are virtually identical in mammals and vertebrates (Panksepp and Biven, 2012). Basic consciousness associated with subjective experience is likely to give rise to simple feelings of pleasure and unpleasure (Solms and Turnbull, 2002). Consciousness, as noted, has been shown to emanate from the phylogenetically ancient brain stem (Merker, 2007), specifically the reticular activation system. This is at least 525 million years old, since it likely emerges with the appearance of vertebrates (Shu et al., 1999). Humans also share some of the older primary emotions with the reptiles. These are located deep down in the brain and include the negative (negatively experienced) emotions “fear” and “anger,” but also the positive (positively experienced) emotions “seeking-desiring” and “lust” (Panksepp, 1998). The oldest distinct emotional system is the seeking-desiring system which is the source of basic anticipatory motivation.

The remaining emotional centers that are currently known essentially consist of components of the “paleo-mammalian complex” (Panksepp, 1998), which involves the limbic and para-limbic systems and is located slightly higher up than the old “reptilian” brain. Typical mammalian emotions such as “care” and “playfulness” (Panksepp, 1998) are at least 225 million years old, insofar as they arise with the emergence of mammalian species (Datta, 2005). Anatomically and neurochemically, these emotional parts of the brain are remarkably similar across all mammalian species (Berridge and Kringelbach, 2013; Panksepp, 1998).

The evolutionary newer neo-cortex, meanwhile, is relatively similar in humans and closely related primates, but the human neo-cortex is notably larger. The special human frontal lobes are only roughly 200 thousand years old and thus phylogenetically rather recent. It is widely accepted that the neo-cortex is essential for making sophisticated human decision-making abilities possible; it is less widely acknowledged that the lower levels of the brain also

play a critical role in sophisticated decision-making (Damasio, 2010). Current theory in affective neuroscience suggests that consciousness and emotional input must be fed from subcortical regions in order for the neo-cortex to be able to function and engage in sophisticated activities (Solms and Panksepp, 2012), such as sequencing time, planning ahead, and representing mental objects. Recent evidence is consistent with the realization “that the neural substrate of feeling states is to be found first subcortically and then secondarily repeated at cortical level” (Damasio et al., 2013: 833).

The triune brain model helps us visualize and understand the brain in terms of evolutionary layers that have evolved to solve different tasks (Panksepp and Biven, 2012; Linden, 2007). Basically, it resembles “three discernable scoops of ice cream” (Linden, 2007), which represent successive *add-ons* that the mindless forces of evolution have come up with because they happened to be useful. By understanding the brain in this layered way, we can also conceptualize that it must work in a bottom-up fashion (Panksepp, 2012; 1998; Solms and Turnbull, 2002), since newer additions are built on older ones.

The concept of the triune brain makes it easier to understand how consciousness works as well, most notably how it is extended from the deeper reaches of the brain to higher reaches, probably as a nested hierarchy (Feinberg, 2000; more details in Ch. 5.2) where higher forms of consciousness is embedded in deeper affective layers of consciousness (Solms and Panksepp, 2012). Thus, affect appears to be a fundamental characteristic of human consciousness (Solms and Turnbull, 2002; Solms and Nersessian, 1999), and the concept of the triune brain aids us in constructing more sophisticated models of human consciousness (Panksepp, 2005).

4.2.4 The primacy of affect

The philosopher, Bergson (1935) suggested that some of emotion precedes cognition. Summarizing this position, he maintains that “alongside the emotion which is a result of representation and which is added to it, there is the emotion which precedes the image, which virtually contains it, and is to a certain extent its cause” (Bergson, 1935: 35). This comes close to anticipating the idea that emotions should be seen in terms of a nested hierarchy, as discussed in the previous section. Although Bergson resists associating his point of view with sentimentalism, this position also hints at an affective reading of primary emotions.

In psychological research, Kunst-Wilson and Zajonc (1980) used stimuli masking techniques to show that emotional responses and specific preferences could be induced by subliminal stimuli that respondents do not consciously perceive. First, respondents were shown a sequence of faces unknown to them, masked behind lines so that they were not aware of seeing them. When subsequently showing a face that the respondents had been exposed to alongside a new unknown face, the respondents were not able to tell which face they had seen before any better than the rate expected by chance. However, when asked which face they preferred, respondents were significantly inclined to select the face they had just seen. The interpretation is that the brain will register faces that are masked without the image of these faces entering into conscious awareness because the stimuli are not consciously perceived (Breitmeyer and Ogmen, 2007). Hence, it seems that people can respond emotionally to stimuli without involvement of cognition. This echoes what we have already seen (Ch. 4.2.1) that patients without cognitive-cortical brain structures are capable of.

With the development of neuroscience, the study conducted by Kunst-Wilson and Zajonc (1980) has been corroborated by concrete evidence concerning the underlying

processes (Frith, 2007). By using brain scanners and various imaging techniques, one can observe that activity in specific affective centers of the brain takes place as a response to subliminal stimuli, although fMRI images need to be interpreted with caution (Feinstein, 2013: 305). The emotional responses generally happen very quickly and are faster than comparatively sluggish cognitive processes would allow; for instance, “it takes less than a hundredth of a second for a fear-potentiated startle reflex to be initiated” (Panksepp, 1998: 33).

Whalen et al. (1998) presented respondents with either a fearful or a happy unknown face immediately followed by a neutral face. The respondents would, on either occasion, report that they had seen the neutral face. However, only in cases where they had subconsciously also seen fearful faces would activity be triggered in the (fear-related) amygdalae of the sub-cortical brain. This and related studies suggest that stimuli must be sufficiently *strong* and also receive directed *attention* (Kouider and Dehaene, 2007). The strength of stimuli can be obstructed by masking techniques, and attention can be obstructed by presenting other stimuli.

In order to understand the interaction between cognition and affective emotions, cognitive neuroscience, as well as cognitive psychology, are clearly important. Behaviorally, the most relevant line of evidence is likely to be the study of instinctual emotional behaviors, which is best done in animals that do not regulate or inhibit behavioral outputs as much as humans (Panksepp, 2005: 159).

The importance of a separate scientific domain that concentrates on *the affective* is particularly evident in relation to *emotions*. Cognitive neuroscience, insofar as it deals only with the cognitive, cannot account for emotions that are qua affective and originate in the affective parts of the brain. Cognitive perspectives that attribute emotions to cognitive parts of

the brain are unable to ground emotions because they fail to account for their affective nature and genesis.

4.2.5 The broader theoretical basis

The emotion location argument rests primarily on empirical documentation provided by affective neuroscience, as presented above. More detail is provided below. The picture that emerges from findings in neuroscience is increasingly appreciated in interdisciplinary research areas (e.g., Northoff, 2011; Prinz, 2004) and in the psychology of emotions literature (e.g., Toronchuk and Ellis, 2012; Izard, 2011; Izard, 1992).

In neuroscience, *affective neuroscience* is specifically dedicated to the study emotions (Panksepp and Biven, 2012; LeDoux, 2000; Buck, 1999; Panksepp, 1998; Davidson and Sutton, 1995; Panksepp, 1992: 554-560). As Panksepp (2005: 159) states: “Affective neuroscience seeks to reveal the causal foundations of basic human emotional feeling through a neuroscientific study of emotional operating systems.” This study encompasses several bodies of knowledge about the mammalian brain; neuroanatomical (Panksepp, 1998: Ch. 4), neurophysiological (Panksepp, 1998: Ch. 5), and neurochemical facts (Panksepp, 1998: Ch. 6), as well as facts and postulations regarding how these neurological correlates correspond to psychological facts about the mind.

In the psychology literature, *affect theory* coheres relatively well with contemporary affective neuroscience (Kernberg, 2006); many of the same general ideas can be found even in relatively early accounts of affect theory (e.g., Izard, 1971; Tomkins, 1962; 1963). For instance, a central idea of Tomkins was that emotions represented the primary motivational system of human beings (Izard, 1991: 1), a point now supported by findings in affective neuroscience (Berridge and Kringelbach, 2008; Panksepp, 2003; Panksepp, 1998).

Furthermore, Tomkins's affect theory (1962; 1963) suggested that there were six *biologically based affects*: (1) "interest-excitement," (2) "enjoyment-joy," (3) "surprise-startle," (4) "distress-anguish," (5) "anger-rage," and (6) "fear-terror." This list fits remarkably well with the seven emotional systems that Panksepp and others identified later. Tomkins's "interest-excitement," for example, relates to Panksepp's (1998) "the seeking system"; "enjoyment-joy" to "play/social joy"; "distress-anguish" to "panic/grief"; "anger-rage" clearly to "rage/anger"; and "fear-terror" to "fear/anxiety." Only Tomkins's "surprise-startle" remains difficult to place within Panksepp's overall structure as of present, but other prominent neuroscientists (e.g., Damasio, 2003: 44) propose to include "surprise" as a primary emotion.

The emotion "surprise," along with other candidate primary emotions, constitutes a subject of current interest in neuroscience (e.g., Mellers et al., 2013), psychology (e.g., Toronchuk and Ellis, 2012), and philosophy (e.g., Prinz, 2004). Other possible primary emotions for which neural substrates might be found include "social dominance," "disgust," "guilt," and "shame." Some of these appear not to meet Panksepp's criteria for inclusion (presented in Ch. 2). For example, "guilt" and "shame" seem to require cognition and are thereby better classified as "secondary emotions" (Damasio, 2003: 45) or "tertiary emotions" (Panksepp, 1998). Prinz (2004: 163) notes that "surprise" does not seem to be valenced like other emotions, which puts its classification as an emotion in question. However, there is still an open discussion regarding what the complete set of primary emotions might look like, since the range of emotions in the human registry is not yet fully mapped.

As my focus in this chapter is on what *grounds* emotions, I am principally interested in the primary emotions. This concentration in no way suggests that secondary and tertiary emotions are unimportant. Cognitive processes play vital roles in regulating, shaping, complementing, and using primary emotions, or what we may call "raw affect" (Panksepp,

1998). Cognition allows us to cope in a variety of sophisticated ways with our external environment. One characteristic is that it transforms and elaborates our mental representations of pleasing and displeasing events (Frijda, 2007; Higgins, 2006). It also allows many distinctive human activities: self-reflection, sophisticated use of language, planning in time and space, mapping, and deliberate decision-making (Damasio, 2010; Solms and Turnbull, 2002). I will now turn to look at the sub-cortical location of affective emotions in the old mammalian brain. I will provide a non-technical overview; the technical details can be found in the cited literature.

4.3 Evidence for specific emotional systems

The primary-process emotion (or affect) systems are neural circuits that function as “hard-wired” responses, which are endowed by evolutionary processes to cope with our surroundings (Levenson, 2011: 379; Panksepp, 1998). According to Panksepp (2012:9), “Primary-process affects are the intrinsic barometers by which animal brains automatically gauge survival issues: If it feels good in a certain way, continue doing whatever you were doing. If it feels good in a different way, continue that type of behavior. If it feels bad, cease and desist from continuing the ongoing behavior, and shift to escape, hiding, and avoidance modes.”

“Circuits” and “systems” are general-purpose concepts for understanding how the brain processes, including emotions, work. These concepts correspond with the underlying physiological reality of the brain. Without going into technical detail regarding how neural circuitry and systems are constructed and how they function, an overall conceptual clarification is in order: “A *circuit* is a group of neurons that are linked together by synaptic connections. A system is a complex circuit that performs some specific function, like seeing or hearing, or detecting and responding to danger” (LeDoux, 2002: 49).

Following the dual-aspect monism perspective discussed in the previous chapter (Ch. 3), neural circuits have two aspects; they can be *observed* from the outside by science (the external perspective), and they can be *experienced* by a subject, an agent, or a decision-maker (the internal perspective). While the internal perspective of experiences cannot be directly accessed by science, we can make conjectures (Damasio et al., 2013; Pankepp, 1998) since affective emotional experiences will always be experienced as rewarding (positive) or punishing (negative). This characteristic makes reasonable inferences possible by observing patterns of behavior in response to stimuli (Panksepp and Biven, 2012).

Evidence suggests that emotional neural circuits are intrinsically affective (Solms and Panksepp, 2012; Solms and Turnbull, 2002; Panksepp, 1998). Affect itself cannot be learned but is instead hardwired into the brain at an early stage of fetal development (Panksepp, 2010; Solms and Turnbull, 2002). Each of the affective primary emotions, or “raw affects,” has their own distinct experiential registry or “flavor” (Rietti, 2009: 41). Furthermore, they constitute building blocks for other, more complex cognitive-affective emotional experiences (e.g., shame, guilt, pride, gratitude).

Emotion and cognition are “separate but interacting mental functions mediated by separate but interacting brain systems” (LeDoux, 1996: 69). Accordingly, current neuroscience suggests (Frith, 2007) that separate brain processes are involved in the *perceptual representation* of an object (cognitive processes) and the *evaluation* of its significance (emotional processes). While LeDoux (1996) views “emotion” as mechanical and automatic, he does not see it as affective. However, the view of emotions as affectively *experienced* is strongly supported by evidence.

Emotions are unlikely to be mere mechanical automata; instead, they are likely to be *felt* because they correlate closely with reported feelings in humans as well as with apparently

motivated behavior in other animals. Evidence therefore suggests that the parts of the brain from which *affect* emanates are subcortical (Damasio et al., 2013; Solms and Panksepp, 2012; Merker, 2007; Denton, 2006; Watt, 2005; Winkielman and Berridge, 2004: 123; Panksepp, 2001; Panksepp, 1998: Ch. 4). Moreover, it has been argued that consciousness itself is likely to be affective (Solms and Panksepp, 2012; Merker, 2007; Panksepp, 2005; Solms and Nersessian, 1999). From this perspective, the subcortical layer of the brain can appropriately be referred to as “the affective brain” because it is the seat of conscious subjective experience. Ruling out subjectivity because it cannot be directly penetrated by science (e.g., LeDoux, 1996) does not do justice to the reality of mental experiences (Panksepp, 1998); rather than explaining them, it explains them away instead.

There is a robust empirical case for locating emotions in the affective systems of the brain (Watt, 2005), and it is quite well documented which systems produce which kinds of emotional responses (Panksepp, 1998), although this ongoing mapping of the brain is still incomplete. Stimulating specific areas of the affective brain with low-current electrodes will consistently produce discrete emotional expressions in mammals, with a reliable valence (positive or negative), as inferred from behavior (Panksepp, 1998); for example, laboratory mice will eagerly self-stimulate to produce positive emotions. Similar effects are reliably produced in humans; patients suffering from depression that is chronic and resistant to medical treatment are cured with an impressive rate of success by deep brain stimulation of specific limbic and para-limbic areas of the brain (Holtzheimer and Mayberg, 2012). This method has recently become an approved clinical treatment (Ressler and Mayberg, 2007).

Panksepp and associates (Panksepp and Biven, 2012; Panksepp and Watt, 2003; Panksepp, 1998) have so far identified and located seven distinct emotional centers in the subcortical brain, which correspond to the following *primary emotions*: (1) seeking (expectancy, desire), (2) fear (anxiety), (3) rage (anger), (4) lust (sexual excitement), (5) care

(nurturance, bonding), (6) panic/grief (sadness), and (7) play (social joy). I present each emotional system except the “lust system” below due to its limited relevance in the present context. The exposition is mainly based on Panksepp and Biven (2012) and Panksepp (1998), and the primary emotion represented by each system meets the classification requirements set forth by Panksepp (as presented in Ch. 2).

4.3.1 The seeking system

The seeking system (sometimes referred to as “the expectancy system”) is a general-purpose emotional system that underpins all the other emotional systems (Panksepp and Biven, 2012; Panksepp, 1998: Ch. 8). Its effect is to make us seek out opportunities and to motivate action. In this sense, it can be seen as the *conation* system, providing the biological basis for Spinoza’s classical concept of *conata*. This system works by making us feel *enthusiastic* and *energized* (Panksepp, 1998). This basic system has extensively influences attention, incentive salience, associative learning, as well as anticipatory predictions (Alcaro, et al., 2007: 283).

Pleasure, or good feelings, generated from the arousal of our seeking system appear to be strictly *anticipatory*. Pleasure arises when perceptual information is used to motivate behaviors conducive to obtaining what we want or need (Levin, 2009: 193). While the “seeking system can motivate animals to pursue a diversity of distinct rewards in the environment” (Panksepp, 1998: 166), the system itself does *not* appear to be teleologically triggered *by* pleasure. Instead, its activation will motivate “blind” exploration and seeking behavior (Panksepp, 1998: 145).

Evidence suggests that activation of the seeking system is generally a *positive* experience (Panksepp and Biven, 2012) although it can generate obsessive behavior and hence be exhausting as well. Behaviorally, animals will almost always be eager to switch this system on, given the chance. Similarly, humans almost invariably report the stimulation of the

seeking system as a positive experience (Panksepp, 1998). The conscious feeling of being “energized” thus seems to represent a fundamental *positive emotion*. On the flip side, over-arousal of the system can have undesirable effects, such as addiction and addictive behavior (Panksepp, 1998: 161), associated with a craving for something but without necessarily deriving any experience of satisfaction from it (Robinson and Berridge, 2003).

Evidence indicates that *the seeking system* is the most central emotional system in the mammalian brain, including that of humans (Watt, 2005: 186; Panksepp, 1998: 145). Sometimes this emotional system is referred to as “the reward system” in the literature (Panksepp, 1998: 53), but this label is a misnomer since (1) there are *several* other “reward systems” in the brain that it is neurochemically more proper to conceive of as rewarding, and (2) *reward*, as indicated above, is not the most crucial characteristic of the seeking system (Panksepp and Biven, 2012); its nature is *appetitive*, not consumptive.

The underpinning of the seeking system is critical for human emotion; without it, “there is no ability to instantiate *any* coherent emotional state, and in a direct sense all the prototype states simply disappear, along with virtually any motivated behavior” (Watt, 2005: 186). In this way, it functions as a platform for all the other known emotional systems in the human brain. Thus, seeking appears to be the most fundamental drive in the human psyche.

The seeking system is innate and not dependent on higher brain structures (Levin, 2009: 194; Panksepp, 1998). It is located deep in the evolutionary old hypothalamic and affiliated mid-brain structures (Watt, 2005; the Lateral Hypothalamus, Nucleus Accumbens, VTA, and PAG; for technical details see Panksepp, 1998: 140–146). Neurochemically it is driven by the neurotransmitter, dopamine, and it is closely associated with the mesolimbic dopaminergic system (Alcaro et al., 2007).

4.3.2 The fear system

Fear is the most extensively researched emotion to date. It is hypothesized to be one of the oldest emotions to come out of evolution. The location of the fear system is deep in the subcortical brain, involving principally the *periaqueductal gray* (PAG) and the *amygdalae*, but also parts of the *hypothalamic structures*; details are available in Panksepp (1998). The fear system is an energy mobilizing system that produces upright freezing postures typically followed by fleeing or immediate withdrawal responses (Panksepp and Biven, 2012; LeDoux, 2002; Panksepp, 1998; LeDoux, 1996). The fear, system, however, does not always involve fear in response to detected objects in the external world; there are also cases of objectless fearfulness as seen in many anxiety disorders (Panksepp and Biven, 2012).

Humans are essentially born “objectless” and gradually come to learn about the external world and represent it as objects. Fear emotions can be attached to these objects in a learning process driven by experience (Panksepp and Biven, 2012), but the capacity for fear is inborn and not learned (Panksepp, 1998). Some fear responses are “hard-wired” and work without any prior learning; examples are freezing when we see a snake or moving away from places that are so steep that they instigate a feeling of fear. Fear is always accompanied by arousal of the autonomic nervous system (Panksepp and Biven, 2012); we may, for example, break out in a cold sweat and become pale.

The fear system is crucial in order to avoid direct harm. It can be over-sensitized, however, and become a problem. Mild fearfulness is likely to be a central input to enable prudent and risk-averse decision-making but can lead to excessive risk-aversion. It can also to some extent be controlled by learning processes (secondary processes; conditioned responses) and by more complex cognitive processes (tertiary processes) (Panksepp and Biven, 2012).

4.3.3 The rage system

While “aggression” (or “anger,” or “rage”) are traditionally associated with inflicting harm and physical violence, “a comprehensive definition of aggression should also encompass more subtle manifestations of effortful, goal-directed behaviors, including proactive aggression (“assertiveness”) and competitiveness, as well as predatory aggression and defensive/offensive aggression” (Panksepp and Zellner, 2004: 41). The rage system encompasses a relatively wide specter of emotional experiences and associated behavioral expressions.

The rage system is important for addressing perceived threats in the external environment (i.e., context, or situation). In evolutionary terms, this system is relatively old and is likely to have arisen out of a need to restrain the fear system (Panksepp, 1998), allowing for somewhat more adaptive sophistication (“fight” or “flight,” rather than just “flight”) in response to cues in the environment. The rage system prompts responses in several gradations, ranging from “mild irritation,” through prototypical “anger,” to “hot aggressive behavior” (Panksepp, 1998). Its location is deeply subcortical in the proximity of the fear system and includes the medial *amygdalae*, the *perifornical hypothalamus*, and the *dorsal PAG* (for details: Panksepp and Zellner, 2004; Panksepp, 1998). Anatomically and chemically, the rage system appears to be shared by humans and reptiles (Panksepp, 1998).

Panksepp and Zellner (2004) suggest that aggression comes in at least two distinguishable forms, “predatory aggression” and “affective attack,” which can be traced down to different, albeit somewhat overlapping, neurological underlying emotional circuitry. Predatory aggression appears to emanate from the seeking system (discussed above), whereas “affective attack” more specifically emerges from the rage system. Predatory aggression appears to be more “controlled,” “proactive,” and “instrumental,” whereas the affective attack

type of aggression is more “impulsive,” “reactive,” and “hostile” (Panksepp and Zellner, 2004: 47).

If we turn to applied ethics, emotion “aggression” appears to play a special and idealized role in some professions but not others. For example, in psychiatry, aggression is normally considered a sign of malfunction in the regulation of the stress and emotion circuitries. Curiously, “aggression” is often portrayed as a *virtue* in the business community to the point that companies enlist “aggression” as a “core value” (i.e., as an “ideal of excellence” or “virtue”; Ditlev-Simonsen and Wenstøp, 2011). Similarly, sales agents are being encouraged to pursue “aggressive sales strategies,” marketing people to devise “aggressive marketing campaigns,” and strategic decision-makers to make “aggressive decisions” and set “aggressive goals.”

Such examples of idealized “aggressive” mind states and their associated “aggressive” actions and rhetoric go hand in hand with the ideals of competition and the pursuit of profit. Porter (1997), for example, discusses “competitive strategy” in terms of choosing between “defensive” or “aggressive” postures. This approach fits with the neurological description of predatory forms of aggression discussed above. However, consciously sought aggression at the decision-making level is strongly modified and restrained by cognitive processes (Panksepp and Zellner, 2004), as well as action-filtering processes in mid-brain structures (*basal ganglia*) between the deep subcortical structures from which aggression arises and the cortical decision-making units of the brain (as elaborated on in Ch. 6).

Cultural and contextual factors, in turn, strongly shape much of what happens concerning cortically initiated regulation, suppression, and modification of aggression (Panksepp and Zellner, 2004). Thus, the cultural idealization of aggression has clear moral

implications since normative emphasis on one emotion is generally at the expense of emphasizing *other* moral emotions.

In contrast to other animals, and in virtue of our enhanced cognitive capabilities, “it appears that some human predators do wish to inflict pain or fear on their victims, and moreover that they nurture aggressive intentions over long periods of time, which would indicate some kind of extended, cognitively-mediated aggressive arousal” (Panksepp and Zellner, 2004: 54). Since the activities of the anger system and seeking system mostly take place beneath the level of cognitive awareness (Panksepp, 1998), fostering intentional aggressiveness may unwittingly contribute to socially insensitive and destructive patterns of behavior.

4.3.4 The care system

Care is an important moral emotion and therefore deserves emphasis in this exposition. The care system is the central basis for *human bonding* and *social relations*. The human care system appears to be an emotional infrastructure that is specific to mammalian species. Certain other species also exhibit care, such as birds, but their neural circuitry appears to be a different construction than the care system in mammals (Panksepp, 1998: 259).

Young et al. (2005) traced the formation of social relations to specific neural circuits and identified the subcortical nuclei where they were localized. The caring system involves several subcortical clusters, including the *anterior cingulate*, *hypothalamic nuclei*, the *ventral PAG*, and the *ventral tegmental area*; for details, see Panksepp, 1998: 249–253. Churchland (2011) argues that caring for offspring is common to all mammals and is the root of *essentially all moral values*. From an evolutionary perspective, caring was initially limited to parenting but has later evolved into a morality of increasing social scope. The basic evolutionary function of the care systems is to make parents, and especially mothers, attach to

their offspring and devote themselves to ensuring their well-being and happiness (Panksepp and Biven, 2012). The care system is also involved in *altruism, empathy, expressions of love, and social tolerance* (Panksepp, 1998).

Hormonal neurotransmitters including *oxytocin* put the brains into caring modus or a “maternal state of mind.” While some argue that this state of mind involves a generally higher capacity for caring (e.g., Churchland, 2011), others have maintained that the effects of the caring system significantly shifts the *distribution* of caring attention so that a person comes to devote *more* focused care *within* caring relations but at the same time *less outside* these relations (Panksepp and Biven, 2012).

In addition to being in the focus of neurochemical research, oxytocin has attracted much public attention recently. The experimental economist Zak (2012) argues that it is a “moral molecule,” which accounts for all forms of altruism, bonding, and trust. Recent animal research (Guzmán et al., 2013), however, suggests that the effects of oxytocin are highly context dependent. In addition to promoting good emotional experiences in positive-social and intimate contexts, oxytocin appears to induce increased fear and anxiety in negatively charged contexts. It remains a central and highly specialized neurotransmitter mediating the care system, but it is far from the only one (Panksepp, 1998).

4.3.5 The panic/grief system

The panic/grief system is a pan-mammalian system that is relatively recent evolutionarily. Its workings and psychological effects are moderately complex. Stemming from of a need to respond to care (i.e., the care system), it involves not just negative emotions like grief, distress, and sadness but also positive emotions related to attachment (Panksepp, 1998). This system plays a major role in sadness, loneliness, depression, and grief (Panksepp, 1998) and is

located in the higher subcortical regions, including the *PAG* and the *dorso-medial diencephalon*; again, see Panksepp, 1998, for details.

The care system and the panic/grief system appear to work in tandem in sophisticated cognitive-affective ways. Taylor (2002) suggests that this tandem system in some situations can represent a third way of responding to threats or stress factors in the external environment. Instead of simply opting for the more masculine fight or flight responses, the panic/grief system in combination with the care system can induce more feminine social approaches, such as appeasement or befriending strategies (Taylor, 2002).

4.3.6 The play system

Play is an unambiguously positively valenced emotion (Panksepp, 1998; Panksepp et al., 1984). The play system, located subcortically and involving the *PAG* and the *dorso-medial diencephalon*, is still being explored as an emotional system (Panksepp and Biven, 2012). It involves an emotional urge for laughter and playful social interaction, followed by a capacity for enjoyment (Burgdorf and Panksepp, 2005).

The play system is probably important for prompting us to explore different solutions to problems, thus giving higher cognitive functions purpose and setting them into motion. It is crucial for individuals to explore the *social context* with all its cultural aspects, social norms, and inter-subjectively dependent social arrangements and understandings; “The play system allows children to learn about social rules of conduct” (Panksepp and Biven, 2012: Loc. 271). The same type of awareness is essential in social and moral decision-making since breaking social norms can sometimes have disastrous results. At other times, the reverse is true; breaking social norms (e.g., setting a moral example by counteracting social norms, such as homophobia) can at times bring about dramatically better social outcomes. Play fine tunes

our ability to anticipate results and the *expected value* of actions, which I will return to in Ch. 6.

4.4 Conclusion

There is strong scientific evidence for locating the origins of emotions in specific areas of the subcortical brain. Thus, Anscombe's challenge (Ch. 1) can be met, not principally by psychological enquiry as envisioned but instead with information from neuroscience that underpins psychology. Current evidence renders it plausible that the sources of normativity, which correspond to the emergence of affective emotions, are located sub-cortically in the upper brain stem and mid-brain structures.

Considerable detail is known from neuroanatomy, neurophysiology, and neurochemistry about the location and workings of specific emotional neural circuits, although many questions remain. Currently, seven distinct emotional neural circuits in humans and other mammals are identified and specified to a relatively high degree of detail. Each emotional center corresponds to scientifically well-documented expressive-behavioral patterns in humans and other mammals and can be induced by means of brain stimulation in specified sub-cortical areas.

It has been robustly demonstrated by contemporary affective neuroscience that the cortical brain structures in humans as well as other mammals are not necessary for the operation of primary emotions or for their expression and behavioral outputs. There seems to be no good theoretical reason, moreover, to suggest that subjects who have these emotions do not experience them subjectively. Evidence, moreover, suggests that basic consciousness and experience is generated subcortically.

Chapter 4 References

- de Aguirre, M. I. (2006) Neurobiological bases of aggression, violence, and cruelty. *Behavioral and Brain Sciences* **29** (3): 228–229.
- Alcaro, A., Huber, R., and Panksepp, J. (2007) Behavioral functions of the mesolimbic dopaminergic system: An affective neuroethological perspective. *Brain Research Reviews* **56**: 283–321.
- Bergson, H. (1977) *The Two Sources of Morality and Religion*. (R. A. Audra and C. Brereton, Trans.). Paris: University of Notre Dame Press. (Original work published in 1935).
- Berridge, K. C. and Kringelbach, M. L. (2013) Neuroscience of affect: Brain mechanisms of pleasure and displeasure. *Current Opinion in Neurobiology* **23**: 294–303.
- Berridge, K. C. and Kringelbach, M. L. (2008) Affective neuroscience of pleasure: Reward in humans and animals. *Psychopharmacology* **199**: 457–480.
- Blackburn, S. (1998) *Ruling passions: A theory of practical reasoning*, Oxford: Oxford University Press.
- Bower, (2007) Consciousness in the raw: The brainstem may orchestrate the basics of awareness. *Science News* **172** (11): 170–172.
- Breitmeyer, B. G. and Ogmen, H. (2007) Visual masking. *Scholarpedia* **2** (7): 3330.
- Buck, R. (1999). The biological affects: A review. *Psychological Review* **106** (2): 301–336.
- Burgdorf, J. and Panksepp, J. (2005) The neurobiology of positive emotions. *Neuroscience and Biobehavioral Reviews* **30**: 173–187.

- Charland, L. C. (2005) The heat of emotion: Valence and the demarcation problem. *Journal of Consciousness Studies* **12** (8–10): 82–102.
- Chomsky, N. (1986) *Knowledge of language: Its nature, origins, and use*. Westport: Praeger Publishers.
- Churchland, P. S. (2011) *Brainthrust*. Princeton: Princeton University Press.
- Churchland, P. S. (2002) *Brain-wise: Studies in neurophilosophy*. Boston: MIT Press.
- Damasio, A., Damasio, H., and Tranel, D. (2013) Persistence of feelings and sentience after bilateral damage of the insula. *Cerebral Cortex* **23** (4): 833–846.
- Damasio, A. (2010) *Self comes to mind: Constructing the conscious brain*. New York: Pantheon Books.
- Damasio, A. (2003) *Looking for Spinoza: Joy, sorrow, and the feeling brain*. New York: Harcourt.
- Damasio, A. (1994) *Descartes's error: Emotion, reason, and the human brain*. New York: Avon Books.
- Darwin, C. (1906) *The descent of man and selection in relation to sex* (2nd ed.). London: John Murray.
- Datta, P. M. (2005) Earliest mammal with transversely expanded upper molar from the Late Triassic (Carnian) Tiki Formation, South Rewa Gondwana Basin, India. *Journal of Vertebrate Paleontology* **25** (1): 200–207.
- Davison, R. J. and Sutton, S. K. (1995) Affective neuroscience: The emergence of a discipline. *Current Opinion in Neurobiology* **5**: 217–224.

- de Waal, F. B. M. (1996) *Good natured: The origins of right and wrong in humans and other animals*. Cambridge: Harvard University Press.
- Denton D. E. (2006) *The primordial emotions*. New York: Oxford University Press.
- Ditlev-Simonsen, C. and Wenstøp, S. (2011) Companies' ethical commitment: An analysis of the rhetoric in CSR reports. *Issues in Social and Environmental Accounting* **1-2**: 65–81.
- Ekman, P. (1992) An argument for basic emotions. *Cognition and Emotion* **6** (34): 169–200.
- Elm, D. R. and Radin, T. J. (2012). Ethical decision making: Special or no different? *Journal of Business Ethics* **107**: 313–329.
- Farber, P. L. (1994) *The temptations of evolutionary ethics*. Berkeley: The Regents of the University of California.
- Feinstein, J. S. (2013) Lesion studies of human emotion and feeling. *Current Opinion in Neurobiology* **23**: 304–309.
- Feinberg, T. (2000) The nested hierarchy of consciousness: A neurobiological solution to the problem of mental unity. *Neurocase* **6**: 75–81.
- Frith, C. (2007) *Making up the mind: How the brain creates our mental world*. Oxford: Blackwell Publishers.
- Frijda, N. H. (2007) *The laws of emotion*. London: Lawrence Erlbaum Associates.
- Green, O. H. (1992) *The emotions: A philosophical theory*. London: Kluwer Academic Publishers.

- Guzmán, Y. F., Tronson, N. C., Jovasevic, V., Sato, K., Guedea, A. L., Mizukami, H., Nishimori, K. and Radulovic, J. (2013) Fear-enhancing effects of septal oxytocin receptors. *Nature Neuroscience* **16**: 1185–1187.
- Hägerström, A. (1917/1953) *Inquiries into the nature of law and morals*. (K. Olivecrona, Ed, C. D. Broad, Trans,). Uppsala: Almqvist & Wiksells.
- Higgins, E. T. (2006) Value from hedonic experience and engagement. *Psychological Review* **113**: 439–460.
- Holtzheimer, P. E. and Mayberg, H. S. (2012) Neuromodulation for treatment-resistant depression. *F1000 Medicine Reports* **4**: 22.
- Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion-cognition interactions. *Emotion Review* **3**: 371–378.
- Izard, C. E. (1992) Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review* **99**, 561–565.
- Izard, C. E. (1991) *The psychology of emotions*. New York: Plenum Press.
- Izard, C. E. (1971). *The face of emotion* (Vol. 23). New York: Appleton-Century-Crofts.
- Kernberg, O. F. (2006) Psychoanalytic affect theory in the light of contemporary neurobiological findings. *International Congress Series* **1286**: 106–117.
- Kouider, S. and Dehaene, S. (2007) Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362** (1481): 857–875.
- Kunst-Wilson, W. R. and Zajonc, R. B. (1980) Affective discrimination of stimuli that cannot be recognized. *Science* **207** (4430): 557–558.

- Lawton, R., Conner, M. and McEachan, R. (2009) Desire or reason: Predicting health behaviors from affective and cognitive attitudes. *Health Psychology* **28** (1): 56–65.
- LeDoux, J. (1996) *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon & Schuster.
- LeDoux, J. (2000) Emotion circuits in the brain. *Annual Review of Neuroscience* **23**:155–184.
- LeDoux, J. (2002) *Synaptic self: How our brains become who we are*. London: Penguin Books.
- Levin, F. M. (2009) *Emotion and psychodynamics of the cerebellum: A neuro-psychoanalytic analysis and synthesis*. London: Karnac Books.
- Linden, D. J. (2007) *The accidental mind: How brain evolution has given us love, memory, dreams, and god*. Boston: Harvard University Press.
- Merker, B. (2007) Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences* **30** (1): 63–81.
- MacIntyre, A. (1999) *Dependent rational animals*. London: Duckworth.
- MacLean, P. D. (1990) *The triune brain in evolution: Role in paleocerebral functions*. New York: Plenum Press.
- MacLean, P. D. (1949) Psychosomatic disease and the “visceral brain”: Recent developments bearing on the Papez theory of emotion. *Psychosomatic Medicine* **11**: 338–353.
- Mellers, B., Fincher, K., Drummond, C. and Bigony, M. (2013) Surprise: A belief or an emotion? *Progress in Brain Research* **202**: 3–19.

- Midgley, M. (1978) *Beast and man: The roots of human nature*. New York: Cornell University Press.
- Narvaez, D. (2008) Triune ethics: The neurobiological roots of our multiple moralities. *New Ideas in Psychology* **26**: 95-119.
- Nesse, R. (1990) Evolutionary explanations of emotions. *Human Nature* **1**: 261–289.
- Northoff, G. (2011) *Neuropsychanalysis in practice: Brain, self, and objects*, Oxford: Oxford University Press.
- Northoff, G. (2004) *Philosophy of the brain: The brain problem*. Amsterdam: John Benjamin's Publishing.
- Pally, R. (2000) How brain development is shaped. In Pally, R. and Olds, D. D., *The mind-brain relationship* (pp. 1-18). New York: Other Press.
- Panksepp, J. (2015) Toward the constitution of emotional feelings: Synergistic lessons from Izard's differential emotions theory and affective neuroscience. *Emotion Review* **7** (2): 110–115.
- Panksepp, J. and Biven, L. (2012) *The archeology of mind: Neuroevolutionary origins of human emotions*. New York: Norton.
- Panksepp, J., Asma, S., Curran, G., Gabriel, R., and Greif, T. (2012) The philosophical implications of affective neuroscience. *Journal of Consciousness Studies* **19** (3–4): 6–48.
- Panksepp, J. (2010) Affective consciousness in animals: Perspectives on dimensional and primary process emotion approaches. *Proceeding to the Royal Society B* **277** (1696): 2905–2907.

- Panksepp, J. (2005) On the embodied neural nature of core emotional affects. *Journal of Consciousness Studies* **12**: 158–184.
- Panksepp, J. and Zellner, M. R. (2004) Towards a neurobiologically based unified theory of aggression. *Revue Internationale de Psychologie Sociale* **2**: 37–62.
- Panksepp, J. (2003) At the interface of the affective, behavioral, and cognitive neurosciences: Decoding the emotional feelings of the brain. *Brain and Cognition* **52**: 4–14.
- Panksepp, J. and Watt, D. F. (2003) The ego is first and foremost a body ego. *Neuropsychoanalysis* **5**: 201–218.
- Panksepp, J. (2001) On the subcortical sources of basic human emotions and the primacy of emotional-affective (action-perception) processes in human consciousness. *Evolution and Cognition* **7** (2): 134–140.
- Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.
- Panksepp, J. (1992) A critical role for “affective neuroscience” in resolving what is basic about emotions. *Psychological Review* **99**: 554–560.
- Panksepp, J., Siviy, S. M., and Normansell, L. A. (1984) The psychobiology of play: Theoretical and methodological perspectives. *Neuroscience and Biobehavioral Reviews* **8**: 465–492.
- Ploog, D. W. (2003) The place of the triune brain in psychiatry. *Physiology and Behavior* **79** (3): 487-493.
- Porter, M. E. (1997) Competitive strategy. *Measuring Business Excellence* **1** (2): 12–17.
- Prinz, J. J. (2007) *The emotional construction of morals*. Oxford: Oxford University Press.

- Prinz, J. J. (2004) *Gut reactions: A perceptual theory of emotion*. Oxford: Oxford University Press.
- Ressler, K. J. and Mayberg, H. S. (2007) Targeting abnormal neural circuits in mood and anxiety disorders: From the laboratory to the clinic. *Nature Neuroscience* **10**: 1116–1124.
- Rietti, S. (2009) Rationalities of emotions: Defending, distinguishing, connecting. *Organon F* **16** (1): 38–61.
- Robinson, T. E. and Berridge, K. C. (2003) Addiction. *Annual Review of Psychology* **54**: 25–53.
- Rorty, A. (1978) Explaining emotions. *Journal of Philosophy* **75**: 139–161.
- Shewmon, D. A., Holmes, G. L., and Byrne, P. A. (1999) Consciousness in congenitally decorticate children: Developmental vegetative state as self-fulfilling prophecy. *Developmental Medicine and Child Neurology* **41**: 364–374.
- Shu, D-G., Luo, H-L., Morris, S. C., Zhang, X-L., Hu, S-X., Chen, L., Han, J., Zhu, M., Li, Y. and Chen, L-Z. (1999) Lower Cambrian vertebrates from south China. *Nature* **402**: 42–46.
- Singer, P. (1982) Ethics and sociobiology. *Philosophy and Public Affairs* **11** (1): 40–64.
- Solms, M. (2013) “The unconscious” in psychoanalysis and neuropsychology. In S. Akhtar and M. K. O’Neil (Eds.), *On Freud’s “The Unconscious.”* London: Karnac Books.
- Solms, M. and Nersessian, E. (1999) Freud’s theory of affect: Questions for neuroscience. *Neuro-psychoanalysis* **1** (1): 5–14.

- Solms, M. and Panksepp, J. (2012) The “id” knows more than the “ego” admits: Neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Sciences* **2**: 147–175.
- Solms, M. and Turnbull, O. (2002) *The brain and the inner world*. New York: Other Press.
- Solomon, R. C. (2007) *True to our feelings*. Oxford: Oxford University Press.
- Taylor, S. E. (2002) *The tending instinct: How nurturing is essential to who we are and how we live*. New York: Holt.
- Tomkins, S. S. (1962) *Affect imagery consciousness: Volume I, the positive affects*. London: Tavistock.
- Tomkins, S. S. (1963) *Affect imagery consciousness: Volume II, the negative affects*. London: Tavistock.
- Watt, D. F. (2005) Social bonds and the nature of empathy. *Journal of Consciousness Studies*, **12** (8–10): 185–209.
- Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., and Jenike, M. A. (1998) Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience* **18** (1): 411–418.
- Wilkinson, G. (1984) Reciprocal food sharing in the vampire bat. *Nature* **308** (5955): 181–184.
- Winkielman, P. and Berridge, C. K. (2004) Unconscious emotion. *Current Directions in Psychological Science* **13** (3): 120–123.
- Young, L. J., Young, A. Z. M., and Hammock, E. A. D. (2005) Autonomy and neurochemistry of the pair bond. *Journal of Comparative Neurology* **493**: 51–57.

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

Zak, P. J. (2012) *The moral molecule: How trust works*. New York: Dutton.

Zellner, M. R., Watt, D. F., Solm, M., and Panksepp, J. (2011) Affective neuroscientific and neuropsychanalytic approaches to two intractable psychiatric problems: Why depression feels so bad and what addicts really want. *Neuroscience and Biobehavioral Reviews* **35**: 2000–2008.

Chapter 5: The affective nature argument

5.1 Introduction to the affective nature argument

The primary concern of this chapter is to address *the nature problem* (Ch. 1.3.2). It argues that the nature of normativity, as a psychological phenomenon, is *affective*; hence the chapter title, “The affective nature argument.” The core of this argument is the position I label the *normativity-as-affect position* (see also Ch. 9.4.2). This position holds that normativity can be identified as a distinct aspect of subjective affective experience. Drawing on contemporary achievements of affective neuroscience, I argue that the basic “pleasure” and “unpleasure” of affective consciousness accounts for the *valence* of normativity and that specific affective emotions shape the *distinctiveness* of normative feelings: their “emotional coloration.”

The normativity-as-affect position suggests that normative deliberation is fundamentally based on emotions and affective consciousness in the sense that affect accounts for the essence of normativity and emotions play a key role within psychologically active bundles of affects. This realization acknowledges that affects constitute a broader set of mental phenomena than emotions (see Ch. 2). As a helpful metaphor, emotions can be thought of as the shifting of “weather,” whereas moods and other background affects can be thought of as a more long-term “climate.” Emotions are also, as noted, more distinctive and thematic than background affects.

In conjunction with the arguments from the forgoing chapters (Ch. 3 and Ch. 4), normativity-as-affect paves the way for a naturalistic, biologically based *causal explanation* of normativity but without claiming that the *nature* of the normative mind itself can be fully accounted for in naturalistic terms (Ch. 3). The psychological mind depends on neural activity in the brain (Ch. 3), and this is further specifiable in terms of physical emotional circuits and neuro-chemistries in the brain (Ch. 4).

Finally, this chapter draws attention to the affinity among three central concepts of the subjective mind: *normativity*, *affect*, and *consciousness*. Central themes of the present chapter are subjective experience, normativity-as-affect, and affective consciousness, and the concept of a nested hierarchy of emotions.

5.2 Psychology of affective emotions

5.2.1 Normativity as affective

In Chapter 1 I introduced the idea of normativity as a *psychological* notion; it involves a sense of “ought” or of “good” and “bad” attached to the mental outlook or *perspective* of a person. As a practically oriented attitude, normativity acts like an aroused “to-be-done-ness dimension” of our state of mind. Normativity can be viewed as an aspect or dimension of a particular type of mind state, and I argue that this is an affective state of mind that involves emotional feelings. As noted (Ch. 1), Gibbard (2012) argues that all of meaning should be seen as inherently normative and hence affective. Conversely, normativity can be thought of as rendering objects in the world subjectively meaningful. This is in line with Modell’s (2003: 15) evidence based view that “*meaning is embodied in our total affective interest in the world*”.

However, many philosophers dispute the contention that normativity is affective. The two major contending views about the nature of normativity in moral philosophy are the cognitive view and the affective view (Prinz, 2007: 13). The central issue that separates these philosophical perspectives is the question of whether normativity requires affect (the affective view) or not (the cognitive view), although there are also non-cognitive views that are not affectively based on notions such as “intuition” construed as a separate mental process. For the moment I will ignore intuitionism; the supposition that intuitions represent separate

neurological mechanisms is regarded as speculative. This omission leaves the affective view and the cognitive view of normativity as the main contenders within moral psychology.

By relating normativity to its underpinning psychological states of mind, the two main philosophical positions regarding normativity can be related to psychological and neuroscientific theory and evidence. Implicitly or explicitly, they involve claims about the nature of the mental processes of the brain and mind. Hence, they can be legitimately assessed in light of empirical evidence. I have already argued (Ch. 1) that normativity has to be seen as internal mental phenomena, if it is going to have a direct bearing on moral motivation and what we care about (Goldman, 2009). It better have such a bearing, because otherwise it would make normativity both irrelevant and inexplicable (which I discuss in more detail Ch. 7 and 8). So the question is whether the internal mental phenomena are correctly seen as affective or cognitive. The literature review in Chapter 2 suggested that they should be seen as affective.

The idea that affective emotions play a necessary role in normative perspective formation and normative deliberation seems pivotal, since this is what cognitivists characteristically deny. One divisive issue is that according to the affective view, normative perspectives cannot be turned on or off at will (Kahneman, 2011: Ch. 1), whereas cognitive views of normativity typically assume free will with regard to normative perspective taking (e.g., Solomon, 2007, Korsgaard, 1996).

Consider this simple example: Assume that you observe an action that you morally and unequivocally disapprove of, say, a salesperson cheating her innocent client. In this case, according to the affective view of free will, it would be impossible for you to choose to nevertheless approve of the action simply at will. You can try to rationalize the behavior of the salesperson, perhaps finding arguments such as that every salesperson cheats; therefore,

they have to cheat to stay competitive as they compete against each other under tough market conditions. However, rationalizations may or may not work efficaciously, so it is not as if you simply *elect* to convince yourself one way or another; it would at the very least involve *uncertainty* regarding the outcome of the reasoning as well as the cognitive *effort*.

In this example, as in many other everyday moral contexts, it seems difficult to separate normative experiences from affective experiences in general. It is also difficult to conceptualize a normative experience without also admitting the presence of *valence* of approach or avoidance – subjective experiences of pleasure and displeasure – provided by primary emotions (Panksepp, 1998) and basic consciousness generated by the reticular activation system in the brain stem (Solms and Turnbull, 2002: Ch. 3); argued (Ch. 2) evidence suggests that the nature of emotions is affective. Normative experiences, therefore, seem to be an organic part of our overall affective experiences, including the involvement of emotions, at any given time. This *inseparability* represents an argument for seeing normativity as affective.

The same difficulty of choice seems to hold true for aesthetic affects (Robinson, 2005). For example, if you hear a piece of music that you really like, it will be impossible for you to simply decide to dislike it. It also seems true for simpler cases of tonal appreciation. Our auditory system and mental apparatus are tuned so as to distinguish chords and discords automatically. This ability seems to provide a sense of harmony and disharmony that is closely analogous with our sense of morality, notably including our normative distinctions between what is appropriate and inappropriate behavior in a given context. However, this ability does not mean that affective attitudes, including normative attitudes, cannot change over time; they are not based upon or even open to momentary decisions.

Morality psychology and, more broadly, normativity, appears to behave along the lines we would expect if it were an affective experience. I suggest that it is. Normative attitudes share the positive or negative valence characteristic of emotions. Like many emotions, they attach to particular perceived or imagined objects in a way that is motivational. And like emotions, normative attitudes are sensitive to our roles and relations. Insofar as emotions are fundamentally affective (Ch. 2), these close analogies count in favor of viewing normativity as affective rather than cognitive.

5.2.2 The subjectivity of normativity

It is widely accepted across philosophy, psychology, psychiatry, and neuroscience that emotional feelings are subjective. My conceptualization of normativity as affective makes normativity subjective as well. From a psychological perspective, normativity is firmly on the side of the *subjective-mind*, rather than that of the objective-mind, in the perspectival division proposed by *dual-aspect monism* (see Ch. 3). While affect may be considered either from its internal subjective aspect (i.e., experienced affect) or from its external objective aspect (i.e., expressed affect), its fundamental aspect is its *internal* aspect. Without a subject *having* an affective experience, we could not appropriately conclude that what was expressed was in fact affect. For example, if a robot merely *appeared* emotional, we would not thereby say that it *really* was emotional, not even if it were a highly sophisticated robot. Therefore, there cannot truly be a science of the affective without its internal, subjective aspect in place; consequently, the same holds true for normativity.

Let me clarify the idea that psychological perspectives can only be experienced by a subject, as Solms and Turnbull (2002) point out. This statement means that psychological perspectives can never be directly accessed by scientific methods. At best, scientists can see different chemical balances and imbalances, which are indications of neuronal activity, such

as blood flows, stimuli inputs, behavioral-expressive outputs, and so on. All these components merely *correlate* with subjective experience. Scientists can identify some of the *structures* that give rise to a perspective but not the *content* of the perspective. Thus, the normative experiences of a person cannot be accessed by science, although science may, under favorable circumstances, be able to determine whether such a normative experience takes place.

From a scientific perspective, therefore, we cannot *qualitatively* understand what normative experiences are like; similar to other affective experiences, these internal states are only “observable” by the subject (Solms and Turnbull, 2002: 295). However, since normative experiences are instances of emotional affective experiences, science should be in a position to infer (i.e., from imaging techniques, electrical stimulation, chemical stimulation, and behavioral observation) some of the classification categories *quantitative* normative experiences fall into, whether they involve positive or negative valence (or ambivalence). Science does not preclude the objective description of subjective states, although it cannot penetrate them. An emotional expression of fear with corresponding activity in the neural fear circuitry (the amygdalae, etc.) would for example classify as negatively valenced and hence “bad” and highly “displeasing” as seen from *any* given internal perspective.

It is important to notice that the objective perspective of science that was just described does not *make* normativity objective. We have not bridged Hume’s famous gap between “is” and “ought.” Instead, the external, objective view of “badness” and “displeasure” *presupposes* an internal, subjective view, unless there is a view from within, there would be no *genuine* badness and displeasure to which the objective category referred. It is in this sense that the internal affective-normative dimension can be said to be fundamental.

All well-functioning human beings have normative perspectives. We can relate to these perspectives subjectively on several different levels (Solms and Panksepp, 2012): levels

that can be differentiated anatomically roughly along the lines specified by the concept of the triune brain (see Ch. 4). Each of these levels may provide important input for practical decision-making. Non-arbitrary decision-making presupposes the involvement of affects (Damasio, 2010), as I will discuss extensively in the next chapter.

In a basic unthinking manner, we have, according to contemporary theory in affective neuroscience (Solms and Panksepp, 2012), direct conscious access to our lower level affects. At higher levels where more sophisticated forms of cognition are involved (drawing on the prefrontal cortex, the association cortexes, etc.), we have normative views with concrete representations of objects, prospective outcomes, and actions being “good,” “bad,” “right,” “wrong,” “brave,” “cowardly,” “pleasing,” and “displeasing.”

Moreover, given the combination our empathetic abilities (Panksepp, 2011) and our own affective normative experiences, we can in many cases correctly understand the basic contours of the normative perspectives of others and respond accordingly. Some of the social understanding of others appears to arise in us automatically by virtue of a special class of neurons that are socially attentive at the subconscious level; this class of neurons is called “mirror-neurons” (Gallese, 2005) because they use sensory information to “mirror” or simulate what goes on inside the heads of other people.

Human brains are also affected by special neurochemical molecules (e.g., oxytocin), whose release is subconsciously activated by social cues in the external environment. By increasing the efficacy of relevant neuronal receptors, these molecules enhance social recognition and intensify social emotional feelings (Donaldson and Young, 2008).

Empathy, as commonly understood, is the ability to detect the emotional states of others (see Ch. 1). While this ability is well documented in humans as well as other mammals, it is still somewhat uncertain whether the underpinning of empathy is cognitive or affective

(Panksepp, 2011). Freedberg and Gallese (2007), however, present evidence suggesting that empathy is fundamentally affective and argue that the general assumption of cognitive science that empathy only involves cognitive abilities is mistaken. Gallese (2005) suggests that empathy and emotional expression together allow us, by means of subconscious embodied simulation of others, to “share” our subjective experiences. We gauge what is taking place in the mind of others through our sensory apparatus that sends information to a special type of neurons, “mirror neurons” (Gallese, 2005), which mimics the mental activity of others in our own brain. This ability suggests a biologically based explanation of a human capacity for socially reflexive normative inter-subjectivity.

5.2.3 Affective consciousness

Accumulating evidence from affective neuroscience suggests that consciousness should be modeled by hierarchically structured *layers* rather than being conflated to a single layer (Panksepp, 2005; Feinberg, 2000; Zahavi, 1999). This arrangement allows for thinking in terms of gradually higher forms of consciousness, culminating in higher cortical types of consciousness that are characteristically self-referential and of which we have awareness.

It also permits the conclusion that there are lower forms of phenomenal experience that humans have in common with other animals. According to Zahavi (1999), primitive, pre-reflective, self-consciousness simply corresponds to the subjective dimension of experience. These “lower” forms of experience are not “unconscious” because subjective experience presupposes consciousness (Mandler, 2002: Ch. 5) and reversely, “consciousness is unavoidably and necessarily subjective” (Northoff, 2013: 218). However, lower forms of consciousness do not necessarily enter the level conscious self-referential awareness. It seems evident that a number of other mammals are incapable of self-referential awareness, but it

would be implausible to assume that they thereby could not have any conscious subjective experience whatsoever (Panksepp, 2005).

As established by clinical evidence in humans (Solms and Turnbull, 2002; Panksepp, 1998) and by experimental research on mammals in possession of closely similar brain structures (Panksepp and Biven, 2012; Panksepp, 1998), the (upper) brainstem (the midbrain and the *pons*) and subcortical regions of the brain are necessary for consciousness (Merker, 2007; Panksepp, 2005). In the subcortical regions of the brain, therefore, we have a basic form of consciousness, which can be appropriately called *affective consciousness* (Solms and Panksepp, 2012; Panksepp, 2005). Philosophically, the idea of affective consciousness is not entirely new; for example, it was captured by Husserl's notion of *Gemütbewusstsein* (see also the excellent historical review of affective consciousness in Tallon, 1997: 147–153). Panksepp, however, arrives at this theoretical position by having to account for concrete empirical data.

Evidence showing that the neocortex is not necessary for human consciousness has already been presented (Ch. 4). The prerequisites for the operation of basic affective consciousness, as well as the full specter of identified primary human emotions, appear to be independent of neocortical activity. Furthermore, as previously noted (Ch. 4), observations suggest that even rudimentary forms of *learning* are possible in humans without cortical involvement (Merker, 2007). Hence, the cortical regions of the brain cannot be the seat of consciousness. Instead, the foundations of consciousness must be subcortical.

It also seems clear that from an evolutionary perspective, consciousness must have arisen out of the older subcortical parts of the brain (Panksepp and Biven, 2012; Damasio, 2010). As Damasio (2010: 29) notes, “whenever the brain begins to generate primordial feelings – and that could be quite early in evolutionary history – organisms acquire an early

form of sentience.” Regarding consciousness, therefore, affective neuroscience presents a picture that is strikingly different from what is traditionally assumed across philosophy, psychology, and neuroscience (Solms and Panksepp, 2012).

The presence of “affective consciousness” (Panksepp, 2005; Panksepp, 1998: 300) or “interoceptive consciousness” (Northoff, 2013) is strongly corroborated by evidence from lesion studies (Damasio, 2010; Solms and Turnbull, 2002; Panksepp, 1998). A subcortical lesion will compromise all forms of consciousness (as already shown by Bailey and Davis, 1942), whereas a lesion in any part of the neocortical brain will leave rudimentary forms of consciousness intact (Panksepp, 2005). This distinction not only suggests that there is a distinct form of affective consciousness but also that it serves as a platform for hierarchically higher cortical forms of consciousness. Consciousness cannot emanate from higher cortical centers of the brain since their shutdown does not interrupt the basic conscious presence.

Recent evidence identifies a small area of the brain (the *claustrum*) whose stimulation reliably closes down all consciousness while leaving waking-sleeping patterns unperturbed (Koubeissi et al., 2014). This subcortical area is the highest region of the brain that is shown to be necessary for consciousness. The *claustrum* is generally considered part of the *basal ganglia* complex (Solms and Turnbull, 2002): the old “reptilian” brain structures (Panksepp, 1998: 66–67), which was discussed earlier (Ch. 4). Moreover, the neuronal structure and composition of the *claustrum* is markedly anatomically different from that of cortical areas, even those of the insular cortex (Rahman and Baizer, 2007).

In addition to the above evidence, deep brain stimulation is demonstrated to produce specific emotional states. Thus, not only affective consciousness in general but also basic conscious emotional feeling states appear to be produced subcortically. While directed neurochemical and electrical stimulation of specific subcortical systems can reliably trigger

different affective emotional experiences and expressions (Ikemoto, 2010; Panksepp, 1998), stimulation of neocortical areas in contrast tends to impair function and is incapable of causing similar emotional effects (Panksepp et al., 2012). The traditional but still popular view held by social constructivists about the mind in which the experience of affect requires an ability to conceptualize or language capabilities is demonstrably incorrect according to Panksepp and Biven (2012: 77).

The affect aspect of subjective experience, moreover, is ubiquitous to consciousness since our emotionally based affective systems are continuously in operation and cannot be turned off (Kahneman, 2011: 105; Panksepp, 1998), not even during sleep (Solms and Turnbull, 2002), as is evident from our often vivid and emotional affective experiences while dreaming. Hence, affect is primary in the sense that it is always present; it serves as a basis for everything else in human psychology. Consequently, all cognitive states of mind will also be infused with affect, inasmuch as affective consciousness is a premise for these mental states.

Affective neuroscience provides an updated understanding of consciousness that paradigmatically displaces former theorizing on the topic: “Consciousness is not generated in the cortex; it is generated in the brainstem. Moreover, consciousness is not inherently perceptual; it is inherently affective” (Solms, 2013a: 106). Some of the basic clinical studies that corroborate these claims have been around for a long time and thus have a long track record in neuroscience. Recent data confirm this picture, allowing researchers to invest in affective theories of the mind with greater confidence and precision.

Although largely ignored by the various mind sciences (philosophy, psychology, psychiatry, psychoanalysis, and neuroscience), the subconscious origins of consciousness have in fact been known for some time in neurophysiology (Panksepp, 1998). Moruzzi and Maguon (1949) demonstrated that states of consciousness, as inferred from brain activity

visible by electroencephalography (EEG) observation, originate in deeply subcortical reaches of the brain, more specifically in the upper brainstem (Solms, 2013a: 104; Merker, 2007), which has been corroborated by recent findings (Solms, 2013a).

Based on the evidence discussed above, the position that Solms, Damasio, and Panksepp, as well as an increasing number of other scientists now take concerning human consciousness seems reasonable, although it implies a paradigm shift for how consciousness is understood. This position suggests that *affect* is quintessentially *conscious*, whereas traditionally affect has been considered *unconscious*. Conversely, we may suppose that the cortex is essentially unconscious unless it is *rendered* conscious by interaction with subcortical affective structures (Solms and Panksepp, 2012).

It is by now accepted by several neuroscientists (e.g., Berlin, 2011) and psychologists (e.g., Bargh and Morsella, 2008) that much of the cortical activity is unconscious. A leading hypothesis is that the cortex is rendered conscious so that the affective self can interact with the external world in explicit terms. Thus, according to Solms (2013b: 14), the cortex “generates stable, representational ‘mental solids’ that, when activated (or ‘cathected’) by affective consciousness, enable the id to *picture* itself in the world and to think.” Here, the “id” denotes the affective self.

5.2.4 The concept of consciousness

The definition of “conscious” is disputed. I will not go into all the facets of this debate, but it can be noted that, despite disagreements over details, most of the current authorities in the study of consciousness seem to agree on a general definition of consciousness as a “first-person subjective experience” (Faw, 2006). Hence, most of the disagreement concerning the definition of consciousness relates to the details entailed by more specific definitions. I will focus on some areas where evidence has shed more light relating to emotions and affects.

As mentioned (Ch. 2), cognitive theories of emotion tend to express or implicitly represent a semantic overlap between “consciousness” and “cognition.” However, the “conscious” is deployed in a variety of ways in different professional (as well as ordinary) settings. At the cognitive end of the theoretical spectrum, “conscious” is considered to be equivalent to “aware,” as in, “I am consciously aware of your presence in the room”; at the other end, you have the neurologists who by “conscious” simply mean that a person is not comatose; cognitive awareness is not what is in question. For the sake of analytical clarity, a more subtle distinction might be helpful.

One crucial distinction that clinical neuroscientists tend to make is that between consciousness and “wakefulness” as opposed to coma; in this view, consciousness is not the same as mere wakefulness (Damasio, 2010: Ch. 1). It becomes meaningful to talk of consciousness once a person has subjective experience; “primordial” subjective experience occurs subcortically, somewhere after wakefulness and before the onset of cognition (Damasio, 2010). Damasio writes of a “triad” related to consciousness that includes “wakefulness,” “mind,” and “self,” which he conceives of as being hierarchically structured.

Panksepp (2005) proposes that consciousness is *layered* and that we should think of a layer of *affective consciousness* as the bottom layer that in humans serves as a foundation for higher (cognitive) levels of consciousness. This view suggests that our sense of self-awareness is *gradual*. Furthermore, it strengthens the plausibility of the view that our *feeling* of normativity and our *perceiving* things in the world as “good,” “bad,” “attractive,” “repulsive,” “right,” “wrong,” etc., are fundamentally based on emotional affects and the corresponding neurodynamics of the brain. Similarly, Turnbull and Solms (2002: 92) suggest that consciousness “consists of feelings (evaluations) projected onto what is happening around us” or what our perceptions cause us to believe is happening around us.

Panksepp's (2005) proposition is interestingly different from Damasio (1999) who, instead of distinguishing between different levels of consciousness, proceeds by dividing it into different selves: "protoself," "core self," and "autobiographical self." Two points that seems to favor Panksepp's version are that layered consciousness (1) makes it easier to conceive of a unified "self," and (2) arguably fits better with the neuroanatomical properties of the brain; it seems more straightforward to point out where different layers of consciousness are located than to find the locations for the different selves. Damasio (1999) and Panksepp (1998) both agree, however, that the essence and origins of emotion are affective and present prior to cognitive awareness and reflection.

5.3 Philosophy of affective emotions

5.3.1 Normative emotions and philosophy

Discussions of affect and emotions have always been central in moral philosophy. The importance of emotions has not always been recognized; in some philosophical traditions, the normative role of affects and emotions has been (and remains) ignored or even denied. A considerable number of prominent philosophers dismiss all biologically based notions, including "affect" and "emotion," in their search for the "proper" (i.e., objective, universal, external) foundations of normative ethics. This concept is what Blackburn (2010: 128) characterizes as the search for "the Holy Grail" of moral philosophy.

Kant's principle-based deontology is in this respect a particularly influential example, with an array of contemporary followers in moral philosophy (e.g., O'Neill, 2014; Korsgaard, 1996) as well as in applied ethics, such as in business ethics (e.g., Smith and Dubbink, 2011; Bowie, 1999). As an example of Kantian dismissal of the affective, Kant (1785/1959: 14) held that actions performed "without any vanity or selfishness" but merely for "an inner satisfaction in spreading joy" for that matter still had "no true moral worth." Roughly,

Kantians argue that “moral worth” requires that agents act in the spirit of and in conformance with principles derivable on the basis of absolute laws (or maxims) arrived at by “pure” logical introspective reasoning. Self-evident “moral truth” must be arrived at in a strictly cognitive fashion without any influence from the “dark” affective forces of emotions.

Kant appears to have understood the nature of emotions as pre-cognitive feelings (Bagnoli, 2011: 70). He viewed emotions as “random, unruly, incorrigible, and unstable” (Bagnoli, 2011: 70) and concluded that emotions and feelings have no moral value and cannot serve as constituents of the normative foundations of ethics. It does not help the moral state of emotions, in Kant’s view, that emotions sometimes happen to conform to (supposed) objective duties. One should note that Kant proposed something like a scientific outside-in view of the *normative* status of emotions but denies entirely the relevance of the internal influence of emotions. Similarly, Wallace (2006: 48) rejects the view that reasoning motivates due to “causal forces within us” as opposed to the guidance of “deliberative reflection” over recognized reasons.

A difficulty with the Kantian approach, and related approaches, is that it assumes that we already have a “valid” normative position from which to make an objective and impartial evaluation. It seems that logical reasoning by itself can only provide structures of thought, but as I have argued, moral evaluation also requires normative *content* about which to reason. The important question is not so much whether emotions and feelings *have* value (as Kant seems to have thought) but whether they *convey* value (equivalently “normative content”). There may be no way to “validate” or justify normativity from the outside as externalism presupposes (see Ch. 8). As internalism argues, it may well be that the only possible route to normative justification is subjective, as when we justify our actions *to* one another (subject to subject) and to ourselves.

From his quasi-Kantian rationalistic eudemonistic position (also against the internalist view I propose) Wallace (2006: 7) suggests that desires can become truly autonomous. He supposes that they become autonomous by virtue of the fact that agents reflectively affirm their normative content, thereby giving this normative content rational independence from the desires from which they originate. As pointed out in Chapter 3, however, this idea of rational independent free will is problematic. Wallace (2006: 30) goes on to criticize the “Humean view” that all desires, including desires to be influenced by rationality, trace back down to basic desires that are beyond the scope of conscious rationalizing. In Chapter 4 it, however, it was shown that precisely this basic Humean view is supported by solid empirical evidence. The brain and mind are appropriately described as, Haidt (2001) famous metaphor, a case of “the emotional dog wagging its rational tail”.

Even the special mental ability of humans to take our own subjectivity as an object of reasoning (Solms and Turnbull, 2002) does not mean that this object, nor the mechanisms by which the self as object is held in mind (i.e. the frontal lobes), nor second-order desires and emotions that take first-order desires and emotions as objects (Frankfurt, 1988: 11-25; 1971), are in fact independent from the influence of first-order desires and emotions. Contemporary neurobiological evidence, as clarified in previous chapters, suggests strong and causal bottom-up dependence relation. The role of affective emotions in decision-making is detailed in the next chapter (Ch. 6).

Besides direct followers of Kant, there are many notable cognitivist moral philosophers who do not recognize a legitimate role for affects and emotions, including Broome (2013; 2004), Parfit (2011), Dancy (2004: 165–170), and Nagel (1986); “Hence, many contemporary philosophers compete to stress and to extol the external nature of reasons” (Blackburn, 2010: 283). Philosophical accounts that seek to explain and justify ethics outside the affective can be called “externalism” in regards to their value or reasons.

Correspondingly, accounts of normative ethics that base the explanation and justification of ethics on the affective can be termed “internalism.” I defend an internalist position (see Ch. 7 and Ch. 8).

In contrast to these “affectively cold” perspectives, internalists view affective emotions as central to morality. It is generally taken to be central in two ways; it is central to the *explanation* of morality and to the *justification* morality. The internalist position has also been influential, at least since the intellectual entries of Hume and Shaftesbury (Nichols, 2004: 29); I mentioned several prominent defenders of this position in the review chapter on the emotions literature (see Ch. 2). According to Blackburn (2010), the four main arguments for internalism are as follows: (1) It offers an adequate motivational explanation of moral thought, talk, and action. (2) An externalist explanation of moral motivation is difficult to defend and hence not credible. (3) An external justification of morality does not seem possible. (4) An external justification of morality is not needed because we can get on just fine without it.

The preceding chapters have provided a solid case for the motivational role of emotional feelings. Since emotional feelings are internal in the right sense, the first point above is supported on the basis of current scientific evidence in addition to other internalist philosophical arguments. An analogous argument can be made for the second point. I will return to point three later (see Ch. 8), but I believe it is defensible as well. The fourth point is somewhat controversial. It does not seem implausible that, if external justification turned out to be possible after all, we might find that it is something we “needed.” As things stand, this possibility is a bit intractable, however, because it may be difficult to imagine what a fully acceptable external justification would be like. I am not sure we “get on just fine” without justification, if justification is something that would cause motivation. There are many

problems in this world that warrant motivated action, ranging from anthropogenic global climate disruption to exploited workers in poor countries.

The key question seems to be motivation, since that (along with smooth social coordination) is in essence what is needed to address what we see as moral problems. This analysis suggests that both the assessment of something as a moral problem, the solution to such problems, as well as the main obstacles to addressing these problems are internally based in our human biological makeup. Our ability to justify our actions to ourselves and to others may then be determined by the degree to which our biologically determined emotional affects feel more like solutions than obstacles to what we see as moral problems. As I will explain further out (see Ch. 9), this perspective does not seem to hold a promise of full moral justification but something more akin to “justification light” instead.

5.3.2 Emotions and desires

Each primary emotion is either negatively or positively valenced (Panksepp, 1998: 48) as inferred from behavior and systematically retrieved from subjective reports; the former type causes us to *avoid* and escape from things, while the latter makes us *approach* things. However, it should be noted that feelings are often complex, involving several emotions at once in addition to non-emotional affects and cognitive elements. Therefore, feelings are often ambivalent rather than unipolar (Goldman, 2009). For example, you might be pleasantly drawn towards sweets at the same time that you have a cognitively initiated repulsion towards sweets because you consciously or unconsciously connect them with the prospect of becoming fat.

Positively valenced emotions can be characterized in terms of “desires,” which are affectively based. The popular alternative view that desires are generated by experience is inconsistent with current evidence from brain research (Panksepp, 1998: 24). According to

Goldman (2009: 7), desires are not in themselves what motivate us but rather states of being motivated. In neural terms, desires appear to be underpinned by the collaboration of different affective systems.

The “seeking system” is central because it appears to serve as a general purpose platform that energizes us and make us engage with the world (Panksepp and Biven, 2012; Panksepp, 1998) without any particular goal in sight. Meanwhile, other affective systems cause us to seek out appropriate objects and situations often in collaboration with cognitive systems. The affective “play system,” for example, can make us seek out opportunities for having fun (Panksepp, 1998), where the scope of what is fun can be regulated and delimited cognitively. This system depends on what our active goals are and our awareness of what is socially acceptable. Desires, in this perspective, are affectively based and cognitively refined.

5.3.3 The nested hierarchy of emotions

In the review chapter (Ch. 2), I addressed some concerns regarding the possibility of constructing a holistic account of emotions. Griffiths (1997) argued that it is impossible to reconcile an account of the basic affective emotions with an account of more cognitive “higher cortical emotions.” His stance was that these accounts represent two entirely different phenomena—“not of the same kind”—and therefore need to be *studied separately*. A similar but more optimistic argument is made by Deigh (2004), who holds that it is difficult to reconcile the “fact” that emotions are intentional (a sticking point for most cognitive theories of emotion) with the “fact” that they are shared by a number of other animal species and babies. Deigh’s solution to this problem is to propose that intentionality and evaluative judgments of “higher emotions” (i.e., the cognitive components of emotion) are learned. However, the involvement of learning in higher-order emotions neither implies that they are unrelated to lower emotions nor that they are of a fundamentally different kind.

Meanwhile, Panksepp presents a holistic theory of emotions, which suggests that emotional and cognitive processes are hierarchically structured (Panksepp and Biven, 2012; Panksepp, 2012). In this theory, the emotional hierarchies of the brain are nested bottom-up (Panksepp and Biven, 2012)—with primary affective processes at the bottom, secondary learning and memory processes in the middle, and tertiary cognitive processes at the top—consequently, there are no purely cognitive emotions. Higher-level emotions are always grounded in the lower affective processes of primary emotions. For example, according to Watt (2003: 86), the emotion “shame reflects activation of a sense of oneself as exposed and defective in the eyes of significant others, and depends on the underlying prototype of separation distress” (i.e., the panic-grief system). It also seems reasonable to assume that the content of the higher cortical cognitive processes is affective and supplied as input by lower affective processes.

According to Panksepp *primary*, *secondary*, and *tertiary* emotions are to be conceived as distinct but at the same time closely inter-related categories and, crucially, higher-level emotions depend on lower-level emotions because they are in part composed of them. The reverse direction dependency, by contrast, is not obvious and, if present, is certainly much less direct.

In Panksepp’s view, emotions at every level involve an affective experiential component. Conceptually, therefore, *affect* can be taken as a central definitional characteristic of emotions. This interpretation of Panksepp’s theoretical position *adds* richness to his main argument for defining emotions, which is that emotions have to be neurophysiologically and neurochemically characterizable, in terms of neural circuits and neurotransmitters (as presented in Ch. 4). This view allows us to hold that emotions are, on the one hand, *objectively* definable in physiological and chemical terms and, on the other hand, *subjectively* definable in terms of affective experiences. The definition of emotion thereby straddles both

the subjective and the objective domain in a way that is aptly captured by dual-aspect monism (as presented in Ch. 3). Hence, this way of defining emotions also corroborates our theoretical understanding of the relation between the mind and the brain.

This perspective on emotions has several further advantages; I will highlight a few points. (1) The affective dimension becomes a reasonable way of understanding why *only some* cognitive processes appear to be *part of* emotions (i.e., tertiary emotions), whereas many other cognitive processes do not. It is the interaction with (the flavors of) affective processes that *causes* specific cognitive processes to be experienced as emotional. For example, an object of resentment becomes resented due to its connection affect based in the anger system or specific affective memories that are stored at the secondary level (i.e., the hippocampus), or both.

(2) We can explain why *so many* emotions appear to have a cognitive component by explaining the prevalence of tertiary emotions. The explanation is that unless urgent action is needed, affective processes are liable to (eventually) be detected in the neo-cortex and will be subject to further cognitive processes there, adding layers of experience and emotional distinctiveness.

(3) For similar reasons we can explain why many emotional experiences come as *episodic* arousals. Emotional episodes come and go with cognitively represented objects that are infused with and motivated by affect. Once attention is no longer needed or is shifted in another direction, emotional arousal dies down. Similarly, we can understand why some background emotions are durable fleeting states. For example, general states of happiness or gloominess are essentially subcortical and surface only episodically into the realm of cognitive object-representation and reasoning.

Building on insights from previous chapters, the current chapter connects normativity and emotional affect. Specifically, I propose that what a person reports as a normative “sense,” “feeling,” “intuition,” “perception,” “inclination,” etc., and categorizes in terms of normative “values,” “virtues,” etc., corresponds to related affective experiences in that person. This position is congruent with dual-aspect monism (Solms and Turnbull, 2002) and the current body of evidence provided by affective neuroscience (Panksepp and Biven, 2012).

To a considerable extent, this reasoning explains why emotional feelings have the special characteristics they do. Emotional feelings are characteristically non-neural, valenced, potentially motivating, and full of subtle sensibilities that are difficult to articulate. They also typically arise in us “uninvited.” For example, we get angry rather than decide to become angry; once we are angry, meanwhile, we can to some extent either let the anger flow, or we can put in an effort to try to prevent it, suppress it, or direct it, but our ability to do so will in part depend reflexively on our overall emotional state.

5.3.4 Second-order desires

Harry Frankfurt (1988) proposes “secondary desires” play a crucial role in human morality and identity. A similar view was argued by Bertrand Russell (1896/1983; 1897/1983). Secondary desires indicate *desires about desires*; we can, for example, approve of or disapprove of our own first-level desires. As Frankfurt (1988: 12) explains, “Besides wanting and choosing and being moved *to do* this or that, men may also want to have (or not to have) certain desires and motives.” Frankfurt (1988) linked second order desires explicitly to identity and moral deliberation.

In light of recent neuroscientific evidence (Berridge and Kringelbach, 2013; Panksepp and Biven, 2012; Robinson and Berridge, 2003; Berridge, 1996), this view appears empirically plausible. Berridge and Robinson (2003) show that different types of desires are

neurologically distinguishable. They argue that “wanting” (e.g., seeking and anticipating something good) and “liking” (e.g., the pleasure of consuming something good) represent separate hedonic emotional centers in the brain. Similarly, Panksepp and Biven (2012) argue that there are not one but several reward centers in the brain, and that the positive emotions of seeking and anticipating are characteristically different and anatomically separate from the various rewards centers that give pleasure (see Ch. 4).

Separate hedonic emotional centers in the brain allow a person to have emotionally *ambivalent* feelings, such as liking what they do not want to like (e.g., unsophisticated soap operas) or wanting what they do not like to want (e.g., one more piece of cake), with the valence pointing in opposite directions (Litt et al., 2010). This behavior implies that there are *levels of normativity* in the moral psychology of our minds. How many levels we are capable of is in principle an empirical question (as opposed to a logical one), but the current evidence supports Russell and Frankfurt’s (1988) assumption that there are two.

This view suggests that many different considerations can shape our internal evaluation of our own desires: for example, what we have learned or have been socialized into, such as our cultural and social identities. Panksepp (1998: 319) provides a related example, which hints at the possibility of an affectively grounded contractual ethics: “For instance, explicit spoken or written contracts between humans help minimize disputes that would easily emerge if one merely followed the dictates of one’s immediate wants and desires.”

Moreover, it seems a reasonable conjecture that considerations at the secondary level are, or at least tend to be, closely tied to our sense of self and self-reflective consciousness. While the core sense of self is essentially emotional (Solms and Turnbull, 2002) and primary desires form a platform for our sense of self, it seems hard to construe primary desires as self-

reflective at all. Secondary desires therefore point in the direction of a socially responsive yet affectively based sense of morality. This self-reflective sense of self, which permits extended moral capacities, is likely to be the result of evolutionary advantage (Solms and Turnbull, 2002).

It is possible that full-fledged human morality is in part a by-product of self-reflective consciousness. It should be emphasized that this point would in no way diminish the importance of morality as seen and felt from the internal, perspectival, deliberative point of view. From an externalist view of morality, in contrast, this possibility would be a reason to distrust and discount the importance of human morality and its psychology, which perhaps then needs to be replaced by some ideal ethical system.

5.4 Conclusions

Normativity can reasonably be viewed as a special type of affective experience that draws on emotions for its distinctive qualitative feel and valence. This theoretical view of normativity can be called the *normativity-as-affect position*. Insofar as affective experiences are never neutral, always possessing some elements of attraction or repulsion, normativity may be an inherent aspect of affect humans as well as other animals to whom we are related.

Evidence from affective neuroscience suggests that the traditional view of consciousness as a unified, cortically based mental phenomenon should be abandoned in favor of a view of consciousness as *layered* in a nested, hierarchical manner. From this perspective, the fundamental layer of consciousness is the *affective consciousness* based subcortically in the brain. Since affect is always conscious, at least in some basic form, normativity can be seen as a fundamental characteristic of consciousness.

Affective consciousness is extended so as to render higher cortical regions of the brain conscious. Higher forms of consciousness include a level that permits observing and learning

in relation to the external environment, culminating in the level of the highest forms of consciousness, which allow sophisticated self-reflexive thought and awareness. All types of consciousness are shaped by and infused with affect, but higher forms of consciousness also serve to regulate various affective impulses.

Chapter 5 References

- Bagnoli, C. (2011) Emotions and the categorical authority of moral reason. In C. Bagnoli (Ed.) *Morality and the emotions*. Oxford: Oxford University Press.
- Bailey, P. and Davis, E. W. (1942) Effects of lesions of the periaqueductal gray matter in the cat. *Proceedings of the Society for Experimental Biology and Medicine* **351**: 305–306.
- Bargh, J. and Morsella, E. (2008) The unconscious mind. *Perspectives on Psychological Science* **3** (1): 73–79.
- Berlin, H. (2011) The neural basis of the dynamic unconscious. *Neuropsychoanalysis* **13** (1): 5–31.
- Berridge, K. C. and Kringelbach, M. L. (2013) Neuroscience of affect: Brain mechanisms of pleasure and displeasure. *Current Opinion in Neurobiology* **23**: 294–303.
- Berridge, K. C. (1996) Food reward: Brain substrates of wanting and liking. *Neuroscience and Biobehavioral Review* **20**: 1–25.
- Blackburn, S. (2010) *Practical tortoise raising and other philosophical essays*. New York: Oxford University Press.
- Bowie, N. E. (1999) *Business ethics: A Kantian perspective*. Oxford: Blackwell Publishers.
- Broome, J. (2013) *Rationality through reasoning*. Oxford: Wiley-Blackwell.
- Broome, J. (2004) *Weighing lives*. Oxford: Oxford University Press.
- Damasio, A. (2010) *Self comes to mind: Constructing the conscious brain*. New York: Pantheon Books.

- Damasio A. (1999) *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt.
- Dancy, J. (2004) *Ethics without principles*. Oxford: Oxford University Press.
- Deigh, J. (2004) Primitive emotions. In R. C. Solomon (Ed.), *Thinking about feeling: Contemporary philosophers on emotions* (pp. 9–27). New York: Oxford University Press.
- Donaldson, Z. R. and Young, L. (2008) Oxytocin, vasopressin, and the neurogenetics of sociality. *Science* **322** (5903): 900–904.
- Faw, B. (2006) Are we studying consciousness yet? *Journal of Consciousness Studies* **13** (4): 94–112.
- Feinberg, T. (2000) The nested hierarchy of consciousness: A neurobiological solution to the problem of mental unity. *Neurocase* **6**: 75–81.
- Frankfurt, H. G. (1988) *The importance of what we care about: Philosophical essays*. Cambridge: Cambridge University Press.
- Frankfurt, H. G. (1971) Freedom of the will and the concept of a person. *Journal of Philosophy* **68**: 5–20.
- Freedberg, D. and Gallese, V. (2007) Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Science* **11** (5): 197.
- Gallese, V. (2005) Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences* **4** (1): 23–48.
- Gibbard, A. (2012) *Meaning and normativity*. Oxford: Oxford University Press.

- Goldman, A. H. (2009) *Reasons from within: Desires and values*. Oxford: Oxford University Press.
- Griffiths, P. E. (1997) What emotions really are: The problem of psychological categories, *Chicago: University of Chicago Press*.
- Haidt, J. (2001) The emotional dog wagging its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* **108** (4): 814–834.
- Huxley, T. H. (1870) ‘Has a frog a soul, and of what nature is that soul, supposing it to exist?’, *Papers for the Metaphysical Society* **1**: 1-7.
- Ikemoto, S. (2010) Brain reward circuitry beyond the mesolimbic dopamine system: A neurobiological theory. *Neuroscience & Biobehavioral Reviews* **35** (2): 129–150.
- Kahneman, D. (2011) *Thinking, fast and slow*. New York: Farrar, Straus, and Giroux.
- Kant, I. (1785/1959) *Foundations of the metaphysics of morals: And what is enlightenment?* L. W. Beck (Ed.). London: Macmillan.
- Korsgaard, C. M. (1996) *Creating the kingdom of ends*. Cambridge: Cambridge University Press.
- Koubeissi, M. Z., Bartolomei, F., Beltagy, A., and Picard, F. (2014) Electrical stimulation of a small brain area reversibly disrupts consciousness. *Epilepsy & Behavior* **37**: 32–35.
- Litt, A., Khan, U., and Shiv, B. (2010) Lusting while loathing: Parallel counterdriving of wanting and liking. *Psychological Science* **21** (1): 118–125.
- Mandler, G. (2002) *Consciousness recovered: Psychological functions and the origins of conscious thought*. Amsterdam: John Benjamin’s Publishing Company.

- Merker B. (2007) Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral Brain Science* **30**: 63–134.
- Modell, A. H. (2003) *Imagination and the meaningful brain*. Cambridge: The MIT Press.
- Moruzzi, G. and Magoun, H. W. (1949) Brain stem reticular formation and activation of the EEG. *Electroencephalograph and Clinical Neurophysiology* **1**: 455–473.
- Nagel, T. (1986) *The view from nowhere*. Oxford: Oxford University Press.
- Nichols, S. (2004) *Sentimental rules: On the natural foundations of moral judgment*. Oxford: Oxford University Press.
- Northoff, G. (2013) *Unlocking the brain: Volume 2*. Cambridge: Cambridge University Press.
- O’Niell, O. (2014) *Acting on principle: An essay on Kantian ethics*. Cambridge: Cambridge University Press.
- Panksepp, J. and Biven, L. (2012) *The archeology of mind: Neuroevolutionary origins of human emotions*. New York: Norton.
- Panksepp, J., Asma, S., Curran, G., Gabriel, R., and Greif, T. (2012) The philosophical implications of affective neuroscience. *Journal of Consciousness Studies* **19** (3–4): 6–48.
- Panksepp, J. (2012) What is an emotional feeling? Lessons about affective origins from cross-species neuroscience. *Motivation and Emotion* **36**: 4–15.
- Panksepp, J. (2011) Empathy and the laws of affect. *Science* **334**, 1358–1359.
- Panksepp, J. (2005) Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition* **14**: 30–80.

- Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*.
New York: Oxford University Press.
- Parfit, D. (2011) *On what matters: Volume 1*. Oxford: Oxford University Press.
- Prinz, J. J. (2007) *The emotional construction of morals*. Oxford: Oxford University Press.
- Rahman, F. E. and Baizer, J. S. (2007) Neurochemically defined cell types in the claustrum of the cat. *Brain Research* **1159**: 94–111.
- Robinson, J. (2005) *Deeper than reason: Emotion and its role in literature, music, and art*.
London: Clarendon Press.
- Robinson, T. E. and Berridge, K. C. (2003) Addiction. *Annual Review of Psychology* **54**: 25–53.
- Russell, B. (1897/1983) The collected papers of Bertrand Russell, Vol. 2. In K. Blackwell, A. Brink, N. Griffin, R. Rempel, and J. Slater (Eds.), *Cambridge essays 1888–1899*.
London: Allen & Unwin.
- Smith, J. and Dubbink, W. (2011) Understanding the role of moral principles in business ethics: A Kantian perspective. *Business Ethics Quarterly* **21** (2): 205–231.
- Solomon, R. C. (2007) *True to our feelings*. Oxford: Oxford University Press.
- Solms, M. (2013a) “The unconscious” in psychoanalysis and neuropsychology. In S. Akhtar and M. K. O’Neil (Eds.), *On Freud’s “The Unconscious.”* London: Karnac Books.
- Solms, M. (2013b) The conscious id. *Neuropsychoanalysis* **15** (1): 5–19.

- Solms, M. and Panksepp, J. (2012) The “id” knows more than the “ego” admits: Neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Sciences* **2**: 147–175.
- Solms, M. and Turnbull, O. (2002) *Brain and the inner world: An introduction to the neuroscience of subjective experience*. New York: Other Press.
- Tallon, A. (1997) *Head and heart: Affection, cognition, volition as triune consciousness*. Bronx: Fordham University Press.
- Wallace, R. J. (2006) *Normativity and the will: Selected essays on moral psychology and practical reason*. Oxford: Oxford University Press.
- Watt, D. F. (2003) Psychotherapy in an age of neuroscience. In J. Corrigan and H. Wilkinson (Eds.), *Revolutionary connections: Psychotherapy and neuroscience* (pp. 79–116). London: Karnac Books.
- Zahavi, D. (1999) *Self-awareness and alterity: A phenomenological investigation*. Evanston: Northwestern University Press.

Chapter 6: The emotional decision argument: Moral decisions require appropriate emotional input

6.1 Introduction to the emotional decision argument

There is increasing recognition that affective emotions play an important role in decision-making (Lerner et al., 2015; Damasio, 2010; 1994; Koole, 2009; Lawton et al., 2009: 57; Salvador and Fogler, 2009; Thagard, 2007; Bechara, 2004; Davidson, 2003: 655; Slovic et al., 2002; Etzioni, 1988: 88–113; Frankfurt, 1988: 189; Zajonc, 1980). Advances in neuroscience have, to a considerable extent, paved the way for such recognition. However, until recently, decision-makers “were assumed to evaluate the potential consequences of their decisions dispassionately” (Loewenstein and Lerner, 2003: 619). Emotions thus attracted little attention in the field of decision-making (Bechara, 2004; Slovic et al., 2002; Damasio, 1994).

Indeed, traditionally, the emphasis has been on emotions as a source of disturbance and bias that *impedes* and *biases* otherwise sound and rational decision-making (Damasio, 2010) to the point that “raw emotions are viewed as antithetical to reason” (Etzioni, 1988: 102). While this picture still persists across philosophy, psychology, and neuroscience, upheld by a cortico-centric view of the mind and consciousness, cumulative empirical research suggests that this negative view of the role of emotions needs to be updated (Panksepp et al., 2012: 30; Damasio, 2010; Panksepp, 2005; Solms and Turnbull, 2002). Far from being antithetical to reason, emotions appear to underpin and permeate reasoning. It remains true that emotions biases decisions in profound ways (Fine, 2008: 58) but they are at the same time essential and indispensable for making good decisions (Damasio, 1994).

The aim of this chapter is to explain the important role that affect and affective emotions play in practical decision-making. As indicated by the discussion in the previous chapter, emotions can be seen as a biological source of normativity, and their primary

function is to serve as a guiding compass for making moral decisions. *The emotional decision argument* posits that input from affective emotions is, under typical practical situational circumstances, a requisite for ensuring the quality of moral decisions.

Depending on the precise circumstances, emotional input is well suited to provide guidance in either implicit-intuitive decision-making or explicit reasoning. Either way, in normally functioning adult humans, emotions are an integral part of moral reasoning rather than a separate entity from it. Affect and emotions, as I see it, provide the *normative content* of reasoning. This normative content represents our concerns and prioritized desires, and to that extent reduces arbitrariness in decision-making. Below I will outline the neuroscientific underpinnings of decision-making.

6.2 The neuroscience of moral decision-making

Building on the concept of consciousness presented in the preceding chapter, it seems appropriate to think of affective consciousness as branching like a tree of neural connectivity, with its roots in subcortical structures and its branches reaching into the higher realms of the brain. Affective content is fed into this cortical region from subcortical emotional structures (Panksepp and Biven, 2012; Damasio, 2010), most notably the *basal ganglia*, *amygdalae*, and *periaqueductal gray* (PAG). Even before affect reaches the higher cortical areas, a considerable amount of cognitive-unconscious reasoning take place (Berlin, 2011). Thus, Damasio (2010: 276) argues that bodily signals (“somatic markers”) reflect an “important reasoning process going on nonconsciously.” Only a small fraction of the totality of reasoning that takes place in the brain ever reaches explicit awareness. However, the motivational aspects of this implicit reasoning affect our attitudes and experiences emanating from lower levels of consciousness.

Once affective content reaches the higher cortical areas of the brain, more complex cognitive operations can take place, including deliberative operations of which we have partial explicit awareness. Since explicit focus requires mental energy (Berlin, 2011), the brain strives to automatize cognitive processes (Guttmann et al., 2006), rendering it unconscious. While explicit cognitive attention may sometimes be required in decision-making for the sake of increased *precision*, unconscious cognition has a vastly greater *capacity* than conscious cognition (Kiefer, 2012; Berlin, 2011; Bargh and Morsella, 2008). Much of decision-making is therefore likely to rely on the comparatively vast capacities of the cognitive unconscious.

Moral decision-making is highly complex, reflecting the social and physical complexity of the decision context. As a consequence, decision-making needs to draw on many parts of the brain at once. Thus, the moral decision-making processes draws on “the frontal lobe (which includes the medial orbitofronal complex (mOFC) and the dorsolateral prefrontal cortex (DLPFC)), subcortical-limbic structures (which include the amygdala and the hippocampus), as well as other brain structures” (Salvador and Fogler, 2009: 6). Brain-anatomically, the *ventromedial prefrontal cortex* (equivalently, the mOFC) appears to be a central convergence zone for moral judgment (Damasio, 2010); “Converging evidence suggests that emotion processing mediated by ventromedial prefrontal cortex (vmPFC) is necessary to prevent personal moral violations” (Moretto et al., 2010: 1888).

Another central aspect of the integration that takes place in the vmPFC is that it draws on stored memory, which allows thought to be extended in time. This process includes contemplating past events as well as anticipating future ones (Wright and Panksepp, 2012: 6); sometimes referred to as “mental time travel” (Gerrans and Sander, 2014; Ciaramelli and di Pellegrino, 2011). As a future-oriented activity, the quality of decision-making is associated with the ability to extend emotions temporally as “delayed rewards, either by computing the

value of future outcomes (i.e., valuation), or by enabling people to imagine and represent future rewards and their consequences (e.g., prospection)” (Sellitto et al., 2011: 565). The frontal cortex thereby allows anticipation of future emotions based on current affects and emotions, and the articulation of “expected emotions” (Loewenstein and Lerner, 2003: 621–626). Affects thereby can be projected onto future hypothetical objects.

Are any parts of the brain specifically designated for decision-making tasks? One possibility that has been investigated is the parts of the brain associated with its “executive functions.” In a functional account of the brain, Luria (1980; 1973) distinguished three major functions: (1) arousal-motivation (limbic and reticular systems); (2) receiving, processing, and storing information (posterior cortical areas); and (3) programming, controlling, and verifying activity (involving, but not necessarily limited to, the pre-frontal cortex). He proposes that the third part of the brain performs the executive role. However, these executive functions need appropriate input from affective emotional systems in order to perform their tasks in ways that are normatively meaningful (Damasio, 2010).

The executive functions of the brain make practical thinking and reasoning possible (LeDoux, 2002: 178). They are therefore indispensable to decision-making. Executive function is also a useful concept for thinking about the brain more generally. However, attempts to localize the executive functions to discrete brain areas have been inconclusive (Ardila, 2008: 93). Evidence from neuroimaging shows that executive functioning involves both cortical and subcortical parts of the brain (Roberts et al., 2002). As previously proposed (Ch. 5), the bottom-up hierarchical structure of the brain and of consciousness are central for understanding the workings of the brain, as the integrated and massively interconnected organ is built on affective foundations.

While the anatomy of the brain explains many of its workings, a more complete view of how the brain works can be obtained by recognizing that there is not only a physical explanation but also a chemical one. As far as brain chemistry is concerned, Litt et al. (2008) suggest that decision-making is an emotional process that uses two distinct neurochemical brain systems: the dopamine system for positive events and the serotonin system for negative events. There are also other neurotransmitters, but these two are most important. Dopamine and serotonin are the most general-purpose neurotransmitters allowing the synapses (small gaps between neurons) to be bridged, thus permitting information sharing between neurons.

The different emotional circuits in the brain that neutrally link specific regions of the brain (Panksepp, 1998; LeDoux, 1996) can be shown to depend on specific neurotransmitters, but there is normally a range of different ones. Serotonin and dopamine alter the *global state* of the chemical “soup” of the brain and thus affect many brain areas at the same time (LeDoux, 2002: 58). Dopamine, for example, has been shown to be important in relation to reward (Di Chiara et al., 2004) as well as associative learning (Schultz, 2002). Several more specific neurotransmitters, in addition to serotonin and dopamine, are involved in the complex cognitive-affective tasks that decision-making requires. Chemical imbalances in the brain can cause decision-making activities to malfunction in different ways, as when specific emotional circuits can be over-sensitized (Panksepp, 1998). I will now turn to look more closely at the role of emotions in practical decision-making.

6.3 The role of emotions in decision-making

Practical decision-making is challenging, especially in relation to moral decisions in socially complex contexts, such as those typically faced by professionals in various professions. The social context surrounding the decision-making process strongly impacts how decisions are made, and much of this influence happens outside conscious awareness (Frith, 2007).

Presumably, some of these influencing factors would be morally endorsed while others rejected had they come to attention.

Often concrete interests by individuals and groups exert pressure on the decision-making process. In the corporate organization setting, for example, there is a typically added complexity due to the high level of normative tension or moral intensity (Jones, 1991) represented by the many, partially conflicting interests various stakeholders associate with a pending decision. Stakeholders, moreover, will to varying degrees see themselves as a legitimate part of the decision-making process, sometimes with accompanying demands that their normative views or claims are respected, expressed, or acted on. Managing emotions therefore becomes an important issue in the organizational setting (Lines, 2005: 22–23).

Both cognitive and affective processes play a crucial role in moral decision-making. The *interaction* between cognition and affect is crucial in practical decision-making (Loewenstein and Lerner, 2003). As Bruner (1986: 113) noted, “there are some simple, probably biologically based linkages between emotion, arousal, drive on the one side, and learning, problem solving, thinking on the other.” Evidence from neuroscience supports this picture of extensive interaction (Panksepp and Biven, 2012; Panksepp, 1998).

6.3.1 The basic role of emotions

Emotions motivate, and this motivation is affective (Panksepp, 1998). As noted earlier (Ch. 2), the traditional view that cognition by itself could motivate (e.g., Lazarus, 1991) seems inadequate as a theory for explaining behavior. It is hard to see how “cold” cognition by itself could motivate (Blackburn, 1998; Izard, 1991); as Robinson (2005: 22) illustrates, “I can think about danger all day long without ever becoming afraid or acting to avoid it. And by the same token a thought of danger does not seem sufficient to produce any of the physiological symptoms of fear.”

In decision-making, the principal function of emotions appears to be to *motivate* decisions by infusing cognitive representations of the external world with internally based emotional content (Loewenstein and Lerner, 2003). This motivation is affective because, as we have seen in previous chapters, it originates in the affective structures of the brain, the most fundamental driver of motivation being is “the seeking system” (Panksepp, 1998: Ch. 8). Often concrete goals feature among the cognitive representations, and emotions attach to their anticipated realizations (Loewenstein and Lerner, 2003: 620), thereby intuitively or declaratively allowing the recommendation of appropriate courses of action.

As suggested by the hierarchical structure of affective consciousness (Panksepp, 2005), the motivation of emotions can occur either at an intuitive level below awareness or within the scope of explicit focus (Creswell et al., 2013). At a higher level of consciousness, which is typically engaged to some extent in decision-making, raw emotional content is fused with cognitive elements of thinking so as to serve as constituents of cognitively elaborated conglomerate emotions (i.e., tertiary emotions).

Conglomerate emotions, such as shame, guilt, and the prospect of regret (Loewenstein and Lerner, 2003), may be especially important guiding elements in relation to goal representation and the expected outcomes of action alternatives. As already shown (Ch. 4 and Ch. 5), emotions arise subcortically, most fundamentally as a basic affective consciousness arising from the upper brain stem brain structures of the brain, which is subsequently shaped by midbrain limbic structures. Emotions need to be connected with the external world, represented to us by our cortical functions (notably the homunculus and association cortexes), to serve their in-built practical purposes (Solms, 2013; Solms and Panksepp, 2012).

Despite the benign role of emotions, their involvement in decision-making is not uniformly conducive to quality; sometimes there is also an element of *interference*. Evidence

amply corroborates the fact that emotions *impede* as well as *promote* decision-making in complex social decision contexts (Davies and Turnbull, 2011). The active focus of practical decision-makers is a scarce resource because of its limited capacity (Berlin, 2011); consequently, it can easily be subverted or distracted by emotional experiences, especially strong emotions (Solms and Panksepp, 2012: 149).

Certain emotions are associated with specific types of decision-making problems or obstacles. *Negative* emotions in particular, such as fear, have the potential to cripple and distort human psychological life (Kitcher, 2011: 95). Based on experimental research, Lerner et al. (2013: 72) hypothesize that “grief” or “sadness increases impatience and creates a myopic focus on obtaining money immediately instead of later” and suggest that this “increases intertemporal discount rates and thereby produces substantial financial costs.” Another negative emotion, “anger,” is also held to be detrimental to decision-making (Lerner and Shonk, 2010), despite the fact that anger sometimes is heralded as a virtue in segments of professional ethics, notably in business. Positive emotions can also distort, for example, by leading to overconfidence (Kahneman, 2011) and framing biases (Kahneman, 2011: 364–367). Discrete emotions also affect the escalation of commitment to decisions (Dang et al., 2014), which in many cases can be considered detrimental.

The appropriate role of emotions in decision-making is a matter of *balance*. Ignoring or excessively suppressing emotions leads to inferior practical decision-making. This seems especially likely when decision-makers face ambiguity, uncertainty, risk, and complexity (Slovic et al., 2002), which are all identifiable characteristics of moral decisions. However, some level of suppression of emotion is necessary because emotions can be overpowering, which also leads to poor decision-making. On the basis of clinical research, Damasio (1994: Part 1) argues that *without* the involvement of emotions and feelings, decision alternatives are experienced as essentially arbitrary. Hence, in order to be appropriately socially responsive

and morally attuned, decision-makers in organizations need to engage emotionally within the social context (Salvador and Folger, 2009).

Moral lapses may often arise when emotions do not serve their role properly. As I argue below, there are many threats to normative-affective content in decision-making beyond clinical malfunction and psychological imbalances. One threat comes from uncritically or unknowingly assimilating and internalizing morally “unhealthy” or detrimental social norms (e.g., wastefulness, opportunism, racism, sexism, misogyny, homophobia). Presumably, most people hold some biases that, even in their own estimation, would improve their moral behavior and attitudes.

Another threat is represented by the acceptance of specific theories and ideas that become cognitive frames of mind. These threats notably include (1) pursuing an ideal or conception of morality that repudiates affective content from the decision-making process; somewhat similarly, (2) pursuing a “morality” consists of following rules, norms, and laws (Smart, 1973); and invalidly inferred positions, such as (3) extrapolating normative conclusions from evolutionary ethics (Farber, 1994). While these threats to some extent overlap, each of them represents a frame of mind that can dominate moral deliberation in ways that diminish, over-rule, or sideline affective content. Awareness of these threats and deliberate efforts directed at dealing with them seem likely to benefit decision-making in terms of expected outcomes.

6.3.2 Threats and opportunities

Making decisions, or choices, about what to do in situations with different and often conflicting values at stake for different stakeholders is an important and challenging part of everyday practical ethics (Wenstøp and Koppang, 2006). Consider for example the business profession: Money changes hands, shipments are made, programs are downsized, new

factories are opened, old ones are closed down, pollutants are spilled into the river, and greenhouse gasses are emitted into the atmosphere, etc. Change is the order of the day in business life, and practically every major decision has a social and moral dimension (Etzioni, 1988). Organizations are continuously faced with emerging threats and windows of opportunity.

To address these threats and opportunities, the decision-makers of a given organization can adopt strategies, plans, policies, and objectives, resulting in practical decisions and specific courses of action. In principal, it is an open question *whose* concerns, wishes, hopes and fears feature in the conceptualization of the threats and opportunities. It could be a threat to the sustained growth of the organization, but it could also, for example, be a threat to the global environment, which is at also issue. This contrast raises the question of what makes something a threat or an opportunity.

Affective neuroscience helps us address this question. In order to see something *as* a threat or an opportunity, deep-seated affective emotions are necessarily involved (Panksepp, 1998; see also Ch. 4), especially the subcortical *seeking system* for detecting opportunities and the *fear system* for sensing threats. Affective feelings are never neutral (Solms, 2013; MacLean, 1994: 110) but always either “agreeable” or “disagreeable” (MacLean, 1994) and thereby “rewarding” or “punishing” (Solms and Panksepp, 2012: 156). In addition to affect, deliberation also involves cognition, notably in conceptualization, the representation of objects, and projecting causal implications of action (Solms, 2013); however, if no affective emotions attach to these cognitions, they will not by themselves be able represent anything as a threat or opportunity.

Since perceived threats instigate negatively valenced fear responses and opportunities seeking urges that are experienced as positively valenced, threats will be “punishing” or “bad”

and opportunities “rewarding” or “good.” This shows a central role that emotions play in practical decision-making. Emotions are crucial for determining threats and opportunities, and they do so by furnishing the subjective normative outlook enabled by affective consciousness.

This perspective also helps shed light on some of the biases that emotions represent in decision-making. For example, it has been argued (Huang and Luo, 2006; Rozin and Royzman, 2001) that there is a considerable and persistent negativity bias in human and animal decision-making; “sticks” are more consequential than “carrots”; or “losses loom greater than gains” (Tversky and Kahneman, 1986; 1981). This concept can be explained by examining the biological evolution of emotions (Panksepp and Biven, 2012); “carrots” are nice to come by, but if you get “hit” hard enough by a “stick,” there will be no more “carrots” forever. There is a biological reason for why the negative emotions represented by threats strike us as more worthy and important.

6.3.3 Emotion regulation

One of the delicate issues involved with looking at emotions in relation to decision-making is the problem of how to understand the affective-cognitive tension between being controlled by emotions and controlling emotions (Damasio, 2010), which is related to the sense in which an agent has a self that exercises genuine authority. Emotion regulation is important in decision-making because, as noted, emotions can sometimes be disruptive.

According to Koole (2009), one of the primary functions of emotional regulation is to enable goal-directed behavior. Goals have to be fixed in the mind by higher cortical processes in order to be extended in time. Therefore, emotions that threaten this fixation must be actively suppressed or repressed subconsciously. Goals themselves are not emotionally neutral since, as noted, they consist of positive anticipatory affects projected onto cognitive representations. Because goals are taken as important by the “extended self” (in the

terminology of Damasio, 1999), which reflects Damasio's set of concerns and values, they also need to be protected.

The hierarchical conception of consciousness proposed by Panksepp (2005) suggests that the self should be understood as involving cognitive as well as affective aspects. Affective and cognitive processes form a nested hierarchic structure in which affect is primary and grounds cognition (see Ch. 5). This setup suggests that affect is an integral part of the authoritative self. Emotion regulation, therefore, it is a matter of self-regulation; the emotional self is at a higher level in the hierarchy, where more cognitive information and resources are available, regulating the emotional self at its lower levels. This is an important point because it is easy to conflate "self" merely with the self we are consciously aware of, but that would be a mistake because our experienced self has much deeper roots than what enters awareness (Panksepp, 2005). Hence, feelings (the experiences of emotions) are likely to play a central role in emotion regulation (Hoeksma et al., 2004: 354).

It can be useful to look more closely at the interplay between affect and cognition in decision-making; in particular, affective emotions as conative forces and the cognitive processes as mechanisms that regulate them. As proposed in the previous chapter, affect is the very fount of consciousness. This idea is consistent with the Freudian conjecture that there is no such thing as an unconscious affect (Solms and Nersessian, 1999: Freud, 1915: 178). Affects correspond to our feelings, including our basic feelings of being alive. However, affects can and would be overwhelming and misdirected without the regulation, suppression, and direction of our senses and our conscious and unconscious cognitive processes. The role that higher brain structures play in regulating emotions is essential for coping in the world and dealing with the practical problems it poses. At the same time, although emotion regulation serves a crucial function, it can also turn into a threat if it becomes too strong and dominating;

it could then stifle the important flow affect that decision-makers needs as a normative compass for navigating.

Koole (2009: 6) proposes emotion regulation can be “defined as the set of processes whereby people seek to redirect the spontaneous flow of their emotions.” While this definition captures the central controlling aspect of emotions, it seems somewhat biased toward conscious awareness and what agents actively do. In contrast, others have suggested that emotion regulation is largely automatic and unconscious (Berlin, 2011; Williams et al., 2009). This idea suggests that the scope practical decision-makers have for emotion regulation may be limited. Perhaps the best decision-makers can do is *not* to place too much effort on trying to regulate their emotions and let the cognitive unconscious take care of this for them, so that they can free resources to focus on other things.

6.3.4 Emotions and reasoning

Decision-making is a complex psychobiological process that takes place within an individual decision-maker in which both unconscious (implicit) and conscious (explicit) processes play their part. Decision-making is of course much more than this because explicit lines of reasoning can be communicated and shared *socially* within a group of decision-makers that is involved in a decision-making process and with other stakeholders of a decision, as is typical, for instance, in the organizational and professional decision-making contexts. Any decision, however, may arise out of the mental processes occurring within the head of the individual decision-maker. These internal processes distinguish decision-making from things that are merely happening.

A subset of mental processes that is of special importance in decision-making is *reasoning processes*. Formal or “theoretical” reasoning is a topic that has been treated extensively by moral philosophy and epistemology. I will not present this vast literature here

since my focus is on the role of emotions in reasoning in relation to specifically practical deliberation. Hence, I focus on *practical* reasoning. Theories of formal reasoning often form a part of more general theories or rationality. One of their primary suggestions is that reasoning can be checked against objective *quality standards* or *requirements*, such as the requirement of consistency. Since I see no objection to this proposal, I will assume that, in theory, reasoning *can* be checked in this manner. However, there is an important distinction between things that can be done *in theory* and what decision-makers can do *in practice* (Harman, 1986).

Let us grant that, given adequate time and effort, the rough structure of important decisions can be checked for consistencies of belief sets and teleological consistencies linking the primary goal of the decision with important sub-goals that serve as a means to achieve the primary goal. Still, we should not be overly optimistic about the practical merits of the rationalistic approach. Below are a few points for illustration:

(1) There will always be *practical limits* to how far one can go in the direction of checking for consistencies (Harman, 1986). (2) As a practical matter, one would have to start checking for consistencies *somewhere*, starting with, for example, beliefs that are present in the awareness of focused attention; this may well be an *arbitrary* starting point. (3) Since beliefs represent a virtually endless map to be charted, it will in practice be *impossible* to obtain anywhere near a *complete set* of beliefs that bear on a given decision, and there will be further complications when many decision-makers are involved. (4) We are only consciously aware of a small subset of our actual set of beliefs; some have been actively suppressed and rendered unconscious, while others have been repressed automatically (Berlin, 2011).

The point is that decision-makers have to work from within their limited attention span and with the limited cognitive resources at their disposal. At the same time, it is clear that

explicit reasoning to some extent serves an important role in decision-making. Besides its direct decision-making relevance, explicit reasoning also allows decision-makers to prepare arguments for their decisions (or against alternative options), sometimes as a means to “justify” their decisions either before or after the fact. How does reasoning relate to affective human nature?

The first thing to note is that there may be many kinds of reasoning (Millgram, 2001: Ch. 1). In moral philosophy, a division between two types of reasoning stand out; reasoning about facts and practical reasoning regarding what to do (Goldman, 2009; Millgram, 2001: 1). Reasoning about what to do is typically concerned with what is factual *and* about the structure of one’s practically oriented desires (Goldman, 2009). However, reasoning in terms of narratives, as a specific type of reasoning, can often be highly fictional with the primary function of revealing something true about what one feels and desires rather than about the external world (Goldie, 2012). Ultimately, all reasoning rests on the foundational underpinnings of affective consciousness since, as noted in the previous chapter, it serves as the premises of our sense of “being in the world.” However, it is not clear whether all reasoning processes require process-internal affective content in a direct manner, although they are premised on lower brain systems and corresponding affective states of mind.

It may be the case that reasoning that involves *beliefs* about facts, for example, requires only *cognitive* content. And although there may be affective content attached the cognitive representations invoked in the reasoning process, this may be irrelevant for the efficacy of the reasoning. Thus, reasoning from one state of belief to another state of belief may not *per se* seem to depend on the affective content. For example, the belief (1) that it is raining, combined with a second belief (2) that if it is raining, the snow will melt, can lead to the third belief (3) that the snow is melting. This picture of reasoning is shared by many philosophers (e.g., Broome, 2013; Blackburn, 2010; Goldman, 2009). The crucial question is

how far this can get us in *moral* reasoning or *practical* reasoning regarding what action to take.

For this question, there is an important split in philosophical opinion. While Broome (2013) and many other philosophers that can be said to be *externalists* about moral motivation are optimistic about what cognitive reasoning involving beliefs can do, *internalists* about motivation, such as Blackburn (2010), Goldman (2009), and Williams (1985), hold that internal affective emotions are necessary for arriving at a decision. From an internalist view, *affective content* is essential. I discuss motivational internalism in more detail later (Ch. 9). Based on the arguments made on the basis of affective neuroscience (Ch. 5), I believe that motivational internalism offers the only plausible explanation of human moral motivation (the alternative view is treated in Ch. 8). I propose that we are motivated in a fundamental way by our affective emotions, and these emotions guide us through their *implicit normativity*, giving us, as Panksepp (2010) aptly phrases it, “intention-in-action.” Intention-in-action, on this view, is prior to and serves as a fundament of any “intention-for-action.” Similarly, Modell (2003: 102) proposes that the human brain is equipped with “unconscious intentionality” which “both selects and deselects perceptual items.”

By focusing on the place of affective content in moral reasoning, we can transform some of the implicit normativity into *explicit normativity*. Explicit normativity can then ground what we can refer to as our intentions *for* action and decisions. While implicit normativity is essentially pre-cognitive and pre-conceptual, explicit normativity is cognitively represented in our mind and can thereby be conceptualized. Explicit normativity can then be deliberately associated with evaluative notions relevant to decision-making, such as values, plans, and goals, as well as more general ethical terms (e.g., virtues, vices, needs, rights, duties).

As noted, explicit reasoning has many advantages, notably including the ability to explicitly communicate, share, and discuss our reasons and reasoning. It also allows decision-makers to seek to justify and explain their decisions. Although undoubtedly many of the affective motives that drive decision-makers will remain in the dark, explicit reasoning involving explicit normativity can at least bring *some* of the motivational basis out in the open.

6.3.5 Emotions and rationality

Given that emotions play an important role in moral decision-making, it is “a fundamental error to regard every surge of emotion against judgment as an uprising of the irrational” (Frankfurt, 1988: 189). But can emotions be *rational*? The claim that they *can* is fairly widespread across philosophy and psychology, as well as the natural sciences (Riotti, 2009: 39). The issue depends, of course, on what precisely one means by the appellation “rationality.” Increasingly traditional theories centering on the notion of rationality are questioned (e.g., Toulmin, 2001). From a neuroscience perspective, Pankespp (1998: 301) talks of “the rational fallacy” are associated with the illusion of being in conscious cognitive control. This way of thinking suggests that the concept of rationality may fail to correspond to its intended mental phenomena.

From a psychological and an organizational perspective, Putnam and Mumby (1993: 36-57) argue that there is a pervasive and deeply entrenched “myth of rationality” in Western culture and institutions. Contemporary philosophy, moreover, suggests that the issues concerning how the term is used and how it should be used are far from settled (Broome, 2013). Finally, it is unclear whether the term “rationality” is in fact required for a coherent theory of mind or for an adequate normative theory of decision-making.

Since the focus is on the role of emotions in decision-making, I will not review the extensive debates about rationality specifically. However, it can be noted that many of the current conceptualizations of rationality differ considerably from each other and are often mutually incompatible. For example, while some hold that rationality consists of a set of normative requirements that go *into* reasoning, others (e.g., Broome, 2013) suggest that rationality is something that comes *out of* reasoning. The latter perspective interestingly suggests that many of the terms that “rationality” has been invested with—for example, “coherence” and “consistency”—can usefully be entangled from the concept of “rationality.” In this case, we can discuss standards, such as “correct” and “mistaken” reasoning, in terms of such qualifiers as “consistency,” *without* drawing on the notion of rationality.

The latter way of thinking about rationality seems to render the question of whether emotions are rational superfluous. After all, if rationality is something we *arrive* at through reasoning, it is not the *rationality* of beliefs and cognitions or values and emotions which is in question but rather their consistency, coherences, and so on. So, for example, we could be in a position to deem a given set of emotional feelings incoherent or state that these feelings pull us in different directions without thereby declaring them “irrational.” Rationality, from this perspective, becomes something like an honorary title, and little else.

Even in the former way of thinking about rationality, however, it seems far from straightforward to talk of emotions in terms of being “rational” or “irrational.” As some philosophers have noticed, the difficulty of characterizing emotions in this manner is connected with the question of whether we can *choose* emotions (Blackburn, 1998: 161–199; Griffiths, 1997; Gibbard, 1990) or choose our way *out of* emotion, which I have already touched upon in the section above; “It is natural to think, and it may even be true, that self-directed psychological states simply have no power to generate emotions towards others, that our emotional life is not under that sort of control” Kitcher (2011: 26). According to Gibbard

(1990: 38), it “may well seem that we can appraise as rational or irrational only what is under a person’s voluntary control. Emotions fail this test, since they are not under a person’s direct voluntary control”; he continues: “What can be appraised as rational or irrational is not an emotion itself, but measures can be taken to nurture or repress it.”

Ultimately, however, the question of whether we can choose emotions is an empirical one that scientific pursuits, such as those of neuroscience, must guide us in finding answers to. The answer partly depends on how the intricate issues relating to emotion regulation (discussed above) are to be resolved. One difficulty here is that emotions seem layered in the brain structures in a manner that makes it difficult to isolate their influence. Another problem is that affective emotions may well turn out to be ubiquitous to consciousness, in which case choosing independently of them becomes impossible.

Cumulative empirical evidence (Frith, 2007; Panksepp, 1998: Ch. 16) suggests that the default answer is negative—that we are generally unable to choose emotions—and that affirmative answers, if at all appropriate, should be viewed as exceptions to this rule. According to Frith (2007; see Ch. 3 above), the traditional sense of rational free choice is at least to a considerable extent illusory. In relation to the remaining element of free choice, supposing there is one, the nature of emotions makes them unlikely candidates as objects of choice in any direct sense, though as mentioned we can make an effort to manage them and sometimes succeed at that. After all, primary emotions, as they arise subcortically, are entirely outside the scope of volition, deliberation, and conscious awareness (Panksepp, 1998; see also Ch. 3). Choice requires focus and conscious awareness, and this is only epistemologically available once emotions reach the higher cortical areas of the brain. However, even cognitive-affective conglomerate emotions seem to escape direct choice. Consider for example simply choosing not to be jealous when in fact you are; I presume most will agree that this does not seem like an easy prospect.

Even the active regulation of emotions and suppression involved in pushing away emotions is likely to be rather limited because, as noted, it is costly in terms of mental energy and the cognitive capacity it occupies. Most of the inappropriate or ill-suited emotional inputs arising from below are regulated and repressed without our conscious involvement (Berlin, 2011). Moreover, some argue that in a fundamental sense, very much of our mental life is controlled from below by our subcortical and limbic brain structures. For example, Solms (2013: 106) argues that “the deep structures that generate consciousness are not only responsible for the *level* (quantity) but also for a core *quality* of consciousness” (Solms, 2013: 106).

Whatever rationality is on a presence definition, assuming that there is a cogent and applicable concept, it seems inescapable that it has to involve consciousness. Hence, the most reasonable position seems to be that, psychologically, emotions are at least mostly *beyond* questions of rationality and irrationality, as some philosophers have plausibly suggested to us (e.g., Blackburn, 1998; Gibbard, 1990; Gauthier, 1986: 327).

6.3.6 A faculty of reason?

In moral philosophy, and in the applied ethics literature (Noland and Phillips, 2010), it is sometimes assumed that human beings are equipped with a “faculty of reason”; often this is referred to as a “faculty of rationality” or simply “reason.” There is a longstanding tradition for making this assumption, such as that visible in the many modern followers of Kant, Descartes, and Plato. In most of these classical accounts, “reason” or “rationality” is conflated with “moral,” and thus the idea has often been presented equivalently as the claim that there exists a specific “moral faculty.” Thus, for example, Reid (1827: 667) proclaimed: “When I exercise my moral faculty about my own actions or those of other men, I am conscious that I

judge as well as feel. I accuse and excuse, I acquit and condemn, I assent and dissent. I believe and disbelieve, and doubt. These are acts of judgment and not of feelings.”

The assumption is typically linked to a sharp distinction and dissociation of emotion on the one hand and cognition on the other (e.g., Parfit, 2011a; 2011b). Neuroscientists, meanwhile, almost uniformly consider the possibility of identifying one unified neural substrate for such a “faculty” to be an outdated quest. The processes that are involved in reasoning involve multiple brain systems with different functions and locations (Levine and Perlovsky, 2010). Philosophically, it is also difficult to cater to the idea of an independent faculty of reason, since the mind supervenes on the brain, and the brain is embodied, while our bodies are socially embedded (Johnson, 2014: 24; see also Ch. 3).

It could be speculated that the faculty of reason resides in the (neo-cortical) *dorsolateral prefrontal cortex*. This slow-maturing brain region is (among other) centrally involved in executive functions, working memory, temporal planning, abstract reasoning, (Miller and Cummings, 2007: 355; LeDoux, 2002), and description-based appraisal (Levine and Perlovsky, 2010). If there were any one dedicated brain region, it would possibly be the most promising candidate. A second candidate would be the *ventral-medial prefrontal cortex*, which is anatomically just below.

Whereas many philosophers still assume that there is a faculty of reason, evidence from affective neuroscience shows that this assumption is implausible. As already argued (Ch. 5), the very consciousness of thought and cognition stems from subcortical regions of the brain, and the agent—ultimately the agent of self-conscious rational thought—emerges in stepwise fashion from below (Damasio, 2010: 10). The notion of a “self” or an “agent” that is autonomous and separated from the rest is in this picture untenable (Ch. 3 and Ch. 5).

Moreover, well before the sensory input arrives in the neo-cortical areas of the brain, the information is already organized by emotional midbrain structures (Panksepp et al., 2012).

The information therefore arrives in the higher brain structures with emotional significance attached. The strict dissociation of emotion and cognition in the neo-cortex (prefrontal cortex) is therefore to some extent artificial. The distinction between emotion and cognition remains essential for understanding what goes on in moral reasoning, insofar as “ethical decision-making appears to be distinct from other types of cognitive decision processes” (Salvador and Folger, 2009: 5). In essence, this difference can be attributed to the role of affective emotions that are fed into the reasoning processes by subcortical emotional systems. This affords little room for an independent faculty of reason or rationality.

6.3.7 Affective content and rational structure

In reasoning and logical inferences, there is plainly such a thing as being wrong (Broome, 2013; 1999; Goldman, 2009: Ch. 2): not wrong in the moral sense but factually. For example, there is something wrong about holding inconsistent beliefs; one cannot very well believe that it is raining and not raining at the same time. There also seems to be a mistake (of rationality) involved when one fails to update one's beliefs so that it corresponds to evidence, at least if these beliefs have practical relevance. If you are outside, and it suddenly starts pouring down, there is something wrong if your beliefs about the weather are not correspondingly updated. The case of beliefs contrasts rather sharply with the case for affects (desires, emotions, values).

Affects can be viewed as content providers for reasoning that attaches to specific information to give it relevance. Information in itself cannot tell decision-makers what to do or decide; first there has to be some goal, aim, or desires present to offer direction. Blackburn (1998) uses the analogy of a map; beliefs provide the means of representing the terrain, while

affective values are the compass that gives direction. Rationality sets certain structural requirements that beliefs have to conform to in order to be reliable; ideally, we want our beliefs to be true. Affects—within or outside our awareness as they may be—identify, reveal, and express what we are concerned with.

From the vantage point of therapeutic psychology, Greenberg (1990: 187) argues that to be “unaware of our primary emotions is to have less information available and to tie up one’s processing resources in internal activity designed to block that very information.” Explicit affective content that is consciously retrievable is accordingly an important decision-making resource. We can then hope to gain better cognitive control with the flow of raw emotional input and regulate it to a greater extent; “The greater people’s ability to access and become aware of primary emotions, the greater their capacity to solve life problems adaptively” (Greenberg, 1990: 188). However, it is not a fault attributable to rationality if we are unaware of the primary emotional affects that exercise control over us. It is more like a biological shortcoming.

There are, however, measures practical decision-makers can take to improve their decision-making given an awareness of their own shortcomings. Generally, anything that reveals emotional affects is likely to be helpful. One concrete way to identify emotional affects is to elicit emotions (Gross and Levenson, 1995). This can be done, for example, by visualizing the consequences of the relevant decision alternatives there are to choose from. If visualizing is difficult cognitively, additional measures can be sought. For example, one might contract an artist to assist directly in making a realistic representation of the different scenarios.

It can be noted that the approach just sketched is concerned with *revealing* or *discovering* affects, not about attempting to decide what they should be. If it were possible to

somehow determine affective content directly and fundamentally, affective content would be subject to rationality. Whatever hidden affects there might initially have been would then be of less importance. But, as argued, affects are not objects of choice in this way.

6.4 Conclusion

Moral decisions require affective content in order to avoid arbitrariness of choice. The emotional infrastructure of the brain supplies this affective content. Prior to arriving in the cortical decision-making structures of the brain, affective content conveys only implicit normativity. Implicit normativity attaches to all our mental images and representations, investing them with subjective meaning. Among others, this process allows us to see various mental objects and prospective actions as threats and opportunities. As affective content enters awareness, we gain explicit normativity about which we can deliberate.

Affective emotions need to be regulated in order to prevent rash and inappropriate forms of behavior. Most of this regulation happens unconsciously without our awareness and is also effortless. Sometimes we need to exert a conscious effort to regulate our affective emotions, but this effort also relies on our subcortical affective brain structures. When active regulation of affect is too broad and too strong, especially in combination with cognitive frames of mind that seek to extinguish affect, there may be detrimental effects on decision-making.

Chapter 6 References

- Ardila, A. (2008) On the evolutionary origins of executive functions. *Brain and Cognition* **68**: 92–99.
- Bargh, J. and Morsella, E. (2008) The unconscious mind. *Perspectives on Psychological Science* **3** (1): 73–79.
- Bechara, A. (2004) The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition* **55**: 30–40.
- Berlin, H. (2011) The neural basis of the dynamic unconscious. *Neuropsychoanalysis* **13** (1): 5–31.
- Blackburn, S. (1998) *Ruling passions: A theory of practical reasoning*. Oxford: Oxford University Press.
- Broome, J. (2013) *Rationality through reasoning*. Oxford: Wiley-Blackwell.
- Broome, J. (1999) *Ethics out of economics*. Cambridge: Cambridge University Press.
- Bruner, J. S. (1986) *Actual minds, possible worlds*. Cambridge: Harvard University Press.
- Ciaramelli, E. and di Pellegrino, G. (2011) Ventromedial prefrontal cortex and the future of morality. *Emotion Review* **3** (3): 308–309.
- Creswell, J. D., Bursley, J. K., and Satpute, A. B. (2013) Neural reactivation links unconscious thought to decision-making performance. *Social Cognitive and Affective Neuroscience* **8** (8): 863–869.
- Damasio, A. (2010) *Self comes to mind: Constructing the conscious brain*. New York: Pantheon Books.

- Damasio, A. (1999) *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt.
- Damasio, A. (1994) *Descartes's error: Emotion, reason, and the human brain*. New York: Avon Books.
- Dang, J., Xiao, S., and Liljedahl, S. (2014) Escalate shamefully, de-escalate angrily or gratefully: The influence of discrete emotions on escalation of commitment. *Scandinavian Journal of Psychology* **55** (4): 380–388.
- Davies, J. L. and Turnbull, O. H. (2011) Affective bias in complex decision making: Modulating sensitivity to aversive feedback. *Motivation and Emotion* **35** (2): 235–248.
- Davison, R. J. (2003) Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology* **40**: 655–665.
- Di Chiara G, Bassareo, V., Fenu, S., De Luca, M. A., Spina, L., and Cadoni, C. (2004) Dopamine and drug addiction: The nucleus accumbens shell connection. *Neuropharmacology* **47** (1): 227–241.
- Etzioni, A. (1988) *The moral dimension: Toward a new economics*. New York: The Free Press.
- Farber, P. L. (1994) *The temptations of evolutionary ethics*. Berkeley: The Regents of the University of California.
- Frankfurt, H. G. (1988) *The importance of what we care about: Philosophical essays*. Cambridge: Cambridge University Press.
- Freud, S. (1915/2013) The unconscious. In S. Akhtar and M. K. O'Neil (Eds.), *Contemporary Freud: Turning points & critical issues* (pp. 21–78). London: Karnac Books Ltd.

- Fine, C. (2008) *A mind of its own: How your brain distorts and deceives*. New York: W. W. Norton & Company.
- Frith, C. (2007) *Making up the mind: How the brain creates our mental world*. Oxford: Blackwell Publishers.
- Gerrans, P. and Sander D. (2014) Feeling the future: prospects for a theory of implicit prospection. *Biology and Philosophy* **29** (5): 699-710.
- Gibbard, A. (1990) *Wise choices, apt feelings*. Cambridge: Cambridge University Press.
- Goldie, P. (2012) *The mess inside: Narrative, emotion, and the mind*. Oxford: Oxford University Press.
- Goldman, A. H. (2009) *Reasons from within: Desires and values*. Oxford: Oxford University Press.
- Greenberg, L. S. (1990) *Emotions in psychotherapy*. New York: The Guilford Press.
- Gross, J. J. and Levenson, R. W. (1995) Emotion elicitation using films. *Cognition and Emotion* **9**: 87–108.
- Griffiths, P. E. (1997) *What emotions really are: The problem of psychological categories*. Chicago: University of Chicago Press.
- Gutmann, A. K., Jacobi, B., Butcher, M. T., and Bertram, J. E. A. (2006) Constrained optimization in human running. *Journal of Experimental Biology* **209**: 622–632.
- Harman, G. (1986) *Change in view: Principles of reasoning*. Cambridge: MIT Press.

- Hoeksma, J. B., Oosterlaan, J., and Schipper, E. M. (2004) Emotion regulation and the dynamics of feelings: A conceptual and methodological framework. *Child Development* **75** (2): 354–360.
- Huang, Y. and Luo, Y. (2006) Temporal course of emotional negativity bias: An ERP study. *Neuroscience Letters* **398** (1–2): 91–96.
- Izard, C. E. (1991) *The psychology of emotions*. New York: Plenum Press.
- Johnson, M. (2014) *Morality for humans: Ethical understanding from the perspective of cognitive science*. Chicago: Chicago University Press.
- Jones, T. M. (1991) Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review* **16** (2): 366–395.
- Kahneman, D. (2011) *Thinking, fast and slow*. New York: Farrar, Straus, and Giroux.
- Kitcher, P. (2011) *The ethical project*. Cambridge: Harvard University Press.
- Kiefer, M. (2012) Executive control over unconscious cognition: Attentional sensitization of unconscious information processing. *Frontiers in Human Neuroscience* **6**: 61.
- Koole, S. L. (2009) The psychology of emotion regulation: An integrative review. *Cognition and Emotion* **23**: 4–41.
- Lawton, R., Conner, M., and McEachan, R. (2009) Desire or reason: Predicting health behaviors from affective and cognitive attitudes. *Health Psychology* **28** (1): 56–65.
- Lazarus, R. S. (1991) *Emotion and adaptation*. New York: Oxford University Press.
- LeDoux, J. (2002) *Synaptic self: How our brains become who we are*. Middlesex: Penguin Books.

- LeDoux, J. (1996) *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon and Schuster.
- Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. (2015) Emotion and decision making. *Annual Review of Psychology* **66**: 799–823.
- Lerner, J. S., Li, Y., and Weber, E. U. (2013) The financial costs of sadness. *Psychological Science* **24** (1): 72–79.
- Lerner, J. S. and Shonk, K. (2010) How anger poisons decision making. *Harvard Business Review* **88** (9): 26.
- Levine, D. S. and Perlovsky, L. I. (2010) Emotion in the pursuit of understanding. *International Journal of Synthetic Emotions* **1** (2): 1–11.
- Lines, R. (2005) The structure and function of attitudes toward organizational change. *Human Resource Development Review* **4**: 8–31.
- Litt, A., Eliasmith, C., and Thagard, P. (2008) Neural affective decision theory: Choices, brains, and emotions. *Cognitive Systems Research* **9**: 252–273.
- Loewenstein, G. and Lerner, J. S. (2003) The role of affect in decision making. In R. Davidson, H. Goldsmith, and K. Scherer (Eds.), *Handbook of affective science* (pp. 619–642). Oxford: Oxford University Press.
- Luria, A. R. (1980) *Higher cortical functions in man* (2nd ed.). New York: Basic Books.
- Luria, A. R. (1973) *The working brain: An introduction to neuropsychology* (B. Haigh, Trans.). New York: Basic Books.
- MacLean, P. (1994) Human nature: Duality or triality? *Politics and the Life Sciences* **13** (1): 107–112.

- Miller, B. L. and Cummings, J. L. (2007) *The human frontal lobes: Functions and disorders*. New York: The Guilford Press.
- Millgram, E. (Ed.) (2001) *Varieties of practical reasoning* (pp. 1–26). Cambridge: The MIT Press.
- Modell, A. H. (2003) *Imagination and the meaningful brain*. Cambridge: The MIT Press.
- Moretto, G., Làdavas, E., Mattioli, F., and di Pellegrino, G. (2010) A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience* **22** (8): 1888–1899.
- Noland, J. and Phillips, R. (2010) Stakeholder engagement, discourse ethics and strategic management. *International Journal of Management Reviews* **12** (1) 39–49.
- Panksepp, J., Asma, S., Curran, G., Gabriel, R., and Greif, T. (2012) The philosophical implications of affective neuroscience. *Journal of Consciousness Studies* **19** (3–4): 6–48.
- Panksepp, J. (2010) Affective neuroscience of the emotional BrainMind: Evolutionary perspectives and implications for understanding depression. *Dialogues Clinical Neuroscience* **12** (4): 533–545.
- Panksepp, J. (2005) Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition* **14**: 30–80.
- Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.
- Parfit, D. (2011a) *On what matters: Volume 1*. Oxford: Oxford University Press.
- Parfit, D. (2011b) *On what matters: Volume 2*. Oxford: Oxford University Press.

- Putnam, L. L. and Mumby, D. K. (1993) Organizations, emotion and the myth of rationality. In S. Fineman (Ed.), *Emotion in organizations* (pp. 36–57). Thousand Oaks: Sage Publications.
- Reid, T. (1827) *Essays on the powers of the human mind: An essay on quantity. An analysis of Aristotle's logic*. London: Thomas Tegg.
- Rietti, S. (2009) Rationalities of emotions: Defending, distinguishing, connecting. *Organon F* **16** (1): 38–61.
- Roberts, A. C., Robbins, T. W., and Weiskrantz, L. (2002) *The prefrontal cortex: Executive and cognitive functions*. Oxford: Oxford University Press.
- Robinson, J. (2005) *Deeper than reason: Emotion and its role in literature, music, and art*. Oxford: Oxford University Press.
- Rozin, P. and Royzman, E. B. (2001) Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* **5** (4): 296–320.
- Salvador, R. and Fogler, R. G. (2009) Business ethics and the brain. *Business Ethics Quarterly* **19** (1): 1–31.
- Schultz, W. (2002) Getting formal with dopamine and reward. *Neuron* **36**: 241–263.
- Sellitto, M., Ciaramelli, E., and di Pellegrino, G. (2011) The neurobiology of intertemporal choice: Insight from imaging and lesion studies. *Reviews in the Neurosciences* **22** (5): 565–574.
- Slovic, P., Finucane, M., Peters, E., and MacGregor, D. G. (2002) The affect heuristic. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–400). New York: Cambridge University Press.

- Smart, J. J. C. (1973) An outline of a system of utilitarian ethics. In J. J. C. Smart and B. Williams (Eds.), *Utilitarianism: For and against*. Cambridge: Cambridge University Press.
- Solms, M. (2013) The conscious id. *Neuropsychanalysis* **15** (1): 5–19.
- Solms, M. and Panksepp, J. (2012) The “id” knows more than the “ego” admits: Neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Sciences* **2**: 147–175.
- Solms, M. and Turnbull O. (2002) *The brain and the inner world*. New York: Other Press.
- Solms, M. and Nersessian, E. (1999) Freud’s theory of affect: Questions for neuroscience. *Neuro-psychoanalysis* **1** (1): 5–14.
- Thagard, P. (2007) The moral psychology of conflicts of interest: Insights from affective neuroscience. *Journal of Applied Philosophy* **24** (4): 367–380.
- Toulmin, S. E. (2001) *Return to reason*. Boston: Harvard University Press.
- Tversky, A. and Kahneman, D. (1986) Rational choice and the framing of decisions. *The Journal of Business* **59** (4): 251–278.
- Tversky, A. and Kahneman, D. (1981) The framing of decisions and the psychology of choice. *Science* **211** (4481): 453–458.
- Wenstøp, F. and Koppang, H. (2006) On operations research and value conflicts. *Omega* **37** (6): 1109–1120.
- Williams, B. (1985) *Ethics and the limits of philosophy*. London: Routledge.

Williams, L. E., Bargh, J. A., Nocera, C. C., and Gray, J. R. (2009) The unconscious regulation of emotion: Nonconscious reappraisal goals modulate emotional reactivity. *Emotion* **9** (6): 847–854.

Wright J. S. and Panksepp, J. (2012) An evolutionary framework to understand foraging, wanting, and desire: The neuropsychology of the seeking system. *Neuropsychoanalysis* 14 (1): 5–39.

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

Chapter 7: The dual relativist position

7.1 Introduction to the dual relativist position

Moral relativism, as will be shown, includes several threats to the establishment of the normative foundations of practical moral decision-making and has often been treated as a unified general position to be avoided in order to safeguard sound normative foundations (e.g., Bowie, 2013: 79–88). In order to determine what the various threats consist of, however, it is necessary to conceptually *unpack* moral relativism (Quintelier and Fessler, 2012). By examining different patterns of reasoning, we can see that moral relativism does not after all denote one single unified perspective. Instead, there are several quite different, often mutually incompatible, positions. Philosophical analysis is a helpful tool in this unpacking.

Moral relativism, in its many forms, features as one of the most prominent areas of debate in contemporary philosophy (Quintelier and Fessler, 2012; Harman and Thomson, 1996). While it needs to be stressed that this area does not belong to the philosophical discipline alone, considerable philosophical effort has gone into elucidating it. In order to distill the important insights philosophical analysis offers on the topic, it is necessary to distinguish the different forms of moral relativism. To do this, I propose the following taxonomy of relativism that exposes the most important theoretical options.

The first important distinction is between *meta-ethical relativism* and *normative relativism*, which is why I call this chapter *the dual relativist position*. In this chapter, I *defend* meta-ethical relativism and suggest that we should *discard* different versions of normative relativism. Meta-ethical relativism and the scientific evidence with which it connects, show that we are equipped with subjective personal values. Normative relativism, I argue, is a threat to morality based on personal values.

I distinguish four analytically separate versions of normative relativism: normative cynicism, normative syncretism, normative nihilism, and normative collectivism. These positions represent theoretical propositions for relativistic normative conclusions to be arrived at by reasoning from a meta-ethical starting point. Hence, meta-ethics clarifies the status of values, whereas normative ethics aims to specify the practical-normative implications this would have. To draw such normative conclusions is problematic, however, due to essentially invalid chains of reasoning.

I have already pointed out that there seems to be no straightforward way to straddle Hume's gap between realms of descriptive "is" and normative "ought" (Ch. 1). Something more has to be *added* to generate normative conclusions from non-normative meta-ethics. This point has practical relevance for decision-makers; it affects the scope what they are entitled to in their practical reasoning. I will show that each of the normative positions needs to be rejected because of circular chains of reasoning (internal inconsistencies) and conclusions drawn at an arbitrary stage in the chain of reasoning.

I argue that meta-ethical relativism forms the core part of explaining moral thought, talk, and action (in normally functioning adult human beings). This position holds that psychological morality is relative to people. This claim follows from the biological basis established in foregoing chapters, centering on the role of affective emotions and their input into moral deliberation. While I defend meta-ethical relativism, I do not mean to suggest that other meta-ethical positions are unimportant. Indeed, I will show that meta-ethical naturalism and meta-ethical cognitivist error theory play important roles as well. Meta-ethical relativism, however, is a central component of the emotivist position, which is presented in detail later (Ch. 9). In this respect, the current chapter does some of the philosophical heavy lifting for the emotivist position.

Cognitivist error theory is accepted as an explanation of the status of beliefs in “moral truth” and “moral objectivity” that some people hold; it points to the error involved in such views. However, I reject cognitive error theory as a general explanation of normative thought, talk, and action because application of error theory would involve a category mistake. Finally, I review meta-ethical views that pose as rivals (meta-ethical naturalism, meta-ethical cognitivism, and meta-ethical nihilism) to meta-ethical relativism and argue that a particular minimal version of meta-ethical naturalism is in fact not a rival of meta-ethical relativism but rather a necessary component of it.

7.2 A taxonomy of relativism

Morality can usefully be analyzed in terms of values (Bishop, 2000), where “values” serve as a general placeholder for other normative concepts (“duties,” “rights,” “deserts,” “desires,” etc.). The *moral status* ascribed to values defines the fundamental differences in normative ethical thinking. I propose a taxonomy of relativism that includes four distinct versions of *normative relativism* with their accompanying *meta-ethical* foundations. While the proposed terminology fits with the general terminology commonly used in moral philosophy (e.g., Prinz, 2007; Wong, 2006; Blackburn, 2005), more specific terms are needed in order to present the meta-ethical and normative positions in detail.

When we unpack “moral relativism,” we find three main theoretical components (Quintelier and Fessler, 2012: 96): (1) *descriptive* perspective of normativity, (2) *normative* perspectives *from within* normativity, and (3) *meta-ethical* perspectives *about* normativity. The first type of perspective is not directly relevant for the present analysis, so I will focus on the two latter types. In order to be clear, I will present a *descriptive* perspective of meta-ethics *about* normativity, *as well as* of categories of normative perspectives, which are subjective, psychologically speaking, and in this sense “within normativity.”

It is essential to distinguish meta-ethical and normative relativism (DeLapp, 2013: 63–65; Long, 2011). Normative relativism feeds on and adds normative premises to the underlying meta-ethical positions that are themselves not normative. Specifically, the *normative* relativist positions arise from *ascribing normative status* to their corresponding meta-ethical positions: essentially urging that one *ought* to try to act on or to try to refrain from acting on the meta-ethical beliefs in specific directions, or that one *ought* to hold specific types of normative attitudes of praise, blame, or evaluative neutrality on their basis. It is important to note that the ascription of normative status in this manner straightforwardly commits the logical fallacy of inferring normative conclusions from descriptive premises, as David Hume famously argued (described earlier). Meta-ethics is itself agnostic about normativity and its questions of “ought.” Thus, from the viewpoint of meta-ethics in itself, one cannot infer that “anything goes” or any other normative conclusion. Instead, that meta-ethics *does not carry any* such viewpoint. This perspective seems to be where different versions of relativism are apt to be derailed.

Because normative and meta-ethical positions are tightly knit yet logically distinct, I will present the versions of normative relativism alongside their respective meta-ethical underpinnings. A single meta-ethical position will in some cases sub-serve several types of normative relativism. In practice, the underlying meta-ethical views are not always explicit; sometimes they are merely implicit and do not necessarily enter into the awareness of the normative relativist.

My present aim is to make the main positions of ‘moral relativism’ explicit. Some of the advantages of making them explicit are that they can be discussed, compared, and assessed more systematically. The taxonomy is meant to cover all generic versions of moral relativism in the sense that the categories presented are logically exhaustive. Hence, while there may be numerous more carefully formulated versions that I have not thought of, they are

presumed to be classifiable as sub-divisions within the generic taxonomy. The taxonomy I present is only one out of several possible ways of sorting between different types of relativism; for a different taxonomy, see Goldman (2012: 197–220), who argues for what he calls “objectivity-based relativism” construed as an intermediary position between “nihilistic relativism” and “objectivism.”

Normative relativist positions are often lumped together into one general category (e.g., Long, 2011). For instance, they are sometimes summarily equated with “subjectivism.” However, it may be meaningful to disentangle some of the different varieties because they do in fact differ markedly from each other. The different varieties disagree with one another regarding their judgment of which set of moral values one should morally endorse. This difference is somewhat controversial because proponents of these relativist views may sometimes deny that they disagree for reasons that I believe are unsound (see below). There seem to be two basic types of disagreement in moral relativism. Essentially, the difference between meta-ethical positions reflects distinctions in *belief*. Differences in normative positions, in contrast, reflect differences in moral *attitudes* that go beyond mere beliefs proper. In the following sections, I present what I take to be the main dyadic relativist positions of meta-ethics and normative ethics.

7.2.1 Meta-ethical and normative nihilism

First, *meta-ethical nihilism* exists regarding values. Meta-ethical nihilism is eliminative in the sense that argues that there are no moral values. Unlike conventional error theory (e.g., Mackie, 1977), however, there is not only the view that there are no *objective* values (see Ch. 8) but there are also no *subjective* values. Concerning the status of values, eliminative materialism (mentioned in Ch. 3) counts as a version of meta-ethical nihilism. Correspondingly, meta-ethical nihilism will reject the dual-aspect monist position I endorsed

(Ch. 3). The discussions in previous chapters (Ch. 2, 3, 4, 5) suggest that meta-ethical nihilism is false because it seems we are, after all, equipped with biologically based affective values that we experience subjectively as feelings. *Normative nihilism*, meanwhile, goes *further than* meta-ethical nihilism and argues that nothing really matters; since nothing matters, there is no point to ethics.

Classically, Friedrich Nietzsche is often suggested as an exponent of nihilism. However, there is much controversy concerning how best to interpret Nietzsche (Blackburn, 2005: Ch. 4), and some sophisticated interpretations (Prinz, 2007: 217; Leiter, 2003: Ch. 4) suggest that Nietzsche may not have been a nihilist about values in any of the senses I just presented but rather in line with *meta-ethical relativism* as described below. According to Leiter's (2003) reading, Nietzsche was an anti-realist about values but no denier of phenomenological experienced subjective values.

By contrast, Arthur Schopenhauer (1818) more clearly appears to have been a meta-ethical as well as a normative nihilist in his seminal work, *The world as will and representation*, arguing that "the world is nothing" and that none of our worldly strivings matter once our "will" has been overcome. His ideal was the "ascetic" who embraced nothingness and abandoned all worldly desire to retreat into mysticism.

7.2.2 Meta-ethical relativism

Meta-ethical relativism represents another distinct understanding of normativity and values. This view holds that personal values are *relative to persons*. This idea exists in terms of what meta-ethical relativism defends; the position can alternatively be stated in terms of what it rejects (e.g., Miner and Petocz, 2003: 13). Other versions of meta-ethical relativism are proposed, with such values being relative to time or place, as argued by Williams (1989); for brevity of discussion, I will leave out such alternative versions presently. As noted earlier (Ch.

6), there does not seem to be any unambiguous distinction to be made between *personal values* and *moral values*. Hence, I shall think of moral values simply as personal values applied to a normatively charged context. With this understanding of moral values in mind, I concur with Helm (2001: 12) in proposing that “*personal values* [...] are relative to the individual in the sense that what personal values it is right for me to hold may well differ from those it is right for you to hold.”

Meta-ethical relativism rejects the notion of objective values (as elaborated on in Ch. 8). Since values are inherently subjective, it is arguably meaningless to think of values in terms of truth and falsity. Regarding their normative status, there is no fact of the matter. Starkly different normative conclusions can seek grounding in meta-ethical relativism, and I have identified three distinctive normative positions that are discussed below. Two of these normative positions can be filed under normative relativism, while the last one stands in opposition to normative relativism in defense of *personal values* as a basis for moral deliberation. The gist of the charge against normative relativism is that it does not take personal values seriously as a basis for normative assertions, expressions, defenses, etc. Below, I explore the two types of normative relativism that are often associated with meta-ethical relativism.

7.2.3 Normative cynicism

Normative cynicism centers on the attitude of *indifference*. According to this line of reasoning, if values are relative to persons, then any view (objectively speaking) must be as good as another. Hence, it argues, since any normative view *really is* as good as any other normative view, it either does not matter which moral view one holds or adopts (permissive normative cynicism) or one is committed to fundamental indifference by not holding or adopting any views at all (dismissive normative cynicism).

A commitment to the ideal indifference may sometimes be explained by an extra-theoretical concern for *equality as fairness*. Therefore, from the view of normative cynicism it is, for example, just as good to criticize corporate fraud as to condone it (permissive normative cynicism); it is also not an open moral option to be judgmental about corporate fraud (dismissive normative cynicism). If one thinks that beheading of the infidel is abhorrent, that is fine; if one holds that such actions are morally required, that is fine, too (permissive normative cynicism). Perhaps one even ought to abstain from being moralistic about beheadings of the infidel (dismissive normative cynicism).

7.2.4 Normative syncretism

The second distinctive position that can be arrived at from meta-ethical relativism is *normative syncretism*. Normative syncretism follows *tolerance* to its logical extreme; one ought to (or must) *accept all* normative views equally. By accepting all views equally, the reasoning goes, one guarantees their equal moral status. “Cultural relativism” in this sense is captured by the idea that “all cultures have the same moral worth” and is a type of normative syncretism (although this is of course itself one among many normative positions). Self-effacing acceptance of all normative statements and manifestations is another type (insofar as standing behind a particular moral view can become difficult while equally accepting every other view).

7.2.5 Meta-ethical collectivism

Finally, there is the third meta-ethical position, which I have dubbed *meta-ethical collectivism*. I will not insist that meta-ethical collectivism is not properly speaking a version of meta-ethical relativism because clearly it does view values as relative; however, it is not relative to persons but instead to collective social entities. Since it is a separate meta-ethical position, which must be kept apart from meta-ethics that concern personal values, I have

chosen to give it a separate name. Meta-ethical collectivism defines a special hinterland between meta-ethical relativism and meta-ethical realism. This position represents the view that there are no subjective or objective values. Instead, it proposes that there are moral collective values (or community values). Often, these are taken to be socially constructed inter-subjectively. Since there are no other values to consider, socially constructed values become the very foundations on which human morality rests (Blackburn, 2005: 124–128).

7.2.6 Normative collectivism

Normative collectivism is the view that the set moral values that are shared by members of the group or community you belong to, are the right ones for you *qua* member (particularist normative collectivism), or for everyone (universal normative collectivism). Particularist normative collectivism, as I portray the doctrine here, is similar to what Midgley (2003: 14) refers to as “moral isolationism”; a doctrine that would aptly character a “sect” or “cult” as ordinarily understood. Possible extra-theoretical explanations for normative collectivism are *group loyalty* and *group identity* (Douglas, 1992). These cognitive-affective attitudes can be held in various degrees. Anthropologist Mary Douglas (1992: 178) suggests that the stronger a group boundary and the more difficult the group entry into a group, the greater the push towards “sects” at one extreme of a continuum. Such instances will be particularly strong cases of normative collectivism to the point that, ultimately, the voice of the individual is extinguished.

There are many examples of normative collectivism in moral philosophy (e.g., Copp, 2001: 103–122; 1997; 1995). In moral philosophy as well as in applied ethics, many versions of normative virtue ethics are representatives of this position. Further examples include absolutist professionalism (where only the professional values, ethos, virtues, norms, codes,

rules, and roles count as normatively guiding), absolutist nationalism, absolutist regionalism, and various absolutist religions positions. By “absolutist,” I mean “pure” and “fundamental.”

7.2.7 Overview of moral relativism

It can be helpful to summarize the positions that are filed under the broad heading of “moral relativism” by visualizing them under one conceptual overview. I present an attempt at this below. By seeing the different views side by side, their differences and kinships can hopefully be more clearly discerned.

Fig. 9—A taxonomy of theoretical positions within relativism

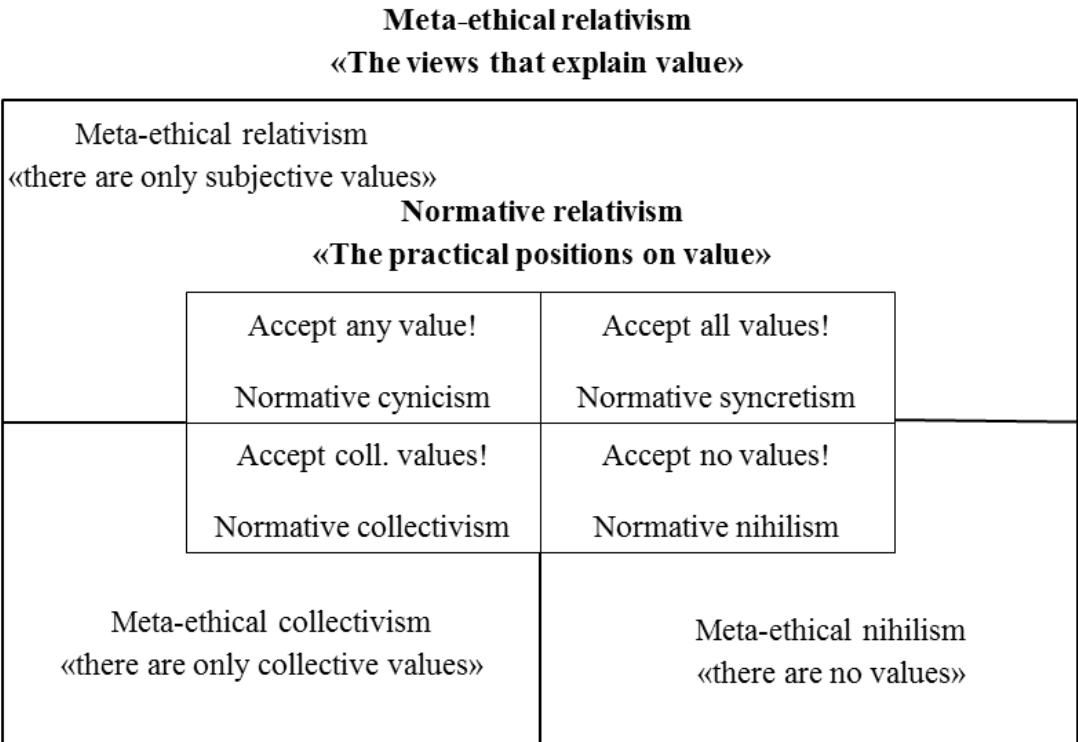


Fig. 9 sums up the three meta-ethical positions and the four normative positions that “moral relativism” covers. To be precise, meta-ethical nihilism is not genuinely relativistic, but I have included it in the overall presentation because it explains normative nihilism, and normative nihilism does issue in relativistic conclusions. As the analysis shows, the positions in *Fig. 9* are distinctively different positions, often with quite distinct decision-making implications and

practical attitudes. The figure shows theoretically archetypical positions which are distinguishable by their different ways of fixating particular conceptualizations of value. *Meta-ethical views* are concerned with the epistemological and ontological *status* of values, whereas *normative views* with which type of values to *endorse*, if any.

7.3 The rejection of normative relativism

Chains of reasoning exist for specific purposes; they must *achieve* something (Wallace, 2006; Blackburn, 2005). Either they provide a way of anchoring our explicit thought processes in empirical evidence and what we take as facts, or they move our thinking from a starting point (premises) to an end point (conclusion). The former purpose has to do with the ideal for *correspondence* to truth (Blackburn, 2005) and *consistency* (Goldman, 2009: 68), which is of practical importance because things tend to go better when we aim at getting things right (Blackburn, 2005: 171–121). The latter purpose is associated with means-end *coherence*, which is important because, insofar as we are able to reason correctly, we can avoid certain errors and undesirable consequences that may flow from them.

In doing either of these things, the activity of reasoning can move us from one mental state to another; for example, we may be moved by (and hence from) a mental state of desire (for a cold beer) combined with a belief (that the cold beer is in the fridge) to a practical or conative mental state (and intention to go and look in the fridge). This particular example is trivial, but reasoning in practical decision-making about moral matters is not.

If, for example, a practical decision-maker has inconsistent beliefs about central facts concerning risks and consequences, acting on one assumption or another may save lives or could put them at stake. Consider the decision of TEPCO, a Japanese energy company, to construct Mark 1 nuclear reactors in Fukushima (a major investment) on the assumption that they were safe to install even in a country that frequently suffers earthquakes. In hindsight, the

Fukushima Daiichi nuclear disaster triggered by a tsunami on March 11, 2011, suggests that this assumption was mistaken. Inconsistent evidence and inconsistent beliefs give practical decision-makers a reason to search for more evidence until their beliefs reflect a stable evidence-based basis for reasoning. The sparse availability of evidence concerning the risks involved in operating this type of nuclear reactor in earthquake-prone counties (this was essentially uncharted water) coupled with considerable knowledge about the perils of radioactive contamination call this particular decision in question.

I suggest that the general problem with several versions of normative relativism is that their chains of reasoning do not get us anywhere; they are viciously circular (tautological). This tautology can be captured by the “recoil argument” (Blackburn, 2005: 25–28). In practical decision-making, vicious circularity is problematic for several reasons. The most obvious point is perhaps that such lines of reasoning do not have any practical value. A more serious issue is that they also can have considerable *disvalue*; they can be costly. Circular reasoning in normative relativism can also be *deceptive*, as when decision-makers and others mistakenly believe they have arrived at a sensible, practical conclusion. Finally, tautologies of reasoning can represent *obstacles* for practical deliberation and discussion of interesting and important issues. Before moving on to look at specific problems with normative relativism, I will present some of its admirable traits, which in part explain why it appears to be so popular and so deceptive.

7.3.1 Admirable and deceptive traits

It seems hard to flatly deny that there is something compelling about normative relativism. In my exposition of the different versions of normative relativism, I mentioned several central virtues: tolerance, equitability, fairness, and loyalty. These are generally considered good things, and specific normative-relativistic positions often seem to uphold such ideals.

Toleration, for example, is normally regarded as a virtue. It is widely considered an admirable disposition of a character trait in a person. It is a “good” that is particularly *social* in its outreach. We are generally inclined to like people who tolerate us and our values, or at least the fact that they tolerate us. Perhaps if more people were tolerant more of the time, the world would be a better place; it might even be happier. Organized human activities would run more smoothly, and the organizations populating our social world may fare better with less friction and obstruction.

Conceivably, then, tolerance is an attitude that in large measure could be beneficial from a social as well as an organizational and managerial point of view. The chief motivation behind most proponents of relativism may well be precisely the value or virtue of toleration (e.g., Prinz, 2007; Wong, 2006).

7.3.2 On tolerating toleration

Is toleration *always* a good attitude? It is easy to find examples where most people would agree that the marginal value of more toleration is negative. For example, it seems that people worldwide are growing increasingly intolerant toward the practice of dumping carbon pollution into the atmosphere by fossil fuel companies; moreover, on average they are supportive of a general intolerance concerning this particular habit. This shows that, as commonly perceived, there are occasions to be tolerant and occasions to be intolerant. The exact threshold that separates these attitudes will of course vary somewhat between individuals. The point is that tolerance, at the margin, does not always have positive valence.

The virtue of *toleration* is perhaps the central unifying notion that a wide variety of normative relativists endorse (Quintelier and Fessler, 2012: 97). While *normative syncretism* seems especially inclined toward this particular virtue, variations of *normative cynicism* can easily do so as well, for example, if it stresses equality and derives an imperative of equal

respect from this. *Normative collectivism*, moreover, is naturally positive to toleration *within* one's in-group, which is essential for group cohesiveness. By extension, normative collectivists may also come to demand toleration from the social world outside the group.

A problem arises with normative relativism if it takes toleration as an absolute value, virtue, or duty. It should be noted that this point is recognized by some normative relativists (e.g., Wong, 2006). The problem is that whenever they attempt to establish their normative position in this manner, it seems bound to fail because it follows that they are unable to tolerate intolerance. If toleration is their virtue, then intolerance becomes their vice. If tolerance is a supreme virtue, its negative, intolerance, will be a supreme vice. In not tolerating intolerance, however, they seem committed to committing their very own supreme vice.

It appears psychologically impossible to be tolerant and intolerant toward an object at the same time. The crux of the problem seems to be that normative relativism easily becomes a slippery slope to what it supposes is its polar opposite, *normative absolutism*. As Williams (2006: 126) puts it: "The problem with tolerance is that it seems to be at once necessary and impossible." If one is to determine when toleration is appropriate and how much, it must be evaluated, and evaluation seems to require a set of values, which as we have seen (Ch. 5) stems from the subjective affective experiences of a particular person (the evaluator).

Normative relativism *trivializes* the normative importance of values precisely *because* they are anchored in persons or, as in the case of normative nihilism, consider talk of personal values as void nonsense. Normative relativism thereby derails the normativity that arises from our affective experiences. Moreover, because normative relativism *also* denies the possibility of anchoring values objectively, hence rejecting any objective values, proponents seem committed to reason about values along chains of reasoning that neither start in some

foundation nor really end anywhere, unless the endlessly circular (i.e., infinitely regressing) chains of reasoning are arbitrarily cut off somewhere.

Finally, the arbitrary normative guidance of normative relativism will not help moral decision-makers to make good choices because “good” or “right,” in the relative sense, is a moving target without any genuine basis (Blackburn, 2005).

7.3.3 On normative collectivism

I will not be able to undertake a full treatment of all arguments in the substantial literature associated with what I call normative collectivism. Hence, I will limit myself to making a few critical points. As noted, *normative collectivism* anchors morality in the collectively held values of a group or community. To my mind, this lifts the central problems that other versions of normative relativism suffer from up from the level of the normative positions of individuals to the normative positions held by groups. A point that is essential for defenders of normative collectivism, but often overlooked, is that valid criteria for identifying individuals as members of a binding normative group or community must be provided (Turner, 2010: 26). This is related to the problem of specifying what notions such as “community” mean (Turner, 2010: Ch. 5), and whether there truly can be anything such as “collective consciousness” (e.g., Emile Durkheim) and “collective intentionality” (e.g., Wilfrid Sellars) .

The position of normative collectivism also raises *further questions*. One question is how individual values are to be *aggregated*; another is the sense in which groups can be said to actually “hold” values. Does a company such as Google hold values? The decision-makers of companies regularly make “value statements,” featuring “core values,” or device “corporate codes.” It seems a trivial point, however, that companies and organizations ultimately are constructions in the minds of people: fictions that aid individuals to get by,

plan, socially orient themselves, and coordinate their activities socially. Beyond this and a range of material manifestations (legal papers, buildings, computers and offices, etc.) that testify to our taking these fictions seriously, an organization cannot be said to exist. And if organizations, as separate “things,” do not have any proper existence that is mind-independent and free from constantly needing social re-confirmation, how can they be said to have values or ever represent an independent set of values?

Only humans, along with sentient animals who possess an emotional brain infrastructure, can, properly speaking, *hold* values (Ch. 4). However, through our social interactions, we can hold values *inter-subjectively*, meaning that our individual values are socially adjusted in a reflexive manner. Thus, while social values can exist among the members *within* a group, it seems they cannot be held *by* a group.

Now, it is sometimes claimed by normative collectivists (e.g., Guilfoyle, 2011; Copp, 1995) that the values of a group are “socially constructed” through processes of social interaction. This idea suggests that they “come into being” somehow and stay that way either forever or until they are “disposed of” through some process. It needs to be stressed that for anything like “social construction” to take place, there must be underpinning affective *biological premises* (Ch. 5). There cannot be such a thing as a genesis of values through social construction, *tout court*. Social values, then, must be socially aggregated from personal values with roots in the affective brains of the individuals experiencing them.

The challenge of addressing normative collectivism and “collective values” illustrates the importance of establishing the empirical biological origin of values scientifically (as was presented in Ch. 4). From this informed scientific angle, it seems evident that humans stand in relation to each other as individuals, not fundamentally speaking as parts of social entities.

It also shows us that the “moral compasses” we have are *internal*, and that if we have a responsibility for how we navigate by this compass, that is also internal. We are not recipients of external normative forces: forces that curiously escape empirical identification and behave in ways that are unaccounted for by science.

Finally, experience is individuated and not genuinely shared. As the neuroscientist LeDoux (2002: 203) notes: “A subjective experience is by definition one that is known directly only by the experiencing person.” Our empathic *indirect* abilities are evidently well-developed, as evidenced by research on mirror neurons (Gallese, 2009; Iacoboni and Mazziotta, 2007) and observable in our general success at identifying the needs and wishes of others. However, we are never literally at one mind with others. We have no compelling evidence to back up such a claim.

A final point is that normative collectivism seems to imply a *coherentist* approach to truth: in this case, truth about values. Consider for example the climate debate. On one side are the “deniers,” who *deny* that we live in a warming world and *deny* that human activity (such as burning fossil fuels) has anything to do with the global average temperature (there are other versions of denial as well, but grant this for the sake of the argument). On the other side, there are the “believers” who believe that the world *is* warming and that this warming *is* significantly human-induced. Both of these sides of the debate may have a coherent worldview that they share with roughly everyone they see as members of their group.

Clearly both groups in this example cannot possibly be correct at the same time as their views directly contradict each other. This contrast shows that to form a coherent system of values that is shared by all the members of a group is sufficient neither for establishing the meta-ethical existence of collective values (i.e., meta-ethical collectivism) nor the moral

legitimacy of such values (i.e., normative collectivism) insofar as meta-ethics is a premise for legitimacy.

There is nothing wrong with coherence *in* itself; however, *by* itself it is insufficient. To establish a scientific outlook, we need something in addition to a coherence theory. What is needed seems to be a *correspondence theory* of truth. In addition to presenting a coherent view and advancing claims that are coherent within this view, we need to ascertain the correspondence relations between the claims and empirical evidence. For claims about values, we need to determine their relation to empirical evidence regarding how values originate in and are caused by subcortical neurological circuitry. In the case of “collective values,” this explanation seems to grind to a halt.

7.4 The error of error theory

Error theory, inasmuch as it identifies cognitive errors of belief as a possible explanation for morality, is a cognitive theory. In philosophy, Mackie (1977) made a seminal defense of this position. His key point was that when involved in moral talk, thought, and action, people mistakenly believe in the existences of “objective values,” whereas in fact “[t]here are no objective values” (Mackie, 1977: 15). I think there are cogent reasons for believing that Mackie is right about the non-existence of objective values, a point on which I will later elaborate (Ch. 8). In this section, I will comment on error theory as an explanation of moral motivation.

Although we grant for now that error theory is correct about the error involved in beliefs about “objective moral values” and “moral truth,” it does not follow that these cognitive errors themselves explain moral thought, talk, and action, and therefore, moral motivation. For this to be the case, cognitive errors of morality must in addition displace the moral motivation that stems from affective emotions (as presented in previous chapters). This

suggestion seems discordant with key points that affective neuroscience suggests about the workings of the human brain.

As discussed earlier (Ch. 5), evidence from affective neuroscience suggests that affect is embedded in the cognitive processes of the brain. This concept is implied by the nested hierarchy of consciousness (Solms and Panksepp, 2012), in which a bottom layer of affect forms the basis of all higher forms of consciousness; “That is a reasonable way for the internally motivated brain to be organized, with more recent neural developments still being solidly grounded in what emerged earlier” (Panksepp, 2012: 6). This idea seems especially reasonable when we consider the fact that brain anatomy is a product of evolutionary forces that gradually resulted in the development of more complex structures.

A cognitive attribution error fails to identify the proper object of attribution, and the meta-belief that this attribution is nevertheless correct in terms of correspondence with truth will only sediment the attribution error. However, unless there is also at least a minimal involvement of affective emotional input, it seems that (1) there will be no conscious deliberation at all and, if the input is of insufficient quality, (2) one will fail to properly assign evaluative weights representing moral importance and urgency, and hence (3) moral motivation will fail. In other words, a belief in “moral truth” will not by itself motivate unless the agent also *cares* about this moral truth and feels drawn by it (or repelled by it). Error theory, therefore, may be a plausible theory of the status of specific beliefs involved in moral deliberation but is inadequate as a motivational theory.

7.5 Inescapable relativism and the brain

Everything humans think and do depends on our brains. Moreover, as was discussed earlier (Ch. 3), only the tip of the iceberg of the processes that underpin what we think and do ever comes into conscious awareness and scope or reflection (Berlin, 2011). This idea is significant

because it means that “our brains” must fend for “themselves,” which has implications for how entrenched the relativism of our deliberation is. The neuroscientist, MacLean (1990), puts it this way:

“Since the subjective brain is solely reliant on the derivation of immaterial information, it can never establish an immutable yardstick of its own. Hence, for these reasons it is left with nondimensional space and nondimensional time for which it must arbitrarily set standards of its own. In this respect, it is saddled with nonyielding relativity” (MacLean, 1990: 571).

This statement also covers the moral standards or the emotionally based priority that our brain instills in us. We might object to the perceived demands that our brain equips us with, but every such objection would itself be based in part on motivational feelings and depend on and results from the activities in our affective emotional systems because these systems are constantly present, working as they do, essentially completely outside the reach of our own awareness. The picture of a “faculty of reasoning” somehow standing back from all this, impartially and objectively, is a misconception. We *are* ultimately our brains or the product of our brains, and the evidence for an independent faculty of reasoning is, as we have seen (Ch. 6), not particularly compelling.

7.6 Moral anesthesia

In this section, I will try to generalize some of the *practical dangers* of normative relativism. Normative relativism represents a threat to the decision-making agent herself as well as the people whose values are at stake in the decision. A full exploration of the threats cannot be pursued here, so I will focus on central subordinate themes.

As far as the decision-maker herself is concerned, it seems evident that, quite independently of the specific normative content represented by a decision-maker’s subjective

values, no decision-maker would be well served to unwittingly and subconsciously arrive at arbitrary practical decisions. As indicated, this can happen when a decision-maker is unaware of the circularities of their relativistic reasoning and simply cut it off or come to a stop at some arbitrary point. Stakeholders, in turn, may suffer the consequences that flow from arbitrary decisions.

One of the most potent types of practical dangers that reasoning along the lines of normative relativism involves seems to be what I call “moral anesthesia.” This term indicates the sensation that moral reasoning seems futile, hopeless, or impossible, and it results either from experiences of circular reasoning or from perceived endpoints in reasoning that suspend moral judgment. In these cases, there may be varying degrees of awareness of the predicament that normative-relativistic reasoning leads to, although the causes of this predicament are not well enough understood to escape the predicament. For example, one line of reasoning has it that, since the universe is vast and void of value, nothing really matters (an exemplification of normative nihilism). It is easy to see that this attitude can be destructive for social business practices, both internally in an organization as well as in relation to external stakeholders (Karmasin, 2002; Miller, 1991).

The failure to picture oneself as a moral agent with personal values that are normatively potent and “valid” threatens to intellectually (cognitively) suppress affective emotions. This effect is especially true of self-reflexive (tertiary) emotions such as pride, gratitude, shame, guilt, and embarrassment. These emotions depend on a sense of moral self (Fontaine, 2009). By denying the moral self, the reasoning of normative relativism can blunt important aspects of our moral lives. These depleting effects can be further exacerbated by social norms in the organizational and professional contexts, as when social norms express normative collectivism and favor collective responsibility over personal responsibility.

When actions are said to be performed by corporations and professional functions (i.e., “corporate actions”) as opposed to moral agents, there is a threat that we may lose or suppress the emotions that reflexively refer to moral agents. Corporations and organizations are not capable of feeling pride, gratitude, shame, guilt, and embarrassment, as normal human beings are; instead, they are metaphorically analogous with psychopaths (Bakan, 2004). Insofar as playing the professional role of business professionals consists of counteracting and downplaying moral emotions, we risk losing these self-reflexive emotions in our interpersonal dealings and as a basis for executing normative decisions.

Our capacity for *cognitive dissonance*—roughly equivalent to the psychodynamic term *defense*—suggests that we are able to dissociate ourselves from the roles that we play (Festinger, 1957). A reduction in uncomfortable cognitive dissonance can be achieved by emotionally blocking out the personal emotional involvement from the sphere of professional conduct. The primary function of cognitive dissonance may be to “organize and maintain mental life in a way that protects the individual from aversive emotional experiences” (Turnbull and Lovett, 2012: 190). Normative relativism, I argue, facilitates this form of moral anesthesia, potentially making it possible for professional decision-makers to make decisions that are morally controversial but which still do not *feel* emotionally uncomfortable. This raises the question: *Should* they feel uncomfortable? This question can only be answered by weighing the comfort of decision-makers against the resulting calamities potentially inflicted on various stakeholders (which is partly empirical and partly normative).

7.7 Conclusion

This chapter presents a dual position concerning relativism about values. “Moral relativism” is unpacked so that the different positions often placed under this term can be assessed separately. With the aim of understanding normativity and the foundations of practical ethics,

it is critically important to distinguish between meta-ethics and normative ethics. Meta-ethics is *about* normative ethics, whereas normative ethics concerns taking evaluative positions *within* ethics (see Ch. 1). It is also, however, crucial to understand how meta-ethics and normative ethics are connected. In order to sift between the different theories of relativism, I developed a taxonomy showing the main positions or theatrical archetypes.

It is argued that we should tentatively accept meta-ethical relativism, which pins normativity to our affective subjectivity (further support is presented later, see Ch. 8). Four different versions of normative relativism are presented. Based on a discussion of their similarities and differences, it is shown that although there are many tempting features of these theories, they do not represent stable normative positions and therefore should be rejected. There are potentially undesirable effects from reasoning along the lines of normative relativism, including the confusion of circular reasoning, the disillusionment of abandoning personal values and convictions, the cynicism of nihilism, the irony of post-modernism, and the comfort of moral anesthesia. None of these effects are likely to improve moral decision-making, and instead they seem to increase the risk of making morally bad decisions (as inter-subjectively evaluated).

Chapter 7 References

- Bakan, J. (2004) *The corporation: The pathological pursuit of profit and power*. London: Constable and Robinson.
- Berlin, H. (2011) The neural basis of the dynamic unconscious. *Neuropsychoanalysis* **13** (1): 5–31.
- Blackburn, S. (2005) *Truth: A guide for the perplexed*. London: Penguin Classics.
- Bishop, J. D. (2000) A framework for discussing normative theories of business ethics. *Business Ethics Quarterly* **10** (3): 563–591.
- Bowie, N. (2013) *Business ethics in the 21st century*. Berlin: Springer Science & Business Media.
- Copp, D. (1997) Does moral theory need the concept of society? *Analyse & Kritik* **19**: 189–212.
- Copp, D. (1995) *Morality, normativity, and society*. New York: Oxford University Press.
- DeLapp, K. (2013) *Moral realism*. London: Bloomsbury.
- Douglas, M. (1992) *Risk and blame: Essays in cultural theory*. London: Routledge.
- Festinger, L. (1957) *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Fontaine, J. R. (2009) Self-reflexive emotions. In D. Sander and K. R. Scherer (Eds.), *The Oxford companion to emotion and the affective sciences* (pp. 357–359). New York: Oxford University Press.
- Gallese, V. (2009) Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanalytic Dialogues* **19**: 519–536.

- Goldman, A. I. (2012) *Reliabilism and contemporary epistemology: Essays*. Oxford: Oxford University Press.
- Goldman, A. H. (2009) *Reasons from within: Desires and values*. Oxford: Oxford University Press.
- Guilfoyle, M. (2011) The ethical subject in poststructural therapy. *Journal of Systemic Therapies*: **30** (4): 1–15.
- Harman, G. and Thomson, J. J. (1996) *Moral relativism and moral objectivity: Great debates in philosophy*. New York: Wiley.
- Helm, B. W. (2001) *Emotional reason: Deliberation, motivation, and the nature of value*. Cambridge: Cambridge University Press.
- Iacoboni M. and Mazziotta, J. C. (2007) Mirror neuron system: Basic findings and clinical applications. *Annals of Neurology* **62** (3): 213–218.
- Karman, M. (2002) Towards a meta ethics of culture: Halfway to a theory of metanorms. *Journal of Business Ethics* **39** (4): 337–346.
- LeDoux, J. (2002) *Synaptic self: How our brains become who we are*. Middlesex: Penguin Books.
- Leiter, B. (2003) *Routledge philosophy guidebook to Nietzsche on morality*. London: Routledge.
- Long, D. (2011) *Relativism and the foundations of liberalism*. Andrews UK: Limited.
- Mackie, J. L. (1977) *Inventing right and wrong*. London: Penguin.

- MacLean, P. D. (1990) *The triune brain in evolution*. Berlin: Springer Science & Business Media.
- Midgley, M. (2003) *Heart and mind: The varieties of moral experience*. London: Psychology Press.
- Miller, H. (1991) The shadow of ethics: A reply to Laurie Pamental. *Business Ethics Quarterly* **1** (4): 395–403.
- Miner, M. and Petocz, A. (2003) Moral theory in ethical decision making: Problems, clarifications and recommendations. *Journal of Business Ethics* **42** (1): 11–25.
- Panksepp, J. (2012) What is an emotional feeling? Lessons about affective origins from cross-species neuroscience. *Motivation and Emotion* **36**: 4–15.
- Prinz, J. J. (2007) *The emotional construction of morals*. Oxford: Oxford University Press.
- Quintelier, K. J. P. and Fessler, D. M. T. (2012) Varying versions of moral relativism: The philosophy and psychology of normative relativism. *Biology & Philosophy* **27**: 95–113.
- Solms, M. and Panksepp, J. (2012) The “id” knows more than the “ego” admits: Neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Science* **2**: 147–175.
- Schopenhauer, A. (2008) “*The world as will*” and “*Presentation*.” In D. Carus (Ed.) and R. E. Aquila (Trans.). New York: Longman. (Original work published 1818)
- Trunbull, O. H. and Lovett, V. E. (2012) Emotion and delusion: Seeking common ground between neuroscience and the psychotherapies. In A. Fotopoulou, D. Pfaff, and M. A.

- Conway (Eds.), *From the couch to the lab: Trends in psychodynamic neuroscience* (186–206). Oxford: Oxford University Press.
- Turner, S. P. (2010) *Explaining the normative*. Cambridge: Polity Press.
- Wallace, J. (2006) *Normativity and the will: Selected papers on moral psychology and practical reasoning*. Oxford: Oxford University Press.
- Williams, B. (2006) *Philosophy as a humanistic discipline*. Princeton: Princeton University Press.
- Williams, B. (1989) *Ethics and the limits of philosophy* (2nd ed.), Harvard: Harvard University Press.
- Wong, D. (2006) *Natural moralities: A defense of pluralistic relativism*. Oxford: Oxford University Press.

Chapter 8: The case against objective values

8.1 Introduction to the case against objective values

The notion of “values” is generally taken to be normative and motivational or action-guiding (e.g., Goldman, 2009; 2008; Helm, 2001: Ch. 1; Etzioni, 1999). In the previous chapters, I proposed that values, conceived of as subjective and internally derived from emotions, motivate from within. As is the case with emotions, values resonate or discord with cues in the *external* world but do so in virtue of the *internal* biological constitution of our emotional infrastructure and its upshots as subjective feelings.

The current chapter is dedicated to assessing the opposing position concerning values. The focus will be on the philosophical position that implies that values are normative and motivational on the one hand, yet neither subjective nor internal on the other. In moral philosophy this area of discussion is conventionally referred to as the “theory of values” (van Fraassen, 1973: 5; Hart, 1971). The central question is what status values actually have—what is the nature of values? The outline of my own account has already been given; this chapter reviews alternative accounts in more detail.

The existence of “objective,” “mind-independent,” or “external” values (Goodwin and Darley, 2008) is frequently argued for in contemporary moral philosophy (e.g., Wedgwood, 2007; Crisp, 2006; Nagel, 1997; Brink, 1989; Nagel, 1986). This position is also found in psychology (e.g., Rokeach, 1968) and in applied ethics as, for example business ethics (Anderson, 1997), but it is often implicit rather than explicit. The views of objectivity, mind-independence, and externality are connected. For instance, Goodwin and Darley (2008) argue that there is a psychology of meta-ethics behind the ethical objectivism we find in normative and applied ethics. This suggests, as noted earlier (Ch. 1), that reviewing the meta-ethics

behind normative ethics and making these explanatory accounts explicit is a matter of practical importance.

The two opposing positions concerning the nature of values are contrasted in this brief chapter. On the one hand there is the meta-ethical position that values are objective (external); on the other hand is the view that values are subjective (internal). This concept represents a central division in the epistemology of values. The question is, as Nagel (1997: 3) puts it: Are moral values and ultimate moral justification to be found in “objective principles whose validity is independent of our point of view,” or “within our point of view—individual or shared—so that ultimately the even the apparently most objective and universal principles derive their validity or authority from the perspectives and practices of those who follow them?”

Arguments for and against the two positions are presented, and the view that values are objective is dismissed. Several problems with the view that values are objective are highlighted, including the problem of understanding their nature, their motivational force, and the psychological relevance to our normative deliberation. The outcome of this discussion, therefore, corroborates the meta-ethical relativist position presented in the forgoing chapter (Ch. 7), which represents the internalist position.

8.2 The case for objective values

In this section I present some of the central philosophical arguments for and against objective values. The position I argue for coincides with that of Goldman (2009; 2008), who provides a more detailed overview of the philosophical debate than what I am going to present here. My debate connects with the discussion in previous chapters, and I relate the arguments for and against objective values explicitly to the perspective of affective neuroscience.

8.2.1 Values and reasons

According to Goldman (2009: 11), there is a tight conceptual relation between the notions of “values” and “reasons.” From the decision-making perspective, practical reasons count for or against prospective actions or decision alternatives. Practical reasons imply values in that they indicate the things to approach or create (the things of value) and also the things to avoid or combat (the things that are of disvalue). Practical reasons can only be independent of the set of motivations, desires, and concerns in an agent’s mind insofar as values or normative facts are. Values also imply practical reasons because if values are to be relevant, they must serve as normative guides to action and practical deliberation (Goldman, 2009: 11). As I have argued previously (Ch. 5), values represent affective-normative guiding (subjective) feelings based in our (objective) emotional systems. Our values, I argue, are best seen as hierarchically structured, affective, subjective, normative, and internal compasses fit to guide moral action and deliberation.

One of the advantages of this type of internalist account of values and reasons is that we can explain normative judgments by referring to our subjective motivational sets (Blackburn, 2010; Goldman, 2009; Williams, 1981: 101-113). This approach depicts a simple and clear view of moral motivation. What is normatively “good” or “bad” (relatively speaking) is a matter of perspective and is determined by *what we care about*, and how much we care about it (Frankfurt, 1988: 80-94); “Caring just *is* being motivated” (Goldman, 2009: 11). Subjective values give internal reasons for action.

Contemporary value objectivists present an alternative theory about motivation. In this view (e.g., Nagel, 1986), “objective values reduce to external reasons” (Goldman, 2009: 11). The central point of this externalist view is that values are independent of the concerns, desires, and motivations agents *actually* have; instead, the question is what concerns, desires, and motivations agents *ought* to have. Normative ethics based on this view is an external enterprise.

8.2.2 Externalism and internalism

There is a close parallel between philosophy and the psychology literature concerning internalism and externalism. In moral philosophy, *externalism* is generally formulated within meta-ethics, particularly regarding various theories that can be filed under meta-ethical naturalism, meta-ethical realism, or more generally, meta-ethical cognitivism (revisit Ch. 7, for details). Thus, for example, meta-ethical cognitivism (e.g., Nagel, 1986) suggests that normativity has an external locus and that it can be accessed cognitively by means of perception and deliberated about in terms of beliefs (sometimes called “evaluative beliefs”). As we saw earlier (Ch. 7), meta-ethical collectivism, at least in its standard version (e.g., Copp, 1995), also presupposed an external locus of normativity. Meta-ethical relativism, in contrast, represents an internalist account of normativity.

The boundary between the external and the internal is normally taken to be definable by the separation between the subjective mind or what can be *experienced* (the internal) and everything else, which is objective (the external). This is the same boundary that I discussed previously (Ch. 3) regarding the distinction between the mind and the brain. Externalism in moral philosophy involves *two claims*, which are theoretically separable but often treated as a unit.

The first claim is that the *justification* of moral decisions and actions is to be found outside the mind of the agent. This is sometimes referred to as “externalism about justification” (Crumley II, 2009: 159). The second claim by externalism is that the nature of the *values* that feature in the normative justification is external to the agent and thereby “objective.” We can, analogously, call this “externalism about values.”

Values, as seen from an externalist theoretical perspective, are objective in two different senses. In the first sense, they are objective because they are external to the mind and

thereby mind-independent. In the second sense, they are objective because these external values are normatively justified, authoritative, or “validated.” Finally, “external” or “objective” values are normally taken to play the central role in externalist theories of moral motivation. A key challenge here is to explain how we can reconcile the mind-independence of values with the normative motivational force they exercise on our minds.

8.2.3 Proof and reasoned belief

At the core, externalism concerns the nature of values. As noted, objective values are embedded in the concept of practical reasons. According to this view, reasons “exist whether or not agents take any notice of them,” as Blackburn (2010: 283) explains, and therefore “do not only exist in the light of contingent desires or mere inclinations.” This belief raises an ontological problem. How can we determine whether “objective values” or the corresponding “objective reasons” they would generate exist?

As a matter of logic, it is impossible to *prove* the *non-existence* of something. For example, it is impossible to prove the non-existence of the *non-natural properties* of nature that Moore (1903) famously postulated and that he claimed constituted the foundations of normative ethics. This type of claim will therefore have to be believed on the basis of faith alone. No amount of empirical evidence can even conceivably show that non-natural properties do not exist because Moore’s claim is not, in the classical Popperian sense, falsifiable.

This dilemma is a general problem for empirical science. For example, it is impossible to *prove* that gene-modified food is not harmful. The fact that there is no evidence of harm does not prove objectively and finally that no harm results from consuming gene-modified food. This is also an inherent problem for building a case against the existence of *objective*

values or *objective reasons for action*. It is an *a priori* fact that facts alone will not be able to provide us with proof of their non-existence.

Proving the non-existence of objective values, therefore, seems to present a problem without a solution. Externalists of normativity could argue that this issue counts in favor of their view. However, it seems to be more of a moot point that is neither for nor against the externalist view.

The conspicuous lack of evidence on which to base the externalist view, however, classifies as a counterargument. This is an especially serious argument against meta-ethical realists since they claim that there is a factual basis for an external account of normativity. Consequently, there are strong arguments against the *belief* in the existence of objective values and reasons. There also seems to be a solid case for skepticism regarding objective values. Since we already have a readily available internalist account of values with a partial biological explanation, we need arguments *for* externalism that are at least as compelling in order to make our skepticism disappear.

8.2.4 The viewpoint of the universe

Another item that must be discussed is the belief in *objective points of view*. Classically, Jeremy Bentham claimed that to get ethics right, we need *impartiality*, which requires standing back and seeing matters from “the viewpoint of the universe.” This, of course, presupposes that there *is* such a viewpoint and that it is epistemologically *accessible* to us. A related classical view (prior to Bentham) can be found in David Hume, who proposed that the needed impartiality could be secured by looking at matters from the perspective of a hypothetical (Hume was an atheist) “ideal observer.”

Bergson’s (1903) metaphysics deserve mention in regards to the issues of perspective taking. Bergson proposed that it is possible to break out of one’s own subjective and relative,

point of view, to reach an absolute point of view. According to Bergson's (1903: 1) common sense philosophy, there are two ways of knowing things: "The first implies that we move round the object; the second that we enter into it. The first depends on the point of view at which we are placed and on the symbols by which we express ourselves. The second neither depends on a point of view nor relies on any symbol. The first kind of knowledge may be said to stop at the relative; the second, in those cases where it is possible, to attain the absolute." A crucial question is whether such an objective and absolute point of view is possible or even sometimes possible in matters of morality. Bergson (1903: 3) hints that this may be possible through acts of "sympathetic intuition," allowing direct access to objects without "expression," "translation," and "symbolic representation."

More recently, Thomas Nagel (1986) launched the similar claim that there exists "a view from nowhere." In his seminal book, bearing precisely that same title, he sets out to deploy the view from nowhere as a means of grounding normative ethics. The gist of his argument here and in later works (e.g., Nagel, 1997) seems to be that all manner of thought, including thinking about moral issues, is impossible without a concept of "objectivity" in mind: "The outermost framework of all thought must be a conception of what is objectively the case—the case without subjective or relative qualification" (Nagel, 1997: 16).

In this line of thinking, Rawls's (1971) "theory of justice" is perhaps the most influential contemporary exponent. Rawls uses the notion of "a veil of ignorance" (an idea borrowed from Nelson Goodman) behind which moral deliberation is to take place as a means for determining what just and fair institutions in society would have to be like. Thus, instead of trying to conceptualize what an omniscient ideal observer must be like, which might be too cognitively demanding to hope to get right, we can instead reduce ourselves to an "impartial observer" but repudiate every trace of subjectivity from our perspective. The test of this exercise is of course whether such a reduction is psychologically possible and whether there is

something left that is reliable and valid as a normative guide to what is morally “right” and “wrong.”

It seems easy enough to understand the basic ideal of impartiality. If human decision-makers and agents could only free themselves from all their narrow partial concerns and desires, impartiality that aims at the “universal good” or “universal opportunity” (for all humans or all sentient creatures, or perhaps even including the natural environment) would emerge. Such an impartial perspective would, it seems, at least move decision-makers and moral agents closer to an objective moral point of view, from which possibly something like objective values would be discernable. This idea raises several critical questions, and I will discuss two of them.

One question is whether it is psychologically and physiologically *possible* to fully—or even almost fully—free oneself from one’s particular and partial personal perspectives and personal commitments. A second important question is the normative question of whether impartial ideals would be *desirable* and on what basis such putative desirability could be uncontroversially established.

Given the preceding discussion of the human mind (Ch. 5 and Ch. 6), the reduction Rawls proposes seems psychologically doubtful; for one thing, most of what drives and motivates us from within is beneath our conscious awareness. The empirical evidence (Ch. 4 and Ch. 5) strongly suggests that human beings are unable to be genuinely impartial. Impartiality seems to presuppose much more cognitive control over our emotional and motivational mental machinery than we are psycho-biologically equipped to exercise. It is not just our private affective feelings and inclinations *per se* that are difficult to expel but also the way our affect resonates with our social interconnectedness and socio-cultural attachment. As Prinz (2007: 143) argues, “ignoring culturally imposed preferences is very difficult, because

preferences are built into our affective responses and these are not amenable to direct or deliberative control.” Moreover, it is well known that even cognitive information processing is strongly influenced by the idiosyncrasies of our social surroundings (Lines et al., 2011: 171).

Impartiality also seems to presuppose that cognition is essentially unaffected by and independent of our affective natures. As I have argued, this is apparently not the case. Although there are control mechanisms both ways (see Ch. 2) from cognition to affect and vice versa, cognition appears to be infused with affect (Solms and Panksepp, 2012). Moreover, the very consciousness of the cognitive mental apparatus seems to depend on a more fundamental underlying affective consciousness (Solms, 2013). Affective forces constantly exert their normative influence on us, and they do so largely outside our cognitive-conscious awareness. Hence, any apparent impartiality is likely to be biased.

Hume’s suggestion of an ideal “impartial observer” (or “juror”) raises the further, highly abstract challenge of determining what the emotional attitudes of this ideal observer would have to be like (Kagan, 1998: 271–280) without falling back on one’s own subjective emotionality (because then partiality would clearly seep in again). We would in other words have to *first* imagine correctly and unambiguously what a full-blooded and impartially benevolent impartial observer would have be like and *then* what her specific evaluations would be *without* at any point resorting to our habit of subjective evaluation. Psychologically, this may seem a tall order. Adding to the difficulties, which relate to our constraints on our cognitive mental abilities, we also have the difficulty of ensuring that the many evaluative processes of our brain (as discussed in previous chapters) that happen beneath our conscious awareness do not bias our extrapolative imagination. We seem to have no way of doing this.

If, in spite of the weight of evidence, it turned out that genuine impartiality was after all possible or at least nearly achievable (so that we could have partial success on this score), we would still have to determine whether an impartial perspective would be *desirable*. In order to determine this, it seems we would have to rely on our partial desires-based concerns again and on the pain of infinite regress. This is a source of subjective bias we appear unable to escape. But should we, on our *subjective* perspective, condone impartiality?

As a subjective matter, this is a question to which conceivably different persons will answer differently. Given the evident plurality of ethical views, it seems unreasonable to expect a uniform answer. However, it is also conceivable that the reflected subjective judgment of many people will land not on the side of Rawls and Nigel but on the side of partiality after all. At least this conclusion cannot be ruled out. Which parents would accept treating their own children's needs and every other child's needs in a perfectly impartial way?

Affective neuroscience points out (i.e., Panksepp, 1998) that we are hard-wired to care (i.e., the care system) in particular for our own children and other children in our care. We can perhaps rationalize ourselves away from the resulting position *cognitively*, but it remains doubtful whether most people are able to also put their emotions behind the practical implementations of such rationalizations. Possibly some people can, but there are likely very few.

A parallel argument can be made concerning business ethics. What business leader would treat the concerns of their own employees impartially and on par with the needs of every other employed person in society and around the world? It seems plausible that decisions would be hard to justify by appealing to this kind of cold-minded impartial ideal. Such a leader might well come across as a psychopathic monster in the media (which, as we know, likes to amplify stereotypical character traits). It seems likely that our hard-wired

attraction for caring for significant others also plays a major role for understanding the moral psychology of business ethics and its distinctive relational aspects.

8.2.5 Business ethics and objective values

In the business ethics literature, it is frequently simply assumed what kind of actions classify as instances of “wrongdoing” (e.g., Palmer, 2013; Balch and Armstrong, 2010), more generally, “right” and “wrong” or “ethical” and “unethical,” without any explanation of *what makes* different actions or action types wrong. Thus, Balch and Armstrong (2010: 292) claim that “wrongdoing is anything but rare” without first presenting any criteria for what counts as “wrongdoing.” Since these actions are simply declared “wrong” or “unethical,” they are implicitly presented as objectively wrong. In such cases, it is as if the ethics that is applied comes from some unspecified “view from nowhere.” This fact testifies to how ingrained the idea of objective values is—they presumably need no explanation—at the same time, it shows that precisely such claims need more explicit formulation and that they warrant serious attention.

Mackie (1977) specifically pointed out that the kind of objective claims in ethics, like the ones cited above, are mysterious, or in Mackie’s term, “queer.” His response was to deny the existence of objective values. How can something outside us, which we appear to have no knowledge about and which we may not even care about or even resent if we did know about it, still enact a normative force upon each and every one of us?

The position I defend is that the claims of ethical objectivity are *not* plausible. Unlike John Mackie, however, I stop short of flat out denying the claims of the objectivists. Instead, I present reasons against believing them. I argue for *skepticism* about objective values rather than “atheism,” but this is a skepticism that on a measuring scale would be much closer to “atheism” than the more forgiving “agnosticism.” From a practical position, I think that

business ethicists should refrain from using unsubstantiated normative terms that imply objective values or external normativity, at least until they can find credible normative foundations on which to base these claims. By deploying unwarranted objectivist terminology, they exert normative pressure and a normative authority that is likely false.

8.2.6 Motivation and reasons for action

One way to conceptualize moral decisions is to analyze decision problems in terms of reasons for action. This way, one can weight reasons for and against different alternatives in a manner that acknowledges that there can be different types of reasons. For instance, one can think that some reasons, if they appear, will be decisive or perhaps decisive only under certain specific conditions, while other reasons are merely contributory (Dancy, 2004). First, however, we need to understand the *nature* of reasons so that we can know where to find them and how to treat them. In particular, we need to determine whether they should be appropriately seen as subjective or objective.

Parfit (1984) has been credited with introducing the terms “agent-relative reasons” and “agent-neutral reasons” (Nagel, 1986). Deploying these concepts presents a way to draw a clearer distinction regarding what is meant by “subjective values” and “objective values.” In particular, subjective values can be thought of as *corresponding to* agent-relative reasons, while objective values correspond to agent-neutral reasons. The search for objective values using these guidelines becomes a quest for the agent-neutrality of reasons. Are what a person takes to be normative reasons for action somehow obliged to relate to a person’s concerns or desires?

Each of the three influential philosophers referred to in this section offer a negative answer to this question. Their ideals, while somewhat different, all allow for objective and impartial reasons in normative ethics. However, it is not clear that it makes sense to see some

reasons being normative *for* a given person at the same time as they are completely *unrelated* to that same person's set of motivations. *Why*, we might ask, should that person accept this reason as a normatively binding reason for action?

“Reasons must be capable of motivating us,” and they should explain as well as justify our actions (Goldman, 2009: 8). They have to be reasons, at least potentially, *for* someone. If our reasons are not established by our internal motivations, it seems that reasons must somehow have objective normative “validity” built into them. This view, however, would be unsatisfactory because it would involve a tautology (vicious circularity). It would amount to telling someone who dislikes vanilla that they have a reason to like vanilla ice cream because it is objectively good. The idea of objective reasons intersect with this problem.

One point needs to be added to this analysis. It may be the case that a person *thinks* she does not like vanilla ice cream, but that in fact she *would* like it if she only tried. Now, it seems that she might have a reason that is *external* to her and not anchored in her current motivations. This point, however, is not unrelated to her extended set of concerns and desires, for it is only a reason for her *because* she *would* like it if she tried. If this were not, after all, the case, it would seem nonsensical to talk about her still having an objective reason to like vanilla ice cream. The underlying question that settles the status of the reasons she has is a matter of her beliefs and how well informed they are, not her recognition and adoption of external or objective values. The values that would allow her to evaluate and eventually appreciate vanilla ice cream would still be her subjective values.

8.3 Science and the subjective

8.3.1 Determining external from internal

As evidence from neuroscience indicates (See *the mind-brain argument*, Ch. 3), the brain creates the subjective mind in which we find ourselves, and it is through the subjective mind that we relate to the external world. We (our minds) do not see the world like it really is—we live in a subjective fiction—but this is a fiction that is adapted so as to coincide with reality in a variety of ways that are relevant for our purposes (Frith, 2007: Ch. 5). Most of the ways in which we relate to the world are through our *brains*, and most of that information is completely inaccessible to us in our *minds*.

This is an important neuroscientific argument for skepticism about objective values. Even if inherently normative objective values existed in the external physical reality, it is difficult to see how they could end up as a conscious normative experience without our brains first having *made* them subjective. It is precisely *creating our subjective world* that the brain *does* for us as far as the mind is concerned (Damasio, 2011; Frith, 2007). Thus, the fact that we have some normative experience cannot be taken as reliable evidence that objective values exist out there in the external world. That would be an unwarranted conjecture, or a “leap of faith,” if you see objective values as a necessary premise for ethics. All we can really say about the nature of values on the basis of our subjective normative experiences—experience of good, right, appropriate, that there is something that ought to be done, and so on—is that values are subjective.

8.3.2 Mind and coherence

The most prevalent philosophical position in defense of the existence of objective values is straightforward *meta-ethical naturalism*. Naturalism regarding values takes values to reside in nature, typically as *normative* and *natural* properties of physical objects. This dualism, however, seems unstable and unsound because as Mackie (1977) pointed out in a response to Thomas Nagel: “there is a mysterious ‘to-be-doneness’ about these supposed objective values, and this does not sit well with the idea that they are objective and independent of us. How can

they have a hold our minds? How can natural properties of objects be at the same time independent of us *and* still necessarily motivating?” (Goldman, 2009: 187).

It can be argued that the normativity of values *requires* them to be *necessarily motivating* is too strong, and that this idea saves value objectivism. Perhaps, instead, objective values merely give us (objective) *reasons to act*? We sense or intuit values, and they provide us with reasons to act; *they* determine how we *ought* to act, whether or not this in fact results in the prescribed sort of action. Weakness of will (*akrasia*), for example, could prevent us from following through from intention formation to eventual action (Broome, 2013).

This move pushes the normative objectivity from values over to reasons for action. Unfortunately, this move is not particularly helpful. As I have argued earlier in this chapter, when we unpack reasons for action, we find two psychological components: a *value component* and a *belief component*. Therefore, we are obliged to return to the fundamental problems that arise out of conceiving values in this manner. This approach in particular offers nothing for the problem of holding together the idea of *mind-independence* (objectivity) with the idea of *mind-motivation* (normativity). At the same time, nothing really depends upon whether we in fact act or not because the question is not action but *motivation* and whether it is sufficiently strong or too weak.

8.3.3 Naturalism and emotion circuitry

There is a science-oriented argument for objective values that warrants attention. An overview of currently identified emotional neural circuits for the subcortical brain was previously presented (in Ch. 4). These emotional centers operate completely or almost completely outside our direct conscious awareness. Some would argue that they are *outside* the domain of the *subjective mind*, since we are not aware of them. It has moreover been argued that these emotional centers instill *biological values* in us (e.g., Panksepp, 1998) and give us *motivation*

in the form of “intentions-in-action” as opposed to the cognitive-consciously accessible “intentions-for-action” (Cromwell and Panksepp, 2011). If each of these points were taken as premises and each were true, it could be consistently argued that values are physiological based on the facts of reality and also that they are therefore objective.

There are several points to discuss in regards to this proposition. Since the relations between the premises are important, I will focus on them as a whole before I comment on separate premises. Firstly, it can be noted that a fully naturalistic reductive account of the mind-brain relation (discussed in Ch. 3), in which the concept of mind is effectively conflated into the physical realities of the brain, would seem to support the conclusion that values are objective in the sense of being outside the mind.

I will not repeat the entire discussion from previous chapters (notably Ch. 3), but it needs to be pointed out that full scientific reductionism is eliminativist about the mind altogether. Consequently, it goes without saying that everything, values included, must be outside the mind; *there simply is no mind* on this account. I have already argued against this view in favor of a different view: dual-aspect monism (Ch. 3 and Ch. 5). A crucial point is that the mind gives us subjective experiences, including a sense of normativity, which belongs to a different “dimension” than what can be described in terms of naturalistic facts.

Secondly, as Solms and Panksepp (2012) suggest, the subjective mind goes much deeper than the layer of cognitive-conscious awareness. The experiential lives of humans and other mammals include much more. Given the scientific evidence, it seems reasonable to claim that humans have an affective consciousness, which gives us a sense of “being in the world” (Panksepp, 2005). In a similar vein, it may also be accurate to hold that we have an affective awareness: for example, when we look at a painting but are not able to conceptualize our feelings, although they are distinct enough. The first premise above, therefore, does not

seem to hold. However, I see no reason to disbelieve the other premises. The falsity of one central premise is enough to make the conclusion unwarranted. Hence, I conclude that a biological basis of values does not conflict with internalism about values.

8.3.4 Values, beliefs, and categorizations

Some researchers who see themselves as discussing *values* (e.g. Schwartz, 2012; Rokeach, 1968) in fact deal only with *value classifications* or *value categories*. Value classifications are useful tools, especially when researchers want to compare people's different decisions, self-reported preferences, or cultural associations. However, it is a category mistake to confuse the nature of classification boxes with the nature of the values or valuations that sort into them. Hence, even if it could be shown that the classification boxes are "objective" in some proper sense and interpersonally "shared" (believed in), it does not follow that the nature of the values sorting into these boxes is objective or that they somehow become objective once classified. The values or valuations are strictly speaking not shared, since their qualia remains subjective, even though the membership to a classification is shared.

The discussion regarding classification has a spill over into the discussion of the nature of values. Schwartz (2012: 3) argues that values are *beliefs*. He may, at least in part, be lead to conceptualize values this way due to his apparent confounding of classification and values proper, as described above. After all, if we *believe* in value classifications, it seems they must correspond to beliefs insofar as the endpoint of the activity of believing must be an object of belief. There is nothing wrong with forming such beliefs; indeed it seems to be a conventional example of how a belief might arise in our mind. A mistake arises, however, with the deliberate theoretical identification of value classification categories with values.

The sense in which this process involves a mistake can be reasonably debated, however, because after all, there is a way to resolve the mistakenness: simply to annex the

concept of “values” and baptize it “belief.” An argument against this option, however, is that it is useful to keep the concepts of “values” and “beliefs” distinct. Making the distinction allows us to see values and beliefs as corresponding to different underlying mental phenomena. Specifically, “values” usefully refer to “affects,” whereas “beliefs” refer to “cognition.” The usefulness of distinguishing affects and cognition has already been argued (Ch. 2). If we conflated the two terms, “beliefs” and “values,” it seems we would need a new concept to replace the concept of “values” that has gone out of fashion. Schwartz may have a reason to keep the old concept, however, because Schwartz (2012: 3) himself argues that values are “linked inextricably to affect.” If we grant that this is the case or an appropriate conception, characterizing values as “beliefs” seems to push the concept of values in two different directions that are mutually incompatible.

A reminder seems appropriate at this point. As argued extensively (Ch. 3), I am *not* claiming that values do not have objective or natural underpinnings. What I claim is rather that this fact, which I think we should suppose for reasons already stated (Ch. 3), does not imply that the values themselves are objective or possible to capture on a fully physical, naturalistic account. I believe it is not known empirically how the conversion from objective physics to subjective experience happens. What we do know, however, is that the subjective, qualitative aspects of experiences appear impenetrable from the external scientific perspective. We also know, however, quite a lot about the objective neural correlates of specific classes of subjective mind states (Ch. 4). What we cannot objectively know by observation is how experience, affects, and values are *like* internally, which is equivalent to saying that these mental states are *inherently subjective*.

8.4 Conclusion

It is argued that, in light of the body of evidence provided by affective neuroscience and theoretical perspectives on motivation, there is a strong case for internalism regarding values.

There is a corresponding reason for skepticism about theoretical perspectives that postulate the existence of objective values. It follows that it is unwarranted to claim that moral motivation and moral justification stem from objective values. It also follows that business ethicists should abstain from deploying normative terms that imply objective values or external normativity.

Chapter 8 References

- Anderson, J. (1997) What cognitive science tells us about ethics and the teaching of ethics. *Journal of Business Ethics* **16** (3): 279–291.
- Balch, D. R. and Armstrong, R. W. (2010) Ethical marginality: The Icarus syndrome and the banality of wrongdoing. *Journal of Business Ethics* **92**: 291–303.
- Bergson, H. (1903) *An introduction to metaphysics* (T. E. Hulme, Trans.). London: Macmillian and Co.
- Blackburn, S. (2010) *Practical tortoise raising and other philosophical essays*. New York: Oxford University Press.
- Blackburn, S. (2005) *Truth: A guide for the perplexed*. London: Penguin.
- Blackburn, S. (1998) *Ruling passions: A theory of practical reasoning*. Oxford: Oxford University Press.
- Bowie, N. (2013) *Business ethics in the 21st century*. Berlin: Springer Science & Business Media.
- Brink, D. (1989) *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.
- Broome, J. (2013) *Rationality through reasoning*. Oxford: Wiley-Blackwell.
- Copp, D. (1995) *Morality, normativity, and society*. New York: Oxford University Press.
- Crisp, R. (2006) *Reasons and the good*. Oxford: Oxford University Press.
- Cromwell, H. C. and Panksepp, J. (2011) Rethinking the cognitive revolution from a neural perspective: How overuse/misuse of the term “cognition” and the neglect of affective

- controls in behavioral neuroscience could be delaying progress in understanding the BrainMind. *Neuroscience and Biobehavioral Reviews* **35**: 2026–2035.
- Crumley II, J. S. (2009) *An introduction to epistemology* (2nd ed.). Toronto: Broadview Press.
- Damasio, A. (2011) *Self comes to mind: Constructing the conscious brain*. New York: Random House.
- Dancy, J. (2004) *Ethics without principles*. London: Clarendon Press.
- Etzioni, A. (1999) Normative-affective factors: Toward a new decision-making model. In A. Etzioni (Ed.), *Essays in socio-economics* (pp. 91–119). Springer Berlin Heidelberg.
- Frankfurt, H. G. (1988) *The importance of what we care about: Philosophical essays*. Cambridge: Cambridge University Press.
- Frith, C. (2007) *Making up the mind: How the brain creates our mental world*. Oxford: Blackwell Publishers.
- Goldman, A. I. (2009) *Reasons from within: Desires and values*. Oxford: Oxford University Press.
- Goldman, A. I. (2008) The case against objective values. *Ethical Theory and Moral Practice* **11**: 507–524.
- Goodwin, G. P. and Darley, J. M. (2008) The psychology of meta-ethics: Exploring objectivism. *Cognition* **106** (3): 1339–1366.
- Hart, S. L. (1971) Axiology—theory of values. *Philosophy and Phenomenological Research* **32** (1): 29–41.

- Helm, B. W. (2001) *Emotional reason: Deliberation, motivation, and the nature of value*. Cambridge: Cambridge University Press.
- Kagan, S. (1998) *Normative ethics*. Boulder: Westview Press.
- Lines, R., Sanz, J., and Aramburu, N. (2011) Organizational learning as a by-product of justifications for change. *Journal of Change Management* **11** (2): 163–184.
- Mackie, J. L. (1977) *Ethics: Inventing right and wrong*. London: Penguin Books.
- Moore, G. E. (1903) *Principia Ethica*. Cambridge: Cambridge University Press.
- Nagel, T. (1997) *The last word*. New York: Oxford University Press.
- Nagel, T. (1986) *The viewpoint from nowhere*. New York: Oxford University Press.
- Palmer, D. (2013) The new perspective on organizational wrongdoing. *California Management Review* **56** (1): 5–23.
- Panksepp, J. (2005) On the embodied neural nature of core emotional affects. *Journal of Consciousness Studies* **12**: 158–184.
- Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.
- Parfit, D. (2011) *On what matters: Volume 1*. Oxford: Oxford University Press.
- Parfit, D. (1984) *Reasons and persons*. Oxford: Clarendon Press.
- Prinz, J. J. (2007) *The emotional construction of morals*. Oxford: Oxford University Press.
- Rawls, J. (1971) *A theory of justice*. Boston: Harvard University Press.
- Rokeach, M. (1968) *Beliefs, attitudes, and values*. San Francisco: Jossey-Bass.

- Schwartz, S. H. (2012) An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture* **2** (1): <http://dx.doi.org/10.9707/2307-0919.1116>
- Skinner, B. F. (1938) *The behavior of organisms: An experimental analysis*. Oxford: Appleton-Century.
- Solms, M. (2013) “The unconscious” in psychoanalysis and neuropsychology. In S. Akhtar and M. K. O’Neil (Eds.), *On Freud’s “The Unconscious.”* London: Karnac Books.
- Solms, M. and J. Panksepp (2012) The “id” knows more than the “ego” admits: Neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Science* **2**: 147–175.
- van Fraassen, B. C. (1973) Values and the heart’s command. *The Journal of Philosophy* **70** (1): 5–19.
- Wedgwood, R. (2007) *The nature of normativity*. London: Clarendon Press.
- Williams, B. (1981) *Moral luck*. New York: Cambridge University Press.

Chapter 9: The emotivist position

9.1 Introduction to the emotivist position

9.1.1 General overview of emotivism

Emotivism is the central philosophical position of the thesis in that it connects the main arguments presented in previous chapters. Its principal claim is that practical ethics should be understood not as concerning matters of fact but rather regarding *expressing* and *acting on* emotions. This is the primary theoretical commitment that emotivism makes. I shall take this to be the *core sense* of emotivism. It suggests that external moral behavior can be explained in terms of internal mind states. I defend a more specific version of emotivism, which holds that human morality is expressive of and motivated by biologically based affect and that this affect specifically accounts for the normative core of morality (as was established in Ch. 5).

A second feature of emotivism that I endorse is psychological *projectivism*, which proposes that affective emotions, aided by cognitive processes, are projected onto perceived and imagined objects. Projectivism offers a way of explaining why these objects are seen as, or felt to be, normatively good/bad, better/worse, appropriate/inappropriate, virtuous/vicious, right/wrong, attractive/repulsive (or similar conative pairs). While I do not claim that projectivism is a necessary theoretical commitment for emotivism, I argue that it should nevertheless be endorsed because it has explanatory values and because it appears to be empirically adequate in light of cumulative evidence about how the brain works (along the lines explained in previous chapters). The perception of objects as “good” or “bad” (etc.) can be either implicit (non-declarative) in behavior and expressions or explicit (declarative) as in consciously formulated language. Most versions of philosophical emotivism have not paid attention to implicitly emotional expressions.

I argue that emotivism represents the fundamental explanatory account of normativity. At the same time, I suggest that this affective normative content can be suppressed or overridden by excessively rigid cognitive conceptualization (or misconceptions) of ethics. Examples of such rigid cognitive structure of thought can be found in fundamentally rule-based, objectivist, as well as relativist approaches to understanding ethics. Along the lines I have argued in previous chapters (especially Ch. 7 and Ch. 8), I consider these rigid approaches to ethics misconceptions essentially resting on a mistaken understanding of the nature and origins of human morality or a failure to appropriately appreciate the implications of such an understanding. I argue that error-theory adequately explains the mistaken understanding of ethics but that error theorists characteristically fail to draw appropriate implications from their own insight.

Furthermore, I hold that by exposing the scientific underpinnings of normativity, the emotivist position is strengthened. Emotivism is thereby rearticulated as a wider interdisciplinary position that can be more readily investigated by the empirical sciences. Thus, emotivism is extended beyond the confines of the philosophical and is rendered accessible to empirically oriented disciplines, such as psychology and neuroscience, subsequently informing practical decision-makers. Similar theoretical and practical reorientation can be found in psychiatry (e.g., Fulford, 2011) under the paradigm “value-based practice.”

In the course of the preceding chapters, a number of connected arguments have been presented. This chapter, I hope, makes the interconnectedness more explicit. While making the case for emotivism, I also emphasize addressing the major opposing views and counter arguments. I argue that there is a case for reviving the long-standing philosophical position that emotivism represents as a scientific position. Emotivism is often called “expressivism” in contemporary philosophy (e.g., Ridge, 2006; D’Arms and Jacobson, 2000; 1994; Blackburn,

1998; Gibbard, 1990). The two terms are normally considered to be equivalent and can be used interchangeably. Below, I explain my reasons for preferring the traditional term “emotivism.”

9.1.2 Terminological and definitional issues

As previously noted, there are two terms for the position in question: the classical term “emotivism” and the more recent “expressivism.” My reason for using the classical term “emotivism” is threefold. First of all, it appropriately highlights the intellectual roots of the emotivist tradition. Secondly, the connection with emotions is of special interest in my explication and defense of the position, and “emotivism” expresses this focus more succinctly. Thirdly, there are some theoretical disputes regarding what counts as “expressivism” that do not affect the term “emotivism” (which will be explained in the last paragraph of this section). Finally, it is important not to think that emotivism is merely concerned with the expression of internal states, such as desires and values, but also with how these desires and values attach to perceived objects in the external world, particularly including the expected consequences of prospective decisions and actions.

This last point is important because emotivism seeks to determine the normative, practical side of moral deliberation, and there is a moral difference between aiming to merely express desires and values and aiming also to promote and realize those desires and values: for example, the difference between expressing emotions of anger, remorse, and resentment that the innocent are killed vs. trying also to prevent innocent killings through decisions and resolute actions.

The term “expressivism” is sometimes used more widely than “emotivism.” For example, Sinclair (2009: 136) holds the following view: “A position is expressivist to the extent that the linguistic function of the target discourse is to express mental states.” On this

definition, expressivism can involve any type of mental state, cognitive as well as affective. In order to avoid any confusion, I shall raise no issue with the terminology of “cognitive” versions of expressivism (e.g., Horgan and Timmons, 2009; 2006), and I will not discuss these particular versions of expressivism. Instead, I reserve the term “emotivism” for expressivism that corresponds specifically to emotional states of mind. As argued extensively in previous chapters (especially Ch. 2, Ch. 4, and Ch. 5), emotions are most plausibly interpreted as affective, and therefore it will be reasonable to consider emotivism as concerning specifically *affective* emotions.

9.1.3 The components of emotivism

Emotivism represents a family of *meta-ethical* positions about ethics. The three basic claims that emotivist theories seek to defend are (1) *psychological non-cognitivism*, (2) *motivational internalism* about moral thought, talk, and action, and (3) *psychological projectivism*. These three claims are interconnected in several ways.

Psychological non-cognitivism holds that normative thought, talk, and action (hence, the topic of ethics) expresses or represents states of mind that, with respect to them being normative, are not merely cognitive (Helm, 2007; Blackburn, 1998). Instead, normative states of mind also involve an affective dimension represented by desires, hopes, wishes, and values, which, as argued in previous chapters (Ch. 2, Ch. 4, and Ch. 5), are all underpinned by affective emotions. Hence, “non-cognitivism understands our sensibilities to be ontologically prior to the value things have” (Helm, 2007: 17).

Motivational internalism, meanwhile, holds that the motivation for moral thought, talk, and action is internal to the subject and as such inherently subjective (Goldman, 2009; Blackburn, 1993; Blackburn, 1984: 324–333). By saying that motivation is internal and subjective, I do not claim that it is necessarily present in the conscious awareness of the

subject; as argued previously, many feeling states do not become explicit in awareness (as in the case of implicit normativity).

There are several versions of motivational internalism. The specific version in focus is broader than versions that only concern the status of moral judgments (e.g., Helm, 2007: 164). Well-articulated accounts of motivational internalism go as far back as to David Hume, who famously contrasted “beliefs” about how the world is with “passions” or “desires” about how the world ought to be. He argued that beliefs by themselves are incapable of motivating action, whereas passions by contrast are necessarily motivating. This, Hume reasons, provides an account of moral judgments because they need to follow motivation. For example, it would seem incoherent to hold the moral judgment: “stealing is morally wrong” together with the motivational attitude, “I approve of theft.

Basic and sincere moral judgments, such as Stealing is morally wrong, on this account, arise from a *disapproval* of theft.

Psychological projectivism explains why objects that feature in our moral deliberation and consideration are perceived as possessing or “radiating” normative qualities and valence (Helm, 2007: 17; Blackburn, 1998). Projectivism holds that these objects reflect our emotions. Recent neuroscientific accounts of how the brain work supports this type of basic psychological projectivism (Solms and Turnbull, 2002); objects in the external world are represented by cognition and held in the mind as “mental solids” but obtain their meaning in virtue of the evaluation afforded by affective emotions and affective consciousness. Hence, I argue that there is a direct connection between psychobiological affects to our realization that *de gustibus non est disputandum*; matters of taste cannot be discussed correctly as matters of facts about the world (Hume’s famous distinction). The claim is that, values and valence only seem to be in the world because we have projected them into it.

Meta-ethics can, as noted (Ch. 1), be contrasted with normative ethics and with descriptive ethics. While descriptive ethics is simply about what ethical practices and thoughts there are, meta-ethics aim to deepen the descriptive surface accounts by presenting coherent and accurate ways of understanding and explaining them. Meta-ethics does not itself represent any particular normative ethical position; it is instead *about* normative positions (Blackburn, 1998). Meta-ethics raises questions such as these: What do normative positions say? What do they mean? What do they amount to? What do they imply? What is the nature of values, virtues, and other normative notions, that positions normatively rely upon?

There is a consideration that slightly complicates matters as far as disentangling meta-ethics and normative ethics in concerned. It is the following. Proclaiming facts *about* what ethics can be (or can consistently mean) can in turn restrict the scope of which normative positions are available as viable positions. This complicating factor makes it appear as if emotivism itself; sometimes issues in normativity, which it is not supposed to do. In the weak, constraining, sense I think it is fair to claim that indeed it does. I shall have more to say about how meta-ethics constrain normativity below in the section about emotivism and moral reasoning.

9.2 A review of emotivism

Emotivism is an epistemological philosophical perspective on ethics that refers to the psychology of the mind. Its central contention is that ethical thought and talk expresses emotions (Blackburn, 2010; 1998; 1993; Gibbard, 1990). The intellectual roots of emotivism are historically very old. It is commonly counted as one of the intellectual achievements of Western philosophy, but in an earlier chapter (Ch. 2) I presented a passage by an ancient Chinese philosopher, Mencius (*Mèngzǐ*), which appears to fit right into an emotivist perspective. It is perhaps best to view emotivism as a position that has many disparate sources, rather than a view that springs from a single fountain head. It would be somewhat

misleading to suggest that emotivism is characteristic of Western philosophy, since Western philosophy has produced at least as influential lines of thought which stand in opposition to emotivist explanations; consider for example all the philosophical footnotes referring back to Plato.

9.2.1 A brief historical note

Though precursors for emotivism can be traced back to ancient times, emotivism as a distinct tradition has roots in the Western philosophy of science and moral philosophy. In the philosophy of science literature, early emotivists (notably Ayer, 1936) were to a less extent motivated by understanding ethics and more interested in pursuing the project of making scientific inquiry “value free” (as part of the wide project of expounding positivism). In contrast, moral philosophers have been concerned more directly with understanding ethics, the place of values in our deliberative lives (Sattris, 1987), and in particular the function of our moral language (Blackburn, 1998; 1984; Stevenson, 1944). Early emotivist moral philosophers (e.g., Stevenson, 1963; 1944) were driven, in part, by a need to understand ethics in a different way than the two opposing meta-ethical positions, naturalism and non-naturalism (Baier, 1967).

Both naturalism and non-naturalism are discussed in this chapter, but I will highlight the essence of the problems related to them upfront. The main problem with meta-ethical naturalism is that it tries to reduce normativity to mere natural facts. As dual-aspect monism (discussed in Ch. 3) indicates, this is problematic despite the fact that there clearly is a correspondence to natural facts that underpins normativity (e.g., the role of facts about the brain in relation to creating the mind). It is problematic because the subjective, experiential quality or “qualia” of normativity, including what it means, cannot be adequately explained on a naturalistic basis alone. If normativity is reduced to natural facts, it is hard to explain its inherent motivational and experiential grip on us (i.e., our inner affective life).

The main problem with meta-ethical non-naturalism, in its classical form (e.g., Moore, 1903), is that it does not discard the commitment to facts in explaining normativity. Instead, it proposes the existence of meta-physical “non-natural facts,” empirically inaccessible save by a special ability (that supposedly some people were well equipped with) called “intuition”. Many philosophers and scientists still believe that the idea of a world of “non-natural facts” supervening on natural facts and the concept of “moral intuition” is overly speculative as an explanation of ethics and normativity. I will return to moral intuitionism below.

In this historical perspective, emotivism may have arisen at least partly out of a need to rethink the explanation and understanding of normative ethics because what was proposed by philosophers already was largely seen as inadequate (Baier, 1967: 140). While the historical perspective explains many of the important philosophical debates that have taken place, we should at the same be careful with pinning a generic type of moral view too tightly to a particular historical stretch of time. Scientific advances sometimes allow us to think in new ways about new facts that were not on the table before, but it is clear that the hunches and speculations about human nature by some of the pre-modern as well as pre-enlightenment philosophers roughly draw the contours of an emotivist perspective.

9.2.2 Variations of emotivism

There are many versions of emotivism. I will comment on some of the most distinctive and well-known varieties and, when appropriate, relate them to some of my own views on the topic. Many emotivist theories are mostly historically important, while some remain important in contemporary philosophy.

I outline my own version of emotivism later in this chapter. One of the important differences between my view and other versions of emotivism is that my version is not merely a philosophical account. I introduce a major scientific component to elucidate how emotions

explain the normative content of ethics. This marks an important departure from many of the central emotivist theories that have been discussed over the last century and into the current one. In particular, most of the early emotivist theories were almost exclusively concerned with the intersection between philosophy of *language* and *logic* (e.g., Ayer, 1936; Stevenson, 1963; 1944; 1937; Urmson, 1968), whereas I hold that emotivism must cover a wider area of human activity than this in order to explain normativity, notably including non-verbal emotional expressions and the internal mental activity of the normative mind. “Research on emotional speech and language has recently regained immense popularity in the neuroscientific community” (Kotz and Paulmann, 2011: 108), alongside rapidly accumulating understanding of the neurobiological underpinnings of emotional expressions (e.g., Panksepp, 1998). This suggests that classical emotivism may need to be updated accordingly.

By providing criticism of emotivism using emotivism, Urmson (1968) helps us to see how emotivism has evolved. He considers Ayer’s (1936) version unnecessarily extreme in its insistence that ethical expressions only express emotions to the exclusion of everything else. At the very least, what ethical expressions express is a complex empirical question for psychology to investigate. Likewise, Urmson (1968) distances his account from the works of Stevenson (1963; 1944; 1937), mostly because he finds Stevenson’s analytical distinctions insufficiently sharp and sometimes inappropriate. For example, he claims that Stevenson’s account of ethics fails to distinguish ethical *argument* from mere *persuasion*, and that he confuses the *vagueness* of emotive terms like “good” with *ambiguity* (Urmson, 1968: Ch. 7).

Urmson (1968) draws attention to an important distinction between different types of evaluations. He proposes that we should distinguish between “good of a kind” and a more general and primitive concept of a “good” fit to serve as an explanatory basis of ethics and illustrates this with a number of examples where such a distinction is appropriate. He calls these two kinds of evaluations “evaluation things as of a kind” and “evaluating things from a

perspective,” respectively. For example, he contrasts statements such as “that is a good apple (knife, car)” (Urmson, 1968: 99) from statements such as “this road (period of weather) is good from the farmers’ point of view” (Urmson, 1968: 100).

He proposes “in judging good of a kind we assess something as members of a class that not everything could fulfill; whereas in judging good from a point of view we judge from a point of view from which everything could be judged” (Urmson, 1968: 104). The former type of judgment seems to depend on a cognitive apprehension of some specified norm, standard, or function, whereas the latter does not. This distinction may perhaps blur in some cases, and it may be unclear how far the notion of “a kind” can be stretched; however, it seems that the distinction points to an important psychological difference between normativity that is conditional on categories and normativity simpliciter.

This is closely analogous with ethics as conceptualized in terms of discrete rules, categories, and standards (e.g., Kant) on the one hand and ethics that is based on emotional evaluation on the other. Pushing too many items into a conceptualization of “goods of a kind” may, as indicated earlier (Ch. 6), represent a threat to affective emotional input to decision-making. For example, should we think of “a good decision” as a “good of a kind” conceived of as one that conforms with rigidly set rules of rationality? If so, it challenges us to find an *objective justification* for these rules.

Criticism of emotivism tends to take this tight philosophical disciplinary focus on language and logic as a premise for discussion. This premise is to some extent justified by the fact that most of the early past-century emotivists in fact *did* constrain themselves to such a narrow focus. However, there is nothing inherent in the basic tenets of emotivism to suggest that it is committed to an exclusive focus on language and logic. Indeed, although formal language usage is important, it can be seen as one among several channels of moral

expression. In practical moral life there is, for example, also non-articulated contextual, situational, and relations cues for the interpretation of linguistic expression, in addition to sounds that may carry emotive meaning but that do not classify as language, as well as a range of intended and unintended behavioral expressions. In general, the interpretation of meta-ethics as a study of linguistics has not been particularly fruitful in moral philosophy (Williams, 1985: 72).

Moreover, concerning understanding of the actual *role* of emotions in moral psychology and moral deliberation, the linguistic focus seems to represent only the tip of the iceberg. As noted earlier (Ch. 5), many of the emotional processes that act on our moral motivation escape awareness and attention. Insofar as understanding normativity is a central aim of meta-ethics, a linguistic-only approach is limiting, as Stroll (1954) rightly concludes in his critical review of emotivism. The underpinning psychological and biological reality surely warrants at least some attention. Even nowadays, little relevant empirical knowledge has informed the discussion of emotivism vis-à-vis alternative theories. Again, I do not mean to say that linguistic analysis is somehow misplaced. However, it is regrettable that it has become the dominating and almost exclusive focus. This is unfortunate because an abundance of relevant information about the nature and role of human emotions is, as I have attempted to show, currently made available by affective neuroscience.

While the emotivist (expressivist) theories of Blackburn (1998) and Gibbard (1990) indeed *do* take the psychobiological underpinnings of ethics seriously, they are forced to battle almost exclusively against antagonists preoccupied with language and logic. Moreover, the empirical basis of Blackburn (1998) and Gibbard (1990) is mainly drawn from psychology without any noteworthy attention from the neuroscience that scientifically grounds the various psychological theories. As I have shown (Ch. 2, Ch. 4 and Ch. 6), neuroscientific underpinnings are crucial for understanding moral psychology in an empirically adequate

manner. Indeed, I hold the view that an emotivist theory should ideally be built in a bottom-up fashion, starting with what we know from biology and neuroscience.

Gibbard (1990) focuses specifically on the expression of “plans” or “planning attitudes.” He plausibly argues that plans are *intentional*. When we use normative expressions, we propose the endorsement not just of particular values and emotions but of wider planning schemes in which such values are couched. This is potentially a useful insight, insofar as it may be useful for practical decision-makers to carefully reflect upon how their normative statements relate to wider plans and policies. Since plans and policies pertain not only to individuals but also the inter-subjective structures of understanding, this association relates decision-making to socially constructed morality and relational commitments.

Blackburn’s (1998; 1993) version of emotivism, which he calls “expressivism,” holds that normative sentences prescribe behavior or express values, emotions, or attitudes, as opposed to merely representing, reporting, or describing facts. He shows that emotivism is entitled to nearly all the objectivist-sounding rhetoric typical of moral realism (an auxiliary theory he calls “quasi-realism”; see Blackburn, 1993). For example, if someone makes the statement “it is objectively wrong to torture kittens,” it expresses both beliefs and values, but whether the beliefs involved are themselves true or false is entirely a different matter. If there is no “moral truth,” the beliefs will be false, but this neither changes the fact that these false beliefs feature among what is expressed nor the fact that there is also an emotional component (normative content, as I would say). Blackburn (1998) argues that it is not the surface meaning of ethical terms that is interesting with respect to considering the merits of the emotivist position but rather what these terms *actually* express; which is to say, the moral psychology and reasoning behind ethical expressions.

9.2.3 Related views

There are several meta-ethical views that bear kinship with emotivism. They have in common that they take emotions or affect as central to the understanding of ethics. The broadest class of related views is perhaps sentimentalism, which has many sub-varieties. Seminal sentimentalist theories include “subjectivism” (Wiggins, 1987), “emotionism” (Prinz, 2007), and the classical “ideal spectator view” propounded by Adam Smith (1759) in his seminal work, *The theory of moral sentiments*. Other related views include “moral factionalism” (e.g., Kalderon, 2005), “prescriptivism” (e.g., Hare, 1952), certain forms of “intuitionism” (e.g., Haidt, 2001), and to some extent, error theories (e.g., Mackie, 1977). Undoubtedly, there are still further theories to add to this list, but the ones mentioned are probably the best known. Below I will consider some of the similarities and differences with emotivism.

According to Smith (1759/2010: 12), “nothing pleases us more than to observe in other men a fellow-feeling with all the emotions of our own breast; nor are we ever so shocked as by the appearance of the contrary.” What makes behavior morally praiseworthy or blameworthy for Smith is how far from or close to the sentiments and the sentiment-based judgments of an ideal observer we are. In this conception, the ideal observer is in some sense remote because he is impartial yet also present in being guided by his sentiments in forming his moral judgments. Emotivism does not have to go by the laborious route of imagining anything like an ideal observer to arrive at the sentiments or emotions that properly ground ethics.

Emotivism differs from “emotionism” (Prinz, 2007) concerning the application of the realist notions of *truth* and *falsity* in ethics. Prinz (2007) argues that we can straightforwardly talk about truth and falsity at the meta-ethical level, as long as we understand this to be a

person-relative truth and falsity. Thus he holds: “A form of conduct is truly wrong for someone if that person has a sentiment of disapprobation towards it” (Prinz, 2007: 139). He furthermore makes the claim that: “We seem to intuit moral facts” (Prinz, 2007: 88). Emotivism rejects these ideas on the grounds that they hypothesize a universe of metaphysical moral properties and facts that do not correspond to those properties and facts of the real world that are supported by empirical science.

However, emotivists do not necessarily hold that people’s ordinary first-order talk of “truth” and “falsity” in ethics is misguided or unjustified (Blackburn, 1998; 1993). The question is what meta-ethical *status* such talk has. According to emotivism, such realist-sounding talk expresses or manifests affective emotions, not genuine facts (metaphysical or otherwise) in themselves. Affective emotions are often *direct at* or *responding to* facts (or perceived or imagined objects) but involve second-order subjective qualities regarding the facts rather than the first-order objective qualities of the facts themselves.

Moral fictionalism (e.g., Kalderon, 2005) proposes that what moral statements express are neither desires nor beliefs but rather what people *pretend* to believe. However, it seems inaccurate to hold that people who believe in “moral truth” merely pretend to believe. It does not follow from the fact, as Kalderon argues, that there *is* no moral truth or that the *beliefs* in such (putatively false) moral truths are somehow not *genuine*. Notions such as “moral truth” therefore are not characterizable as “fictions” as Kalderon (2005) and other fictionalists presuppose; “moral truths” are not “made up” but are genuinely *believed* in by moral realists. Emotivism does not need to postulate such moral fictions in order to accept the presence of erroneous beliefs. That said, emotivists can usefully lean on *error theory*—which I believe to be fully compatible with emotivism—in order to explain such putatively misattributed beliefs.

It is important to distinguish between intuitionism and emotivism. While the classical philosophical intuitionists (G. E. Moore, H. A. Prichard, W. D. Ross) and their contemporary followers (e.g., Dancy, 2004; Audi, 2004; 1996) tend to differ markedly from emotivism, certain variations of intuitionism overlap significantly with the emotivist theoretical domain (e.g., Haidt, 2001; Bergson, 1946). Bergson (1946: 200), for example, views intuitions as *experience* conceived of as an “infinite series of conscious acts.”

Unlike in the classical philosophical accounts of intuitionism cited above, intuition according to Bergson is not perception. Instead, it denotes an affective, effortful, and experiential phenomenon required to live in an internal and subjective sense of “duration” (as opposed to the external reality of the objective passage of time). Without claiming that Bergson himself was an emotivist, his views concerning intuition do seem consistent with the basic tenets of emotivism. Care needs to be taken to differentiate between the versions of intuitionism that implicitly or explicitly reject emotivism from versions that, at least in principle, are open to its acceptance.

Tersman (2008: 391), for example, defines moral intuition as “a judgment that is accepted by someone not merely on the grounds that he realizes that it follows from some moral theory or principle that he also accepts.” This general definition of intuition is in principle open to an emotivist explanatory account of what, more specifically, an intuition consists of. For example, an “intuition” could be characterized as a *normative verdict* emanating from an affective emotional process and thus as the endpoint of this process. Similarly, Haidt (2008; 2001) uses the terms “moral intuitions” and “moral emotions” more or less interchangeably. One advantage of relating intuition closely to emotions is that one is not committed to the metaphysical baggage of classical intuitionism (Tersman, 2008). Another advantage is that intuitions can be connected to emotions as *natural* phenomena with natural substrates in the brain, many aspects of which can be investigated using scientific approaches.

When it comes to *actual* moral practices, the variety of usages of moral terms that different individuals engage in at different times, in different contexts, and in different states of mind is an *empirical* question for psychology and neuroscientists to map (at the group level, sociology will also be relevant). Given the plurality of individual values, social identities, and cultural-linguistic norms, it seems prudent to assume that there is considerable variety in the usage of moral terms. As argued earlier, mapping out what descriptive moral practices are prevalent does not permit us to draw direct normative conclusions about what practices people ought to engage in. However, it represents an important corrective input for *meta-ethical* analysis.

As an approximation to actual moral practices, we can assume that individuals sometimes use moral terms to *express beliefs* in moral truths, to *prescribe behavior*, as *pretenses* or guises, to *express desires* and values, to express *plans or intentions*, or simply *without any particular motive* that they are consciously aware of, such as when people merely follow linguistic conventions. An important argument that I wish to emphasize is that there is a more fundamental sense in which all the above are based on the normativity embedded in our affective natures.

For example, believing in moral truths seems to presuppose the holding of some externalized conceived objects as morally *important* and *worthy* of attention. Prescribing and pretending are two very different strategies in which moral terms serve as tools, but in each case there is an underlying *purpose*. Importance, worthiness, and purpose are all normative and affective notions. Since desires, values (i.e., second-order desires), plans, and intentions are all inherently affective, it follows that their expressions must also be. Arguably, even more thoughtless linguistic rule-following springs out of some measure of affective attachment to these linguistic rules. If my argument that the normative and the affective are approximately the same is correct, and my account of human affective nature as presented in previous

chapters is roughly right, it seems to follow that there is a deeper level of emotivist explanation behind all linguistic and overt moral practices at the surface.

9.2.4 Cognitivism and non-cognitivism

Emotivism implies *non-cognitivism* about the nature of values or value and, more specifically, that the non-cognitive nature of values is captured by *emotions*. The term “value” (or disvalue) in this context serves as a placeholder for inherent normativity and can be linked to natural phenomena, such as mental states satisfaction or eagerness, excitement, or any other state of mind that is positive (or negative, in the case of disvalue). It should be noted that emotivism is not the only form of non-cognitivism for values. It is the specific type of non-cognitivism that grounds morality and moral motivation, or psychological normativity, in emotions.

Emotivists or expressivists typically take the nature of emotions to be *affective* (e.g., Blackburn, 1998; 1993). However, some argue for a broader usage of the term “expressivism” on which it is *not* committed to non-cognitivism and affectively based accounts. Sinclair (2009: 136), for instance, suggests that a “position is expressivist to the extent that the linguistic function of its target discourse is to express mental states.” Since any mental state will do here, there can also be *cognitive* forms of expressivism, and this version is exemplified by authors such as Horgan and Timmons (2006).

Emotivist theories conventionally refer to states of mind, but it is important to note that their claim need not be and that these states of mind, psychologically complex as they may be, are *purely* non-cognitive or affective (Ridge, 2006). The important point is that certain *aspects* of the states of mind are non-cognitive and that these aspects play a *critical role* in moral thought, talk, and action. While emotivism therefore does not imply a denial of cognitive mental processes, it is *affect* that accounts for normativity by generating the content

of our normative experiences: their vividness, positive or negative valence, and distinctive emotional coloration.

What do emotivists object to in cognitivism regarding values? In everyday conversation, people often say that things are “right” or “wrong,” with various grades of moral intensity. Since moral realist cognitive talk like this seems quite common, what could be wrong about taking it at face value? Consider an ethical statement such as “stealing is wrong.” The philosophical-scientific question of dispute is what type of *state of mind*, or *brain function*, is involved in this sort of claim. Cognitivists hold that what ethical statements refer to are certain beliefs. Thus, saying that “stealing is wrong” either *expresses* a belief that it in fact *is* wrong or *describes* what one believes are *facts* about the external world. Beliefs, from this perspective, can be true or false (or, more or less accurate), and so statements such as “stealing is wrong” are either true or false.

Emotivists, by contrast, hold that phrases like “stealing is wrong” on the one hand do *not* necessarily express any cognitive beliefs at all but on the other hand *do* necessarily express states of mind (and corresponding brain processes) that are non-cognitive or, more specifically, *affective*. While moral language presents the clearest cases, analogous interpretations can be made for moral action. For example, a moral act of theft prevention (say, blocking a thief at the door) could be interpreted as being motivated by an affect.

Let us look at a particular example that is usually taken to support the cognitive view. *Psychopaths* are considered to “know” what is “wrong,” but they do not *feel* this wrongness; for clinical reasons, they also may not be able to. It seems that if a psychopath says “stealing is wrong,” that would be a straightforward case of expressing a belief about what is morally wrong, just as cognitivism claims. However, to explain a psychopath is a small victory, for we

do not think of them as genuine moral agents, and it is of course moral agents that a meta-ethical theory must explain.

In addition to not coming close to explaining moral agents, it can be questioned whether even the explanation of psychopaths is adequate because, after all, do they really *know* what is morally wrong? They have beliefs about *what other people accept* as morally wrong, but the psychopaths themselves seem to fall short of genuinely accepting something as morally wrong. They are *unable* to access moral wrongness directly, and so they are left merely guessing about it. Emotivism provides a straightforward explanation for what it takes to morally accept something; it requires the involvement of *affective emotions*. This is what the psychopath *lacks*.

The essence of the dispute between cognitivism and non-cognitivism is related to the notion of (objective) “moral truth.” First of all, cognitivists often claim and seem committed to the claim that beliefs *track* some sort of *moral truth*. It is not necessarily known what precisely the moral truth is and, correspondingly, beliefs can be taken to track truth both imperfectly and imprecisely. The important thing for (most) cognitivists is that there is a “moral truth” out there in the external world and that it is possible with some rate of success (however small) to capture it through intuitive mechanisms or our perceptual-sensory apparatus in combination with our cognitive mental apparatus.

Several reasons for questioning the idea of a moral truth and the usefulness of this notion were presented previously (Ch. 8), so they will not be restated in full here. It will suffice to discuss some of the key objections that are particularly central in emotivism. Emotivists are generally skeptical about the claim that something like moral truth exists. One of the main reasons for this skepticism is that there is no credible body of evidence that establishes the existence of moral truth. It seems to elude any form of direct empirical

investigation by the sciences. A further reason is that it is difficult to provide a coherent and empirically satisfactory, and at the same time non-moralizing, idea about what this evidence could be. For example, if someone held a book in his or her hand and claimed “the moral truth is stated here,” we would not thereby be the slightest bit closer to anything that could reasonably be counted as scientific evidence.

It is striking that many philosophers (see Ch. 1) state their philosophical positions without any reference to psychology or evidence provided by the life sciences, even when they implicitly depend on such assumptions. There are indications that this practice is changing as philosophers increasingly opt for wider and more interdisciplinary approaches that relate to psychology, biology, and neuroscience (e.g., Prinz, 2012; Northoff, 2011; Prinz, 2007). However, the majority of philosophers still operate from *a priori* assumptions and conceptual foundations of normativity that do not intersect with the psychological facts about human nature. I now turn to briefly discuss the scientific underpinnings of emotivism.

9.2.5 Scientific underpinnings

The scientific underpinnings of emotivism follow from the discussions of previous chapters (especially Ch. 4 and Ch. 5). This can be readily seen if we consider the components of emotivism described above. *Motivational internalism* is underpinned and supported by the emotional circuitry of the brain. Current theory in affective neuroscience suggests that while many external perceptions are filtered by cognitive processes, it is the affective emotional structures of the brain that motivate responses (Panksepp, 1998). As noted in this chapter as well as in previous ones, motivated action does not require the involvement of conscious aware cognition. Moreover, I have argued that normativity corresponds to inherently affective phenomena that start out pre-cognitively as implicit normativity. It follows that *non-cognitivism* about moral action should be the default theory. Finally, *projectivism* gives an

empirically adequate account of how cognitive and affective processes in the brain cooperate for moral thought and deliberation.

Emotivism has traditionally been seen as a branch of *meta-ethics* or *epistemology*. However, as I have shown, emotivism is *not just* a philosophical perspective because it relies on *psychological* assumptions about the workings of the human mind, and as discussed (Ch. 3), these assumptions about the mind in turn rest on assumptions regarding the workings of the human *brain*. Philosophy and psychology of the mind therefore have to conform to and build upon scientific evidence specifically concerning the human brain (neuroscience) and, more broadly, human nature (the life sciences). These links between philosophy, psychology, and neuroscience are crucial for understanding normativity as a part of human nature, yet they are seldom drawn (Joyce, 2008; 2006: Ch. 2). The position of *neuroemotivism*, which is outlined below, aims to unify these disciplinary perspectives on normativity.

It is also important to recognize morality as *evolved*, although concrete evolutionary psychology cannot be pushed too far. Kitcher (2011: 74), for example, argues “the capacity for normative guidance, was an important step in the transition from hominoids to human beings.” This seems plausible and may well be correct, but clearly the details remain in the darkness of evolutionary history. While specific accounts of evolutionary psychology can be charged by overstating their case and presenting “just-so-stories” that cannot be verified scientifically (Panksepp and Panksepp, 2000), it seems clear enough that the human brain and the capacity for morality that it affords is a product of evolution. How these built-in capacities play out socially, moreover, also depends of course on cultural, social, and physical factors in a given person’s external environment. Each person has a life story that is full of random events that shape the unique psychology of that person. This psychology includes the particular storage and arrangement of cognitive content as well as which parts of the affective

infrastructure become sensitized and which emotions become central contributors to that person's experience of life.

9.2.6 Emotivism and moral reasoning

Meta-ethics can restrain the normative domain in a similar way as can, more generally, rational requirements (e.g., the requirement of consistent beliefs). Broome (2013: 1–22), among others, argues that there are specific rational requirements that regulate correct *reasoning*. Following the lead of Broome (2013), I shall hold that rationality in this sense concerns the *structure* of practical normative reasoning. In this view, there is a fact of the matter (which, however, is not fully known) as to what is correct and what is mistaken reasoning. I assume, along with Broome, that this last point about rationality and reasoning is correct.

However, practical normative reasoning, even when structurally correct, is in itself insufficient as a means to settle or justify normative questions because it is confined to dealing only with structural issues and not normative content. *Meta-ethics*, I argue, can potentially bring us closer to settling normative problems by imposing *restraints*, not on reasoning and the structure of normativity but on the *content* of normativity. However, meta-ethics represents only a weak constraint, which although it counts as improvement is still insufficient as a means to settle normative decision problems. The residual is the territory of normative ethics, which I argue should be fundamentally based on the affective experiences provided by our hard-wired affective consciousness and our emotional systems.

Emotivism, along with meta-ethics in general, does not make any direct normative claims about what one ought to value, do, or decide in a practical situation. Normative claims would need justification, and they are the domain of normative ethics. In cases where we want an internal justification, the inputs of affective content are required. As noted earlier (Ch. 5),

however, this content is normative only in the *psychological* sense. It is a *felt* sense of normative authority rather than a built-in objective normative authority; and in the final level of assessment, it has only the moral authority we *grant* it. In this sense, ethics is fundamentally subjective, and we are faced with the “problem of subjective normative authority” (Bratman, 2006: 95), to which I will return later in this chapter.

When we are *on the path* to justified ethics, content based on our biological affective emotions is brought to bear on our practical decisions, and the restrictions imposed by meta-ethics are in place, as are the regulation mechanisms afforded by the cognitive structures of practical reasoning. Each of these three elements can be blocked or fail to be harmonized properly. To the extent possible, given our limited or “bounded” free will (see Ch. 3), we ought to try to cultivate harmony.

Other routes to justification are external insofar as their justifying criteria lie outside the skin of a person’s body. This does not mean that many of them do not draw on internal mechanisms, such as states of mind however. Broome (2013) presents an admirably clearly articulated version of this type of theory, which is about internal motivation in the sense that the “motivation question is about your mind only” (Broome, 2013: 1). What justifies, however, is assumed to be aspects of the external world. Rationality, he holds (Broome, 2013: 3), requires people to intend to do and be disposed to do what they believe they ought to do.

From this perspective, moral motivation starts with a “belief about what one set out to do.” No affective underpinnings whatsoever enter the equation. Beliefs, according to Broome, are intuition-based. However, this moral intuition is not of the emotional kind like Haidt (2001) proposes but is rather a supposedly “cold” and “unmoved” perception of “goodness” in the external world, guided by a special “faculty” of reason.

In contrast, emotivism suggests that motivation is affectively based. As we have seen (Ch. 2 and Ch. 4), this is also the conclusion neuroscientists increasingly seem to arrive at, and for which there is ample neuroanatomical and neurochemical evidence. The point is not that cognitive processes do not play their part in motivation but that the part they play is not a solo act. Affective neuroscience, as presented in detail earlier, shows that affective input is required in order to provide direction as a compass on the map that cognition presents.

Broome (2013) appears to present a perspective that is purely philosophical, although it is also intended to have practical decision-making implications. Its supreme strengths lie in its clarity and analytical precision. Its main shortcoming, in my view, is in its failure to draw upon and suitably incorporate relevant evidence from the empirical sciences. I cannot find a single reference to either psychology or neuroscience. It advances a philosophical tradition using focus and clarity, but it fails to recognize that the empirical foundations of the tradition are unsound and some of the main supportive pillars of the theory lack any empirical basis altogether.

For example, on the first page and in the first sentence, Broome asserts: “When you believe you ought to do something, your belief often causes you to intend to do what you believe you ought to do” (Broome, 2013: 1). This is an empirical claim about how the psychology of the mind works, yet no references or any other sources of evidence are cited. Proceeding in this manner, it is easy to make a wrong step along the way. In this particular case, do your beliefs alone often cause you to intend anything at all?

The relevant difference between the normative and the meta-ethical perspective is whether we stand *within* our own evaluative sensibilities, as in the former perspective, or *outside* it, as in the latter (Blackburn, 2010: 29). Decision-makers involved in practical

business ethics make decisions from the internal, normative perspective because (revisit Ch. 6) they are faced with questions of what to do and decide and what to assign moral weight to.

Thus, if our aim is to *understand* normative positions, we are in the territory of meta-ethics. Emotivism represents an epistemological framework for understanding what normativity is and the moral psychology concerning moral deliberation, discourse, and action. “Moral psychology is the study of how the mind makes ethical judgments” (Thagard, 2007: 369), which according to the preceding discussion (Ch. 3 and Ch. 4) implies an aim of understanding the relevant cognitive and affective processes in the brain.

As the term suggests, *emotions* constitute the central explanatory theme for what ethics is fundamentally concerned with. In philosophy, emotivism also goes by other names such as “expressivism” (Blackburn, 1998; Gibbard, 1990) and “quasi-realism” (Blackburn, 1993). I will stick with the classical term, emotivism, because it more readily illustrates the historical lines. Another more important reason is that I wish to prevent the possibility of misunderstanding emotivism as *merely* being expressive; in some cases it plays a justificatory or an action guiding role. Terminology is not crucial; it is just a matter of communicative clarity.

This chapter presents a detailed account of emotivism. Some historical lines are drawn to get a sense of where this body of theory comes from. Emotivism is distinguished from and contrasted with related views as well as opposing meta-ethical views. Different versions of emotivism are also discussed; based on this discussion, a refined version of emotivism is proposed. While the suggested position does not detract from the account presented in Blackburn (1998), significant empirical detail is added to expose the biological foundations of emotivism. The most important aim is to show how emotivism connects with, builds upon,

and integrates the scientific and philosophical input from previous chapters. Consequently, I will start by drawing together some lines of thought.

9.3 Addressing arguments against emotivism

The critics of emotivism are many. Unfortunately, I will not be able to do justice to all of them. Due to space considerations, I have to limit myself to a selection of opposing positions. I will try to make up for this by including the most important and well-articulated opposing views and highlighting their central arguments succinctly. While the focus is mainly on philosophy because almost all the criticism directed at emotivism stems from philosophical accounts, I will also highlight some of the relevant psychological and scientific issues that underpin and inform the points of dispute. In some cases an understanding of human biology, albeit incomplete, provides resources for addressing objections to the emotivist view.

9.3.1 *The abandonment of moral truth and objectivity*

The fact that emotivism implies the abandonment of the notions of truth and objectivity in morality is perhaps the single most widespread point of criticism held against emotivism (e.g., Parfit, 2011; Kramer, 2009; Brink, 1989: 43–50). Moral realists and cognitivists of ethics tend to find this unacceptable either as a theoretical premise or as a conclusion. The specific reasons for rejecting emotivism differ considerably, however, between different opposing theories.

A number of central positions that defend or (implicitly or explicitly) rely on a notion of *moral truth* and *objective values* have already been discussed (Ch. 7 and Ch. 8). It was shown to be difficult to establish a solid foundation for these moral realist notions. Perhaps most critically, it appears that nothing can be reliably said about what would establish the truth or objectivity of these claims about normativity. Relatedly, it seems difficult to determine what exactly may count as moral truths and objective values and how decisions

about what counts could be made without any biasing involvement of subjectivity. As argued (Ch. 5), we cannot *escape* our subjectivity at will, so we are forced to live our subjective lives and see things as they appear to us.

The abandonment of the notions of “moral truth” and “objective values” can be seen as counting in favor of the emotivist position rather than against it. The reason for this is twofold: (1) the objectivist notions of normativity are questionable because they are not empirically verifiable nor in principle falsifiable (Goldman, 2009), and (2) we do not actually need them in practical ethics and moral decision-making (Blackburn, 2010) since we can rely on internal resources to establish normativity. It is less ambitious but also far simpler to do so without the claims to objective normativity. Emotivism therefore offers a sound alternative to what appears to be cases of wishful thinking and endorsement of overly speculative premises.

9.3.2 Thoughtlessness and gut reactions

Emotivism is sometimes criticized for presenting a picture where reasoning and meaning is substituted with thoughtless gut reactions (Millgram, 2005: 245). I hope to have shown by now that this dichotomous way of presenting affective emotions (i.e., gut reactions) and normative reasoning and meaning is problematic. As accumulated evidence suggests (see Ch. 6), affective emotions *can* indeed disrupt “orderly” reasoning and meaning in moral deliberation. This is especially true of strong and sudden bursts of emotions, or “eruptive emotions” as they are sometimes called.

Even quite literally thoughtless affective emotions nevertheless play a vital role in practical moral deliberation (see Ch. 6). Although all raw emotion is initially non-cognitive and non-declarative, it serves as input precisely to *infuse* reasoning processes with subjective meaning. Primary emotions, moreover, constitute necessary raw material by which cognitive processes shape more complex self-reflective and moral emotions (i.e., tertiary emotions) as

well as affiliated second-level moral attitudes and desires (e.g., toward and for cognitively represented objects).

While it can be pointed out (correctly) that the biological construction of normativity is “messy” or “imperfect,” this is arguably what we have. Alternative stories (1) of “neat,” “orderly,” “structured,” “chunks of normativity” somehow appear in our minds in unexplained ways; (2) normativity seeps into our minds like some special type of radiation that is not empirically accounted for; or (3) normativity can be constructed on the basis of pure logical inferences while entirely ignoring the pulls and pushes of affectively based motivation. These three viewpoints seem correspondingly implausible (see Ch. 8).

Finally, the messiness and imperfection of input factors into reasoning processes does not preclude these processes from allowing us to *impose* a helpful structure and order on our moral reasoning. As evident from our ability to give and receive rather detailed arguments, human decision-makers appear to have sophisticated abilities for creating “order” out of “chaos,” although some imperfections are liable to remain. Thoughtlessness at the foundational level of normativity, therefore, is what should be expected once we come to terms with the science of our biological origins and question alternative, more speculative, accounts of the origins of normativity. This may make morality less “glorious” but also more intimately human in return.

9.3.3 Motivation does not follow moral acceptance

Brink (1989: 46-47) presents what may on the surface look like a powerful argument against emotivism (a similar argument is made by Copp, 1995). As noted earlier in this chapter, one of the central theoretical pillars of emotivism is *motivational internalism*. In this view, moral talk and action are motivated internally. Brink (1989) holds that amoralists (e.g., psychopaths

or sociopaths) who know what is “morally right” show that the theory motivational internalism represents does not work.

I have already argued elsewhere that objective notions of what is “morally right” are flawed (see Ch. 5 and Ch. 8), but let us for the sake of argument imagine an amoralist who thinks that others in society perceive that a certain type of action in a certain type of situation is the morally right thing to do and also believes this action establishes what is objectively morally right. This person “knows” what is “morally right” but nevertheless fails to be appropriately motivated by what she thinks is morally right.

A reminder that seems appropriate here is that motivational internalism does not in fact claim or somehow require that moral motivation is successful. There may be other motivations that weigh in stronger, for example (Goldman, 2009), so a given moral motivation need not be decisive. However, a problem *still* seems to remain concerning the amoralist, for this person is assumed to lack moral motivation altogether.

Emotivists have come with various replies to this objection; I will not list them here. Instead, I will focus on what I consider to be the strongest reply. To my mind, the most powerful reply comes from looking at the psychological and neurological processes that underpin “acceptance.” In contrast with Brink (1989), I argue that *genuine* acceptance involves more than merely *recognizing* what (e.g., case-based dilemma resolutions, principles, laws, norms, virtue ideals) people in society or in the community seem to accept morally. Genuine acceptance involves not only the *cognitive* abilities of empathy and mentalizing but also *affective* attitudes. The cognitively aware psychopaths and sociopaths Brink (1989) alludes to *lack* these key affective emotional experiences. They lack the *guidance* that these normative-affective experiences would have provided. It is this shortcoming that, in these particular instances, explains the absence of moral motivation.

Hence, it is not a shortcoming of emotivism but a biologically based and psychological shortcoming of particular agents instead that *prevents* these people from experiencing normativity and thereby from being moral agents. Since the emotivist explanation concerns normatively functional moral agents, the specific type of agent in question (complete sociopaths and psychopaths) fall outside of the explanatory scope of the theory. The version of emotivism that I propose below, which I have dubbed “neuroemotivism,” is capable of addressing these types of objections in a concrete and empirically informed manner.

9.3.4 Emotivism as viciously circular

MacIntyre (1981) represents one of the central theoretical works within the neo-Aristotelian current of moral philosophy. In this theoretical exposition, the enlightenment thinkers and the emotivists are singled out for attack. His central argument is that contemporary moral discourse is in bad shape, and he directs attention particularly to influences made by the famous intuitionist philosopher G. E. Moore and the emotivists who came after him, notably A. J. Ayer. Here, I will focus specifically on his criticism of emotivism.

The problem, as MacIntyre sees it, is that the *rational* basis for ethics has been lost. Enlightenment thinkers such as David Hume and Edmund Burke, and the early emotivists inspired by them, hold that we do not need to base normative ethics on rationality. MacIntyre claims that (1) the influence of emotivism has been destructive, and (2) that emotivism is a viciously circular theory that has to be abandoned. MacIntyre (1981: 12) also holds that emotivists are unable to identify what is of relevance in ethics:

“For all attempts so far to identify the relevant types of feelings or attitudes it is impossible to avoid an empty circularity. ‘Moral judgments express feelings or attitudes,’ it is said. ‘What kind of feelings or attitudes?’ we ask. ‘Feeling or attitudes of approval,’ is the reply. ‘What kind of approval?’ we ask, perhaps remarking that

approval is of many kinds. It is in answer to this question that every version of emotivism either remains silent or, by identifying the relevant kind of approval as moral approval—that is, the type of approval expressed by a specifically moral judgment—becomes viciously circular.” MacIntyre (1981: 12), *After virtue: A study in moral theory*.

The claim here is that the arguments made by emotivists or that are available to emotivism necessarily lead to self-reflexive circularity. However, the argument MacIntyre makes is one of his own construction. The relatively obvious answer to his question about what kind of approval (or disapproval) emotivism focuses on is that it is the kind of approval (or disapproval) that is based in the agent’s *emotions*. This is presumably why emotivism is called “emotivism”; it refers to human emotions. It can be further specified what these emotions are and where they come from (see Ch. 4). By referring to an empirical basis in emotions, emotivism is not viciously circular in the manner that MacIntyre suggests.

Subjective feelings of approval and disapproval are based on bodily emotions, and the moral thought, talk, and action of individuals are an outcome of feelings that we are consciously aware of as well as subconscious affects. Why are these streams of reflective and non-reflective affects morally important? One available answer for emotivism is because they are constitutive of *what we care about*.

9.3.5 Emotivism implies positivism

Emotivism is often attacked as an ally of positivism or as part of the theoretical positivist scientific paradigm and also because it sprung out of the “Vienna circle.” Specific emotivist philosophers, notably Ayer (e.g., Ayer, 1936), have contributed to this view of emotivism. Ayer was a vocal supporter of logical positivism and focused his intellectual energies almost exclusively on the scientific debate and almost nothing on normative ethics.

However, emotivism neither implies positivism nor arose exclusively as a result of this scientific movement (Sattris, 1982). The conflation of emotivism and positivism is therefore misconceived. Sattris (1982) argues that many early emotivists, including notable figures such as Hägerström, Meinong, and Brentano, had nothing to do with the positivist movement. Furthermore, some of those who see themselves as positivists do not view themselves as emotivists. There is, moreover, nothing that commits emotivism to the scientific paradigm of logical positivism. The fact that the two perspectives are compatible does not in itself entail anything about there being a relation between them.

9.3.6 Doubts based on natural selection

Olson (2010) argues that *natural selection* makes it difficult to understand why our moral practices should be interpreted as emotivism (expressivism) proposes. According to Olson (2010), emotivism faces two explanatory challenges:

“Number one is that of explaining why natural selection – which, by expressivism’s own lights, favoured moral thought and talk because of their socially useful regulative and coordinating functions—did not favour a stance that would make moral thought and talk more effective in fulfilling these functions. Number two is that of explaining how moral thought and talk survive in cultural evolution, despite the prevalence of the freshman objection and related worries” (Olson, 2010: 87).

The freshman objection to emotivism is as follows: “If expressivism were true, anything would be morally permissible” (Olson, 2010: 88). *The freshman objection* implies erroneous reasoning from meta-ethics to normative conclusions precisely along the lines I previously discussed under the label *normative cynicism* (Ch. 7). Olson (2010) raises the question: If this is a common inferential error to make, why is that so? Why are human beings not adapted so as to avoid making this kind of “trivial” mistake of inference?

The two closely related points that Olson (2010) makes can be considered as relatively weak and indicative arguments against emotivism. After all, they are not actually about errors in emotivist theory itself but rather regarding errors of those who try to evaluate emotivism from the outside. If we idealize the forces of natural selection, where utility undeniably is often favored, perhaps we should thus expect such errors to have been polished away by the tooth of time? This view presents a picture of natural selection as highly efficient.

While it is evident that natural selection has come up with many “brilliant” and surprisingly efficient solutions to pressing practical problems facing the human species, it should also be emphasized that there are distinctive limitations, striking examples of maladaptation, and numerous byproducts of advantageous adaptations that are disadvantageous themselves. The human capacity for reasoning is a striking achievement of natural selection. It would not be surprising if there are side effects and limitations pertaining to this remarkable adaptation, which has taken place in a relatively short time span in evolutionary terms. As previously noted (Ch. 4), the remarkable cognitive abilities that humans have had to build on the older brain structures that were already there, like “scoops of ice cream on a cone.”

Regarding the point that questioned how expressive thought and talk could have survived cultural evolution, I think there are a number of socially complex items that could potentially feature in a complete explanation. This explanation is likely to include points about emotion, expressive thought, and talk either being effective and practically useful or not being too ineffective and costly. One possible explanation is that emotion-expressive thought and talk are so deep-seated and imagined in the human psyche and underpinning neuronal structures that the forces of cultural evolution were never strong enough to fundamentally alter these linguistic and epistemological features. The fact that (see Ch. 2 and Ch. 4) we still share rudimentary emotional affects with a wide selection of other mammals, perhaps even all of them, supports this explanation.

9.3.7 Emotivism implies incorrect reasoning

Previously (Ch. 6), I noted the cleft between internalists and externalists about motivation and their differences concerning practical reasoning. This difference in perspective also shows up in relation to emotivism. Broome (2013: 276) argues that we “may say that reasoning is correct exactly when it looks correct,” and that we therefore should take propositional statements at face value. Emotivism (expressivism) fails to do this, he argues, insofar as they express “flavorings” or “plans” beyond the formal propositional content of the statements that are made. To my mind, this criticism is valid as far as it goes.

Emotivism can produce two good replies that are replies rather than count-arguments. The *first* reply is that sentences serve important *purposes other than* stating propositional content that is to be taken at face value. Indeed, I think there are many social contexts, including decision-contexts, in which it would be generally regarded as somewhat strange or misplaced to take normative statements at face value. This sense of strangeness can in some cases be intensified by careful interrogation into the supposed meaning of normative statements. Normative language, moreover, is often used in a loose rather than strict manner, and in such cases, a loose interpretation may be the correct response.

The *second* reply is that normative arguments are distinctly *different types of arguments* than those that can be presented in terms of propositions and truth conditions. Blackburn (1993) shows this in the theory he calls “quasi-realism.” The formulations of normative statements contain normative linguistic markers that allow for reasoning from normative premises (combined with ordinary propositions) to arrive at normative conclusions. Descriptive or propositional statements alone cannot do this, as Hume showed, because it would make invalid inferences from “is” to “ought.” Quasi-realism shows how normative statements resemble propositional statements but differs from them in kind. Hence, expressivism is supported as an explanation of normative language use.

9.3.8 Emotivism as denying the expression of beliefs

Among those attacks on emotivism that are centered on the linguistic function of moral terms, many operate with a dichotomy between desires and beliefs. Their claim is that emotivism mistakenly insists that moral terms only express desires and never beliefs. MacIntyre (1981: 11), for example, defines emotivism as “the doctrine that all evaluative judgments and more specifically all moral judgments are *nothing but* expressions of preference, expressions of attitude or feeling, insofar as they are moral or evaluative in character.” Similarly, claims about emotivism can be found in Jackson and Pettit (1998) and Brink (1997). These accounts portray emotivism as *denying* that moral terms can express or represent beliefs.

It should be noted that these characterizations represent extreme forms of emotivism, which *none* of the major contemporary emotivist philosophers (e.g., Simon Blackburn, Allan Gibbard, Michael Ridge) hold. Each of these proponents of emotivism is concerned with the non-cognitive element moral terms positively *do* express, rather than the denial that they also in some cases can express as elements of belief. In advocating his “ecumenical expressivism,” Ridge (2006) is very explicit and vocal on this particular point. According to Ridge, this stylized dichotomy between moral terms either expressing desires or expressing beliefs is simply a “false dichotomy” (Ridge, 2006: 303–305).

By presenting emotivism as making absurd or implausible claims, which are *compatible with* emotivism but to which emotivism is *not committed*, critics such as those mentioned above make emotivism an easy target for themselves. In the defense of MacIntyre (1981), it should be said that he makes references mainly to the works of Charles Stevenson, but as indicated earlier in this chapter, this literature in many respects represents both an off-center and outdated version of emotivism.

How does my version of emotivism stand up to the criticism? In the chapters leading up to this one, I have argued for a position that allows for a variety of moral linguistic practices so that expressions of beliefs and desires are both undoubtedly made. The type of attack presented above therefore also fails in regard to my version. I am in line with *ecumenical expressivism*, which “allows that moral utterances express both beliefs and desires but denies that a moral utterance is a guaranteed to be true just in case the belief(s) it expresses is (are) true” (Ridge, 2006: 307–308).

In my view, expression of belief in moral statements will be *accompanied* by expressions of affect. When moral terms are used in order to express beliefs about ought, the objects that these beliefs are directed at happen to be factious. While proper beliefs are supposed to track truth but in these cases almost certainly track falsity, they are making the error identified by error theorists (e.g., Mackie, 1977). The fact that beliefs of this kind are accompanied by affective attitudes does not need to enter the conscious awareness of the person holding the beliefs, and so it is not strange if it is sometimes overlooked. Some people may systematically overlook the presence of the affective content that goes into their moral statements.

9.3.9 The egoism fallacy

The charge of egoism is sometimes made against emotivism. In this section I point out that this line of criticism corresponds to a fallacy, specifically a fallacy of inappropriately conflating “subjective” with “selfish.” Hence, I refer to this charge as *the egoism fallacy*. The charge of egoism is formulated in different ways and directed at different conceptual targets in the theoretical structure of emotivism. My aim is to deal with the most serious charges and place some of the less serious charges to the side. The most serious charge of egoism seems to be the one that is aimed at *meta-ethical relativism* as a central pillar of emotivism. Hence, most of the discussion will revolve around this particular type of objection to emotivism.

It should be noted that the egoism fallacy potentially includes *several* mistakes, which can be combined and co-committed in different ways. What causes the inappropriate conflation of “subjective” and “selfish” may be a number of different things. The principle type of error seems to involve reasoning and can therefore be characterized in terms of *fallacious reasoning*. Other causes of walking into the egoism fallacy appears to have to do with a lack of *conceptual clarity*, with *faulty premises*, or with a mistaken or significantly incomplete or lack of understanding of the human being as a biological organism.

In order to better understand the various charges against emotivism imputing egoism, *conceptual disambiguation* is necessary. The first thing to take note of is that, with a few rare exceptions (e.g., Locke, 2006; Rand, 1970), critics and holders of opposing views maintain that egoism is something “bad,” “improper,” “unwarranted,” or “undesirable” (Hinman, 2012: 112; Kohlberg, 1984: 5; MacIntyre, 1981). In contrast with the overwhelming majority of thought on the subject, Rand (1970) glorifies egoism as an ethical ideal. I shall assume that the majority of critics are correct in their assessment of egoism as a poor starting point for normative ethical thought and theory. This means that the criticism has a *valid case* as far as its central premise for *initiating* criticism is concerned. So, while I do *not disagree* with the thought that various versions of egoism suffer from problems of moral justification, I do *disagree* that this affects emotivism or that emotivism implies egoism.

In order to expose the errors of the criticism, I present a conceptual analysis of egoism and distinguish the usage of different terms that relate to it. First, however, it should be noted that the theoretical levels of ethical analysis (i.e., descriptive, normative, meta-ethical) are important because they provide different meanings to what an accusation of egoism implies. Implicitly or explicitly different criticisms file their complaints under some interpretation of emotivism; some of these interpretations are correct, while others are not.

A charge of *descriptive egoism* is an empirical matter concerning whether emotivism *describes* human beings in general or decision-makers specifically, correctly or incorrectly. Since emotivism is *not* a descriptive theory, this type of criticism misses its target. Whereas emotivism aims to *explain* ethical activity and psychology in terms of their emotional origins, it is not committed to any particular view that specifies the subjective normative content in the mind of a certain human being at any given time. It says that whatever this content is, it is captured in terms of a specific person's set of values, and that these values reflect emotions. However, it does not describe what these values are, such as whether they are more egoistic and self-serving or more altruistic and socially attuned. Emotivism equally explains the motivation of more egoistic individuals and more altruistic individuals, regardless of their specific evaluative inclinations. This does not mean, however, that emotivism denies that there might be answers to such descriptive questions; this is just not the subject matter of emotivism as a meta-ethical theory.

Independently of emotivism, I can add that descriptive egoism as a general theory is most probably incorrect (Batson, 2011; 1990) because, as previously noted (Ch. 4 and Ch. 5), evidence suggests human beings are in fact subcortically hardwired to be social and relational (Panksepp, 1998: Ch. 13) to the point of blurring the distinction between harm to oneself and harm to one's friends (Beckes et al., 2013). While the *theoretical status* of emotivism is *unaffected*, it can be *furnished* by richer and more informative descriptive theories that provide details about what general categories normative content can be filed into. Regarding the relation to descriptive theory, one more concern has to be pointed out to ward off potential misunderstanding. As an *explanatory theory*, emotivism is *not* independent of the empirical descriptive facts on which the explanation *itself* rests.

The charge of *meta-ethical egoism* is more direct and therefore more interesting. It can be suggested that emotivism, due to its footing in meta-ethical relativism, can either imply or

be equivalent to egoism at the meta-ethical level. Specifically, it can be reasoned against as follows: (1) emotivism relativizes ethics (morality) to the desire-based interests of persons, and (2) the desire-based interests of persons are by definition “self-interests,” and (3) “self-interest” can only express egoist motivation, (4) *therefore*, emotivism entails meta-ethical egoism. In other words, if morality is internally based, it is, despite any appearances to the contrary, doomed to reflect a fundamental form of egoism.

While I think this piece of inferences is *valid*, I believe it reaches a conclusion that is *false*. The reason that the conclusion is false is that one of the premises is false. The *second* premise is auxiliary; it does not add any content to the reasoning, but it is also not completely innocent. For the sake of taking this criticism in its strongest version, I shall grant that we can make use of the equivocation between the interests *of* a person and discuss her “self-interest,” assuming a liberal use of this latter term. But if we grant this interpretation, under which the *second* premise becomes acceptable, I argue that it now becomes impossible to also accept the *third* premise. The problem is that the *wide* and liberal notion of “self-interest” just granted in order to save the second premise discords with the *narrow* notion of “selfishness” that egoism invokes.

In relation to egoism, the terms “selfish” and “self-interested” can usefully be distinguished. In everyday linguistic practice, these terms are used in a variety of overlapping ways. One potential problem with the casual usage of these terms is that the overlap becomes so substantial that it is hard to differentiate the terms. I will propose a stricter usage that may be helpful in systematic deliberation and formal discussion regarding normative questions, especially when it comes to the issues surrounding “egoism.” Differentiating the two terms is important for discerning the difference between meta-ethical relativism and egoism. An analogous way to do this is to distinguish between “narrow self-interest” and “broad self-

interest,” which is the strategy used by Goldman (2009: 238); I shall therefore think of “selfishness” as corresponding to “narrow self-interest.”

In my view, selfishness can be straightforwardly seen as the idea that one is never self-denying and where the only concern is maximizing benefits and minimizing harms to oneself. The term “selfish” is clearly and unambiguously related to egoism under this description. In descriptive ethics, a “selfish” person is a person who is egoistic. In normative ethics, “selfishness” designates a “virtue” that is particular to *normative egoism*, as most famously defended by Rand (1970), but is otherwise taken to designate a “vice,” a “problem,” or a “departure” from ethics in the vast majority of normative theories (e.g., Aristotelian virtue ethics, classical utilitarianism, Kantian deontology).

A liberal interpretation of “self-interest” covers all the interests of a given person: all the interests of the “self,” as it were. Clearly, the notion of “self-interest” then has to include *far more* than mere selfish interests. One may be interested in the welfare of one’s mother. One may be interested in the happiness of one’s spouse. One may want to satisfy the needs of one’s clients simply because one wants them to be happy. One may take an interest in the plight of poor homeless children. One may desire to restore the Gorongosa wildlife habitat in Mozambique. One may be interested in the continuation of life and livelihood on the planet a hundred years from now, or in the distant future. One may even, as routinely reported in the media, be interested in sacrificing one’s own life for some particular cause.

The end-station here is *alter*, not *ego*, although it is ego taking the interest in alter on their behalf. In short, “self-interest” may incorporate any kind of altruistic interests. The broad sense of “self-interest” leaves the agent free to take an interest in whatever she likes (Goldman, 2009), whether this is self-oriented or other-oriented, and whether it is on behalf of oneself or someone else. These are *interests of the self* and interests of a person that this

person defends, protects, and promotes based on her *personal values*, which form the basic normative content in decision-making. Nothing, in principle, suggests that what these values reflect and what this person cares about is herself only.

At this point it can be claimed—the descriptive charge again—that as *a matter of fact*, people care only about themselves, fundamentally speaking. Dawkins (1976), for example, argued that we *must* be selfish creatures because, as he claimed, our *genes* are selfish. Again, this would not affect emotivism as an *explanation* of human morality. However, there seems to be ample empirical evidence (e.g., Batson, 2011; Decity and Chaminade, 2003; Sober and Wilson, 1998; Batson, 1990) to suggest, quite independently of emotivism, that descriptive egoism is false.

From a biological and scientific angle, the mistake of underappreciating the social attuned nature of the subcortical machinery of our brains is sometimes made (see Ch. 4). Due to their central role, affective emotions emanate *from* the brain of a particular person, which does not imply that this stream of affect only, or even predominately, caters to desires, needs, and wants that *serve* that particular person. In evolutionary terms, the survival of the species does not reduce neatly to the survival of the individual because the survival of the family, the group, and even of strangers must also be considered as part of the equation.

There are, moreover, abundant examples of individuals selflessly sacrificing their well-being, safety, and even their very own lives for others, sometimes even for the sake of mere belief: for example, a belief in absolute moral truth or in the particular will of some deity. Undeniably, there are also ample examples of *genuine caring* for others, as is almost invariably visible in parenthood. MacLean (1990: 562) argues “a parental concern for the young generalizes to other members of the species.” Thus it is not surprising to see that genuine caring is also found in relation to complete strangers: for example, helping people

who have gotten their car stuck or picking up the lost wallet of complete strangers and assuming the cost involved in returning it.

These selfless acts can in many cases make us feel good about ourselves, but it seems implausible to thereby call them “selfish” or “egoistic” because that would conflate not only the term “selfish” with “self-interested” but also with “altruistic.” If we do this, we debar ourselves from being able to talk about altruism as a special kind of act, and egoism therefore becomes tautologically true, which would not be helpful for sorting out practical moral issues.

9.3.10 Other arguments against emotivism

It is sometimes claimed that the *moral words* that are frequently used do not convey emotion at all. For example, a term such as “good” is sometimes taken to be neutral as far as emotions go. This criticism relates only to the traditional linguistic domain of emotivism, but it makes up an important part of what the emotivist position is supposed to be able to defend.

As a response, the first thing that needs to be pointed out is that the core of the claim is an empirical question. Neuroscientists could use various imaging techniques to potentially determine the truth or falsity of the claim by observing subjects expressing moral phrases and words. Another possibility is to quarry the subjects’ subjective experiences, but such quarries are easily malleable and not particularly reliable (LeDoux, 2002: 202). It should also be pointed out that experiences themselves are one thing, and our awareness of them is another. Subjects may well be experiencing emotions affectively *without being aware* of these experiences cognitively.

Beyond the relevance of nitty-gritty science and the difficulties of determining the involvement of emotions, it can be independently doubted whether the claim can be true. If one asks about the valence of a term such as “good,” the answer is likely to be that it is *positive* and not negative. This response indicates what sort of emotions might be involved in

the basic usages of this particular term. It does not preclude that a different spin can be put on it, for example, in phrases like the emotionally ambiguous “evil is my good.” However, there are clear indications that moral terms have references to our affective emotional experiences and apprehensions.

Finally, any use of words will involve cognitive processes, but if affective neuroscientists, notably including Mark Solms and Jaak Panksepp (see discussion in previous chapters, especially Ch. 3 and Ch. 4), are correct, the cognitive processes and cognitive consciousness *depend on* deeper affective emotional processes in a nested hierarchy of consciousness in which affective consciousness is primary (Solms and Panksepp, 2012). Emotional input from affective levels of consciousness is a requisite for higher levels of consciousness. In this perspective, cognition is embedded in affect, and it becomes empirically meaningless to talk of “pure cognition” and “purely cognitive states of mind” (Gallese, 2013; Panksepp, 2012; Solms and Panksepp, 2012).

Another quite frequent type of counterargument is that *moral principles* represent morality or ethics that are completely decoupled from emotions. A wide number of normative ethical theories have invested in this position by proposing principles that rely on rationality, reasoning, intuition, introspection, or some combination of these. Some of these positions have already been discussed (e.g., Broome, 2013), so here I will simply elaborate on some general points about the status of moral principles vis-à-vis emotions.

First of all, it can be doubted whether moral principles are necessary for ethical decision-makers. There are many reasons to think that decision-makers can do well without them. Particularists about ethics and normativity articulate this type of view. Dancy (2004), who modernized the early particularist position of Ross (1930), argues that ethics can work completely without reliance on moral principles and that the effects of moral principles in

most cases will be to obscure ethics. Dancy (2004) presents a particularist theory where direct “intuition” of the specific moral context is relied on in order to make moral decisions. However, there appear to be no obstacles for developing an emotivist version of particularism, in which emotions substitute for “intuitions” as a means of making sense of and making decisions.

Despite the prospects of particularism, even *at* the level of endorsing principles, the *endorsement* itself is a mental psychological process that is explained by emotivism. As we have seen in the discussion above (and in earlier chapters), emotivism offers a theoretically plausible and empirically adequate account of the general picture of what is going on as a person approves of or endorses something, including when one endorses something as normatively authoritative. We have seen that there is a more detailed picture, too; the one presented by affective neuroscience.

If moral principles are to pose a threat to the emotivist position, it must therefore be shown that normative moral principles first come to *exist* and subsequently come to *motivate* agents and decision-makers psychologically. But as we have seen in previous chapters (especially Ch. 8), the existence of objective moral values—and we can, temporarily suspending particularism, grant for the sake of argument that these values can be captured in terms of principles—is metaphysically speculative and not epistemologically compelling (Goldman, 2009). We have also seen that even if objective moral values or principles existed, it would be difficult to find a compelling story regarding first how we came to know about them and subsequently how they would be necessarily motivating (Goldman, 2009; Mackie, 1977). This task would be all the more difficult if this motivating story also had to ignore emotions and the affective constitution of human nature.

9.4 Neuroemotivism

9.4.1 *The need for a revision of emotivism*

Over its history, the philosophical tradition of emotivism has sustained a considerable amount of criticism from a variety of philosophical angles. However, as my discussion shows, its central pillars still appear to be standing. Moreover, the three pillars presented in the beginning of the chapter raise critical questions of their own and are directed at rivaling positions. Although the central pillars of the theory remain intact, emotivism has evolved considerably over the past century and the beginning of this century. With the benefit of hindsight, early classical statements (e.g., Ayer, 1936; Stevenson, 1944; Urmson, 1968) of emotivism stand out as incomplete and not fully satisfactory.

One of the most important stages in the evolution of emotivism is the change from a narrow preoccupation with surface-level linguistics to an emphasis on the *moral psychology* that underpins moral practices, including linguistic practices. This shift in orientation is to some extent brought out in the theoretical developments of modern-day expressivist theory, notably Blackburn (1998) and Gibbard (1990). However, this *psychological turn*, it seems to me, has not yet fully matured. A mature psychological turn would involve a full integration of emotivism with the empirical basis found in *affect theory* in social psychology, represented by researchers such as Izard (2013; 2009; 1991), Ekman (2003; 1973), and Griffith (1997). Chapter 2 shows how close the lines of arguments presented by emotivists and defenders of affect theory are. However, the review chapter also reveals that this specific theoretical integration still has a long way to go.

Several philosophically and psychologically important points that are helpful for developing emotivism into a mature and adequate theory follow from the psychological turn. For example, recognizing the epistemic inaccessibility of the cognitive unconscious and the

fact that the affective conscious experience often escapes explicit awareness (Ch. 5), meta-ethical theory needs to shift its focus from only what moral agents overtly *intend* to include what moral agents *express more broadly*. As Panksepp noted (as discussed in Ch. 6), there can be implicit affective *intention-in-action* that precedes an explicit intention for action. Blackburn (1993: 5) draws attention to “mechanisms whereby what starts life as a non-descriptive psychological state ends up expressed, thought about, and considered in propositional form.” I suggest that it is in the interest of emotivists, and indeed anyone who wants to understand normativity, to specify as correctly as possible what these mechanisms are as well as the entire causal chain of mental events that leads up to moral expression and behavior.

To this end, there is potentially a *second evolutionary stage* for emotivism that has become possible over recent decades, namely what can be called the “neuroscientific turn.” In psychology, affect theory is already well under way making this turn (e.g., Lerner et al., 2015; Izard, 2013; 2009; 1991), and it seems incumbent on emotivism to do so as well. Affective neuroscience, as shown in detail in the previous chapters, contributes by detailing the affective processes and emotional systems in the human brain. This would present emotivism, the most sophisticated scientific understanding of human emotions available, as a descriptive and causal theory that has bearing on the explanations it offers. Moreover, affective neuroscience, as represented by scientists such as Panksepp (1998), appears to strongly corroborate the three pillars of emotivism described above. I propose to develop a version of emotivism that harnesses this potential, which I choose to call *neuroemotivism*.

One fear that emotivists may have about the project of neuroemotivism is that it leads to full material reductivism and ultimately a version of moral realism. But as I have argued in Chapter 3, drawing on the thoughts of Mark Solms and others, such doubts would be unwarranted because in philosophy of mind, specifically concerning the theory about the

relation between the subjective mind and objective brain, affective neuroscience turns out to be compatible with *dual-aspect monism*. Dual-aspect monism resists material reductivism, and therefore avoids placing emotivism in the territory of moral realism. As previously argued (Ch. 3), psychology cannot simply be reduced to biology because none of the theoretical concepts in biology are able to capture the internal aspect of the phenomena of subjective experience.

The facts about human nature are essential in order to spell out a theory of emotivism (Gibbard, 1990). When seen this way, emotivism implies a fundamental *naturalism* but not regarding conjectured “moral facts” or “normative properties,” but about *human nature*. This naturalism regarding human nature both grounds emotivism and determines the forms it can take as a theory of normativity. The compatibility of meta-ethical emotivism and a naturalistic worldview, in the sense just presented, is also noted by other contemporary emotivists (e.g., Ridge, 2006: 302). I propose, then, that meta-ethics needs to be based on a scientific understanding of human nature and that affect theory (social psychology), affective neuroscience (neurobiology), and dual-aspect monism (philosophy of science) make substantial contributions to such an understanding.

Hence, I view it as paramount that emotivism is continuously updated in accordance with relevant scientific evidence. The affective revolution that takes place in neuroscience and psychology is of special relevance, including its increasingly more sophisticated understanding of human emotional systems. While this is unlikely to change the basic ideas and structure of emotivism, it could affect the scope of emotivism as well as many of the specific claims it can make. The further emotivism can move in the direction of neuroemotivism, the better scientifically informed it will be and the more helpful it will be as an explanation of moral talk, thought, and action.

9.4.2 Neuroemotivism outlined

Traditional accounts of emotivism tend to look only at the part of human psychology that is within conscious awareness, and they are predominately concerned with *linguistics* and *logic* in relation to the special function and meaning of moral statements, focusing on terms such as “moral judgments,” “evaluations,” “meaning,” and “intentions.” As I have pointed out in this chapter, a focus on *conscious expressions* of sentiments or emotions only covers part of the complete explanatory story of human morality. Specifically, traditional emotivists tend to leave the *foundational part* of the story absent.

What people explicitly *think* they mean, or thought they meant, by their normative statements is one thing. What they truly mean or meant and why they actually made the statements they did is another matter. It is often unclear what people in fact mean by their expressions, even though it remains true that these in part do reflect their emotions. Perhaps sometimes they do not even mean anything at all. Neuroemotivism proposes that more attention needs to be paid to mental processes that do not enter conscious awareness, what social psychologists (e.g., Bargh and Morsella, 2008) and neuroscientists (e.g., Berlin, 2011) sometimes refer to as “the cognitive unconscious”.

As many psychologists accept, it is not reliable to ask people what they meant after the fact, as this is apt to give rise to *post hoc* rationalizations and “just-so stories” (e.g., Haidt, 2001; Weick, 1995). Similarly, as psychotherapists know well (e.g., Solms and Turnbull, 2002), people are good at fabricating (confabulating) suitable reasons for their statements once asked, but these cannot be taken as trustworthy accounts of actual meaning conveyed in these statements. Consequently, emotivism and others interested in moral psychology would do well to develop broader explanations for moral terms to show what causes them, what they express, and what their functions are; the same could be done for explanations of motivated moral action.

The ambition of emotivism is to provide an explanation of moral thought, talk, and action. This explanation should cover the whole normative process, not only its surface-level manifestations (e.g., moral objections) and subsequent inter-subjective social constructions (e.g., moral agreements). In particular, *foundational normativity* associated with sub-cortical neurochemical processes in the affective brain also needs to be taken into account. *Foundational normativity* underpins all ensuing moral activity, including outwards socially responsive and interactive activities such as proposing, defending, endorsing, accepting, creating, and co-constructing morality in inter-subjective social space.

Over the course of the previous chapters, I have argued that *normativity* is best seen in terms of *emotional affect*, the “normativity-as-affect hypothesis”. This view is consistent with the view “that the experience of moral emotions is constitutive of the exercises of practical reason” (Bagnoli, 2011: 62) as long as we recognize that moral experience is not always, and perhaps typically not, something the experiencing agent is fully aware of. Normativity can, to varying degrees, be implicit. Affect, in this view, represents a naturalistic explanation of normativity. The ultimate biological sources of normativity appear to be *affective consciousness* (Ch. 5.2.3). Primary emotions, as identified and specified by Panksepp (1998), accounts for the valence of normativity as well as its distinctive “coloration”. For these reasons emotions feature centrally in the neuroemotivist explanation of morality. This view, moreover, places *emotional feelings* at the center of morality.

Neuroemotivism seeks to detail this neurobiological explanation of normativity. Central to this explanation are the ideas that the neurochemical and neuroanatomical characteristics of emotional systems explain which emotional feelings we have and that these emotional feelings correspond to normative feelings of “ought” and “ought not,” “good” and “bad,” and “right” and “wrong” in a moral context. The emotional “flavor” as well as the “valence” of normative feelings are explainable in terms of biological emotions. Thus, moral

emotions that include both cognitive and affective elements, such as “guilt,” “pride,” and “shame,” depend on primary affect from subcortical brain areas for their distinct feel and pleasantness or discomfort. Cognitive object representations, meanwhile, account for the way these moral emotions attach us to the external social world.

Neuroemotivism respects the fact that an explanation of human normative manifestations will have to relate to what happens, not only in the cortical areas of the brain where emotions sometimes become declarative or explicit but also in the deeper emotional structures in the sub-cortical and limbic areas of the brain. Normativity, I argue, does not start out as explicit or declarative elements of explicit thought but instead as *implicit normativity* at a *pre-cognitive stage* as feelings that are only *affectively* conscious. This is what the evidence provided by *affective neuroscience* suggests. Neuroemotivism will be an empirically adequate scientific position as long as my identification of normativity is affective and holds, and I have pointed out a number of empirically oriented theoretical reasons for accepting this affective interpretation of normativity.

Neuroemotivism is *layered* according to the anatomy and functional stratification of the human brain in the manner explained previously (Ch. 4). While many of the classical and contemporary versions of expressivism are only concerned with the higher cortical functions of the brain, such as those involved in the formulation and expression of explicit moral judgments, neuroemotivism also looks at what happens sub-cortically, accounting for the entire process by which affects are transmitted and ultimately expressed.

The idea of emotivism or expressivism as layered is not new. For example, Blackburn (1998: 8–14) demonstrates the usefulness of thinking of emotivism in terms of layers in his idea of “emotional ascent.” However, the process of relating emotivism to specific brain layers and electrochemical processes in the brain, as I suggest, appears to represent a new way

of conceptualizing emotivism. What I have presented here is merely an outline of what neuroemotivism would look like and what it would say. The conceptual details will have to be worked out.

9.5 On moral justification

9.5.1 Moral justification in practice

Moral justification of decisions is important in practical ethics, especially the practice of giving reasons and arguments for one's decisions and actions (and sometimes against their alternatives). Consider for example the role of moral justification in an organizational change process: During organization change—which is typically psychologically painful for the parties affected—timely and well-adjusted justifications can help to promote “a number of positive outcomes, including perceptions of procedural fairness, organizational citizenship behaviours, preservation of organizational commitment, trust in leadership” (Lines et al., 2011: 168), as well as positive attitudes and behavior towards change itself (Lines, 2004). Moral justification can be expected to have positive social effects, such as the reduction of normative tension between affected parties to the change process within the organization, but only in relation to external stakeholder groups outside the organization and to society at large.

Reasons and arguments that people give when they are being honest and sincere reflect their *moral judgments*. Since moral justification uses reasons and arguments as its means, moral justification depends on moral judgments. Moral judgments tend to be *controversial*, however, because they are inherently *subjective*. Another matter, which often complicates practical ethics, is that the people who give reasons and arguments often have vested interests in the outcomes of the relevant decisions, how the decisions are previewed, their reputation as professionals, or their own self-image. This involvement often gives agents and decision-makers *further* reasons, which can bias the reasons and arguments offered for the decision at hand. Often incentives are structured so that these further reasons provide reasons for *not* necessarily being honest and sincere, which are also reasons for stakeholders involved in the

decision to expect the reasons and arguments that are given to not necessarily be fully honest and sincere, allowing some reason for *distrust*.

Moral justification in complex social settings is typically fraught with difficulties and room for misunderstanding. I will place these further complicated matters aside for the moment and instead return to the difficulties concerning *subjectivity*. The different parties to a moral decision may of course be in fully inter-subjective agreement concerning a decision and the reasons and arguments for it, including the underpinning moral judgment. However, we may suspect these to be rare occasions.

Even when people come to agree upon the relevant *facts* that have bearing on a moral judgment, there may be significant evaluative disagreements reflecting subjective differences in taste. As emotivism suggests, there will often be a residual practical disagreement that cannot be resolved by mere argument: an irreducible element of *de gustibus non est disputandum*. And even in cases where there *is* an agreement regarding what to try to achieve (i.e., agreement about values) and what to aim for (i.e., agreement about goals), there can be further disagreement concerning *how* best to meet the chosen goals, the *prioritization* of different sub-goals, as well as *who* has *what* responsibility for taking action towards arriving at the end states that the goals represent. Hence, almost invariably there will be some room for disagreement concerning moral decisions. Practical approaches to moral justification, therefore, are probably well advised not to assume that fully acceptable moral justifications are available.

One of the specific characteristics that is particularly apt to make moral judgments controversial and provocative is when they involve some claim to “moral objectivity” or “moral truth,” or have the appearance of involving such a claim. I have shown that the basis for such claims is shaky and their meaning is unclear (Ch. 8). Moreover, I have argued,

alongside Blackburn (2010) and other internalists (e.g., Goldman, 2009; Joyce, 2009) that motivation and value are not only speculative but in fact also *unnecessary* to postulate such grand realist claims. Practical decision-makers and professionals need internal, emotionally based confidence in *what is good* rather than obscure beliefs about what moral truth might be.

Emotivist meta-ethics suggest that it is possible to put talk of moral truth and objectivity in quotation marks; it is not at the heart of the matter of morality. At the same time, emotivism also explains realist beliefs in objective moral truth; it shows that the (putatively erroneous) beliefs in moral truth and moral objectivity separate from the internal, emotionally based motivation for the actions and decisions (putatively wrongly) associated with these claims and beliefs by the agent. By bringing in the perspective of neuroscience, moreover, it can be explained why these motivations sometimes escape the awareness of the person holding them.

9.5.2 Justification points in two directions

Justification can, in a certain sense, be said to point in two directions. On the one hand, there are the prospects of justification towards others in our external social environment (and, for religious people, perhaps towards one or several deities). On the other hand, there is inward justification towards oneself.

The latter direction of justification is often forgotten. However, evidence from experimental psychology suggests that it is psychologically important; *self-image* is of central psychological importance for individuals acting as moral agents (Bandura et al., 1996). According to Bandura et al. (1996: 365), “People do not ordinarily engage in reprehensible conduct until they have justified to themselves the rightness of their actions.” Thus, much of what people choose to do and to refrain from doing appears to be guided by an implicit (sometimes explicit) aim of building and retaining a positive self-image: a positive picture of

oneself. In other words, the findings suggest that agents in general or most agents appear to be describable in terms of some sort of implicit *virtue ethics*.

While the costs of lacking *external* justification (reasons and arguments) for a given moral decision or actions typically *are* apparent, there are often additional costs related to the lack of *internal* justification (Miller, 2013: 256), and these internal costs may *often not* be apparent. These costs are often not apparent socially to the external milieu, but they may not even be apparent to the agent herself, as in some cases, the shortfall of justification represents *implicit negative normativity*.

In general, it appears to be psychologically costly for individuals to trespass against moral standards they identify strongly and affectively with. This picture of the moral agent, insofar as it is correct, corroborates the position that it is fallacious to characterize the general moral agent as ruthlessly selfish. Individuals are inclined to respect many internalized social norms, even when they know their behavior is not being monitored. This behavior partly involves respect for others, but it also partly involves a measure of self-respect.

9.5.3 The authority of desires

It has often been a default assumption in moral philosophy that desires need some sort of normative stamp of approval in order to be normatively “valid” or “justified” (e.g., a claim recently set forth by Parfit, 2011). As was shown in a previous discussion (Ch. 7, Ch. 8, and in this chapter), however, other would-be sources of moral justification are either unsound (e.g., meta-ethical cognitivism), in need of validation (e.g., Moore’s intuitionism), or themselves in need of moral justification, as in the case of normative theories. For example, it was shown (Ch. 8) that candidate’s justifying notions, such as “objective values” and the “objective moral point of view” are difficult to establish. Meanwhile, *desires* based on

affective emotions represent a potential alternative source of moral justification (Goldman, 2009; Frankfurt, 1988).

Humans desire all sorts of things: money and wealth, power, fame, food, fancy cellular phones, expensive French wines, friendship, love and affection, sex, a glass of water, a good night's sleep, etc. It is widely agreed upon that such desires are not particularly attractive as a basis for moral justification. Yet some philosophers (e.g., Goldman, 2009), including most emotivists (e.g., Blackburn, 2010; 1998), argue that moral justification has to *relate* to our subjective desires somehow; they cannot simply be abandoned as the Stoics and Kant would have us do. If desires are to provide moral justification, however, they need to be *qualified*. What we would need, it seems, is *qualified desires*. This topic raised questions about what the criteria by which desires are to be qualified should be and could be. In this section, I explore this issue.

Frankfurt (1988) has plausibly suggested that we should first look at how a given person's desires are coupled. Some desires and concerns seem more central than others (Goldman, 2009), while some desires are merely instruments for meeting other desires whose satisfaction is more dearly sought after. As mentioned previously (Ch. 5), Frankfurt proposes that we have *desires about desires*, which he appropriately refers to as “second-order desires.” This idea is not entirely new with Harry Frankfurt; it is also found in the early writings of Bertrand Russell (e.g., Russell, 1897/1983: 100–104), which have been usefully developed by Lewis (1989: 113–137). More recently, this type of perspective is elaborated upon and defended by Goldman (2009).

“Human beings are not alone in having desires and motives, and in making choices” (Frankfurt, 1988: 11), but it may be unique to humans that they are able to form *second-order* desires; “Besides wanting and choosing and being moved *to do* this or that, men may also

want to have (or not to have) certain desires and motives” (Frankfurt, 1988: 12). This tendency shows some of the complicities in human motivation and hints at how different emotional systems can be in conflict with one another. Because our affective lives are complex and involve multiple processes, modeling desires on a single plane may be insufficient in order to obtain a reasonable overview. By introducing two levels, however, we can conceptualize how different affects interact, regulate, compete, and merge with one another. The normative authority of particular desires and affects seems to depend on their relative position in the affective web that constitutes “self.” As both Damasio (2010) and Panksepp (2005) propose, “self” is not one single phenomenon; it is a multifaceted, dynamic, affectively based chain of phenomena that is hierarchically ordered in layers with corresponding levels of consciousness.

This explanation suggests that normative authority is divided among several desires and affective feelings that are concurrently active. In this picture, some affective forces will feature more centrally than others in relation to the various levels of “self.” It cannot be taken for granted that the higher level affects (i.e., second-order affects) are always in control, but that may often be what is required in order to achieve what we call morality.

9.5.4 Justification light

Here I summarize the above discussion under a position I will call “justification light.” The essence of this position is that, while “full” and “objective” moral justification is indefensible because it lacks a credible basis – an external source of normative authority – practical agents and decision-makers can, as explained in terms of emotivism, deploy a *subjective* justification. This subjective justification is expressive of the agent’s emotional affects as it comes out of the explicit part of their moral reasoning processes. This sort of justification

expresses a “psychological ought”, as opposed to the conceptualization of an “ought” that is independent of the particularities human psychology.

As classically argued by Hume, a valid normative conclusion cannot be drawn from descriptive premises alone. The psychological “ought” arises from the characteristics of our affective animal natures combined with our cognitive abilities to structure our emotional affective upshots according to information about the external world, and according to logical structures that ensures consistency and ends-means coherence (much of which the human brain is capable of doing without us knowing, Berlin, 2011). This sense of “ought” is therefore based on what is (the descriptive), although it equips us with normative experiences of “what ought to be,” and what is “good” and “bad” (etc.).

In moral life, we face the guidance provided by our own emotional nature and the indirect guidance provided by others, as well as the social reciprocal understanding that has been generated on this social interactive basis. Because Hume’s syllogistic logic is correct about the invalidity of inferring normative conclusions from purely descriptive premises, and since the premises we have to work with are all descriptive, a *fully justified objective morality is out of reach*. Moreover, it is out of reach on a permanent basis.

In its place, however, we have a human morality that is underpinned by our psychological and biological premises and supported by the social structures we (and others before us) have created and continue to defend, modify, and develop through our social interaction. This process incidentally gives us what we need in terms of moral justification: a subjective sense of moral justification that I have termed *justification light*. Many philosophers have voiced dismay with this type of conclusion. For example, Bagnoli (2011: 64) notes: “When normative authority is reduced to motivational power, it looks arbitrary and spurious”.

However, it is also possible to argue that the result is not so gloomy (e.g., Blackburn, 2010; Williams, 1985: 93), because it turns out that even without the grandeur of objective justification we still have sufficient resources to continue our practices of justification. We can justify our decisions morally towards ourselves and others, with reference to our own emotional natures and our subjective concerns, feelings, and desires. These practices of justification are *normatively meaningful*, even though they are much less ambitious ones than some have hoped and wished for.

It needs to be stressed that all moral justification is prone to being distorted by ex-post rationalizations (Fine, 2008; Weick, 1995). As noted earlier, when we think about ourselves as object, we remove ourselves from ourselves as experiencing subjects; the cost of cognitive self-representation is inhibition and suppression of the affective layers of self (Solms and Panksepp, 2012). Post-hoc rationalizations may easily fail to pick out our genuine motivations, and it seems impossible to certify that we have identified the true motivations insofar as our motivations to a considerable extent are inaccessible at the cognitive-conscious level. This view suggests that while moral justification serves important social functions, it should be used and interpreted with caution.

9.6 Conclusions

In this chapter I have argued for emotivism in general as well as a more specific, empirically updated version of emotivism that I have labeled *neuroemotivism*. Neuroemotivism draws on neuroscience for its naturalistic underpinnings, notably on evidence from *affective neuroscience*. In the theory of mind, it is supported by *dual-aspect monism*. In psychology, it is particularly corroborated by *affect theory*.

I have argued that none of the major lines of criticism of emotivism is convincing. Many of the charges against emotivism presently and historically tend to be false or

inappropriate. One component responsible for the objections to emotivism is that the theory itself has not always been properly stated by its early proponents. I have shown that emotivism should not be conflated with any version of egoism. One of the “costs” of accepting emotivism is that the idea of “moral truth” must be abandoned. However, I have argued that we do not need notions such as “moral truth” in practical ethics, and thinking along such notions can moreover have undesirable practical implications.

Finally, I have argued that we have to abandon the idea of objective moral justification and instead must settle more modestly for *subjective justification* unto others in the inter-subjective social space. Subjective justifications can be related to a theory of desires that identifies secondary desires as a means to justify moral decisions and actions.

Chapter 9 References

- Audi, R. (2004) *The good and the right: A theory of intuition and intrinsic value*. Princeton: Princeton University Press.
- Audi, R. (1996) Intuitionism, pluralism, and the foundations of ethics. In W. Sinnott-Armstrong and M. Timmons (Eds.), *Moral knowledge?* (pp. 101–136). New York: Oxford University Press.
- Ayer, A. J. (1936) *Language, truth, and logic*. New York: Penguin Books.
- Baier, K. (1967) Fact, value, and norm in Stevenson's ethics. *Noûs* **1** (2): 139–160.
- Bagnoli, C. (2011) Emotions and the categorical authority of moral reason. In C. Bagnoli (Ed.) *Morality and the emotions*. Oxford: Oxford University Press.
- Bandura, A., Barbaranelli, C., Caprara G. V., and Pastorelli, C. (1996) Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology* **71** (2): 364–374.
- Batson, C. D. (2011) *Altruism in humans*. New York: Oxford University Press.
- Batson, C. D. (1990) How social an animal? The human capacity of caring. *American Psychologist* **45**: 336–346.
- Bargh, J. and E. Morsella (2008) The unconscious mind. *Perspectives on Psychological Science* **3** (1): 73-79.
- Beckes L., Coan, J. A., and Hasselmo, K. (2013) Familiarity promotes the blurring of self and other in the neural representation of threat. *Social Cognitive & Affective Neuroscience* **8** (6): 670–677.

- Berlin, H. (2011) The neural basis of the dynamic unconscious. *Neuropsychanalysis* **13** (1): 5–31.
- Bergson, H. (1992) *The creative mind*. (M. L. Andison, Trans.). New York: The Citadel Press.
(Original work published in 1946).
- Blackburn, S. (2010) *Practical tortoise raising and other philosophical essays*. New York: Oxford University Press.
- Blackburn, S. (1998) *Ruling passions: A theory of practical reasoning*. Oxford: Oxford University Press.
- Blackburn, S. (1993) *Essays in quasi-realism*. New York: Oxford University Press.
- Blackburn, S. (1984) *Spreading the word*. New York: Oxford University Press.
- Brink, D. (1997) Moral motivation. *Ethics* **108**: 4–32.
- Brink, D. (1989) *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.
- Broome, J. (2013) *Rationality through reasoning*. Oxford: Oxford University Press.
- Bratman, M. E. (2006) *Structures of agency: Essays*. Oxford: Oxford University Press.
- Damasio, A. R. (2010) *Self comes to mind: Constructing the conscious brain*. New York: Random House.
- Dancy, J. (2004) *Ethics without principles*. Oxford: Oxford University Press.
- D'Arms, J. and Jacobson, D. (2000) Sentiment and value. *Ethics* **110**: 722–748.
- D'Arms, J. and Jacobson, D. (1994) Expressivism, morality, and the emotions. *Ethics* **104**: 739–763.

- Dawkins, R. (1976) *The selfish gene*. Oxford: Oxford University Press.
- Decety, J. and Chaminade, T. (2003) Neural correlates of feeling sympathy. *Neuropsychologia* **41**: 127–138.
- Ekman, P. (2003) *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*. New York: Owl Books.
- Ekman, P. (1973) *Darwin and facial expression: A century of research in review*. San Diego: Academic Press.
- Fine, C. (2008) *A mind of its own: How your brain distorts and deceives*. New York: W. W. Norton & Company.
- Fulford, K. W. M. (2011) Neuroscience and values: A case study illustrating developments in policy, training and research in the UK and internationally. *Mens Sana Monographs* **9** (1): 79–92.
- Gallese, V. (2013) Bodily self, affect, consciousness, and the cortex. *Neuropsychoanalysis* **15** (1): 42–45.
- Gibbard, A. (1990) *Wise choices, apt feelings*. Cambridge: Cambridge University Press.
- Goldman, A. I. (2009) *Reasons from within: Desires and values*. Oxford: Oxford University Press.
- Griffiths, P. E. (1997) *What emotions really are: The problem of psychological categories*. Chicago: University of Chicago Press.
- Haidt, J. (2008) From philosophical thinking to psychological empiricism. *Perspectives on Psychological Science* **3** (1): 65–72.

- Haidt, J. (2001) The emotional dog wagging its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* **108** (4): 814–834.
- Hare, R. M. (1952) *The language of morals*. Oxford: Oxford University Press.
- Helm, B. W. (2007) *Emotional reason: Deliberation, motivation, and the nature of value*. Cambridge: Cambridge University Press.
- Hinman, L. (2012) *Ethics: A pluralistic approach to moral theory* (5th ed.). Boston: Cengage Learning.
- Horgan, T. and Timmons, M. (2009) Expressivism and contrary forming negation. *Philosophical Issues* **19**: 92–112.
- Horgan, T. and Timmons, M. (2006) Cognitivist expressivism. In T. Horgan and M. Timmons, *Metaethics after Moore* (pp. 255–298). Oxford: Oxford University Press.
- Izard, C. E. (2013) *Human emotions*. New York: Springer.
- Izard, C. E. (2009) Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology* **60**: 1–25.
- Izard, C. E. (1991) *The psychology of emotions*. New York: Plenum Press.
- Jackson, F. and Pettit, P. (1998) A problem for expressivists. *Analysis* **58**: 239–251.
- Joyce, R. (2009) Patterns of objectification. In R. Joyce and S. Kirchin (Eds.). *A world without values: Essays on John Mackie's moral error theory* (pp. 35–54). New York: Springer.

- Joyce, R. (2008) What neuroscience can (and cannot) contribute to metaethics. In W. Sinnott-Armstrong (Ed.). *Moral psychology volume 3: The neuroscience of morality: Emotion, brain disorders, and development* (pp. 371–394). Boston: The MIT Press.
- Joyce, R. (2006) *The evolution of morality*. Boston: The MIT Press.
- Kalderon, M. (2005) *Moral factionalism*. Oxford: Oxford University Press.
- Kitcher, P. (2011) *The ethical project*. Cambridge: Harvard University Press.
- Kohlberg, L. (1984) *The psychology of moral development: The nature and validity of moral stages*. New York: Harper & Row.
- Kotz, S. J. and S. Paulmann (2011) Emotion, language, and the brain. *Language and Linguistics Compass* **5** (3): 108–125.
- Kramer, M. J. (2009) *Moral realism as a doctrine*. New York: Wiley.
- Lerner, J. S., Li, Y., Valdesolo, P., and Kassam, K. (2015) Emotion and decision making. *Annual Review of Psychology* **66**: 799–823.
- Lewis, D. (1989) Dispositional theories of value. *Proceedings of the Aristotelian Society* **63**: 113–137.
- Lines, R., Sanz, J., and Aramburu, N. (2011) Organizational learning as a by-product of justifications for change. *Journal of Change Management* **11** (2): 163–184.
- Lines, R. (2004) The structure and function of attitudes towards organizational change. *Human Resource Development Review* **4**: 1–24.
- Locke, E. (2006) Business ethics: A way out of the moral morass. *Academy of Management Learning and Education* **5** (3): 324–332.

- MacIntyre, A. (1981) *After virtue: A study in moral theory*. Paris: University of Notre Dame Press.
- Mackie, J. L. (1977) *Ethics: Inventing right and wrong*. London: Penguin Books.
- MacLean, P. D. (1990) *The triune brain in evolution: Role in paleocerebral functions*. New York: Plenum Press.
- Miller, C. B. (2013) *Moral character: An empirical theory*. Oxford: Oxford University Press.
- Millgram, E. (2005) *Ethics done right: Practical reasoning as a foundation for moral theory*. Cambridge: Cambridge University Press.
- Moore, G. E. (1903) *Principia Ethica*. Cambridge: Cambridge University Press.
- Northoff, G. (2011) *Neuropsychanalysis in practice: Brain, self, and objects*. Oxford: Oxford University Press.
- Olson, J. (2010) The freshman objection to expressivism and what to make of it. *Ratio* **23** (1): 87–101.
- Panksepp, J. (2012) What is an emotional feeling? Lessons about affective origins from cross-species neuroscience. *Motivation and Emotion* **36**: 4–15.
- Panksepp, J. and Panksepp, J. (2000) The seven sins of evolutionary psychology. *Evolution and Cognition* **6** (2): 108–131.
- Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.
- Parfit, D. (2011) *On what matters: Volume 1*. Oxford: Oxford University Press.

- Prinz, J. J. (2012) *The conscious brain: How attention engenders experience*. Oxford: Oxford University Press.
- Prinz, J. J. (2007) *The emotional construction of morals*. Oxford: Oxford University Press.
- Rand, A. (1970) *The virtue of selfishness: A new concept of egoism*. New York: Springer.
- Ridge, M. (2006) *Ecumenical expressivism: Finessing Frege*. *Ethics* **116** (2): 302–336.
- Ross, W. D. (1930) *The right and the good*. Oxford: Clarendon Press.
- Russell, B. (1897/1983) The collected papers of Bertrand Russell, Vol. 2. In K. Blackwell, A. Brink, N. Griffin, R. Rempel, and J. Slater (Eds.), *Cambridge essays 1888–1899*. London: Allen & Unwin.
- Satris, S. A. (1987) *Ethical emotivism*. Berlin: Springer.
- Satris, S. A. (1982) The theory of values and the rise of ethical emotivism. *Journal of the History of Ideas* **43** (1): 109–128.
- Sinclair, N. (2009) Recent work in expressivism. *Analysis* **69** (1): 136–147.
- Smith, A. (1759/2010) *The theory of moral sentiments*. New York: Digireads.com Publishing.
- Sober, E. and Wilson, D. S. (1998) *Unto others: The evolution and psychology of unselfish behavior*, Cambridge: Harvard University Press.
- Solms, M. and Panksepp, J. (2012) The “id” knows more than the “ego” admits: Neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Science* **2**: 147–175.
- Solms, M. and Turnbull, O. (2002) *The brain and the inner world*. New York: Other Press.

- Stevenson, C. L. (1963) *Facts and values: Studies in ethical analysis*. New Haven: Yale University Press.
- Stevenson, C. L. (1944) *Ethics and language*. New Haven: Yale University Press.
- Stevenson, C. L. (1937) The emotive meaning of ethical terms. *Mind* **46**: 14–31.
- Stroll, A. (1954) *The emotive theory of ethics*. Berkeley: University of California Press.
- Tersman, F. (2008) The reliability of moral intuitions: A challenge from neuroscience. *Australasian Journal of Philosophy* **86** (3): 389–405.
- Thagard, P. (2007) The moral psychology of conflicts of interest: Insights from affective neuroscience. *Journal of Applied Philosophy* **24** (4): 367–380.
- Urmson, J. O. (1968) *The emotive theory of ethics*. London: Hutchinson.
- Weick, K. E. (1995) *Sensemaking in organizations*. Thousand Oaks: Sage.
- Wiggins, D. (1987) A sensible subjectivism? In S. Darwall (Ed.), *Needs, values, truth: Essays in the philosophy of value* (pp. 185–214). Oxford: Blackwell Publishers.
- Williams, B. (1985) *Ethics and the limits of philosophy*. London: Routledge.

Chapter 10: Summary and conclusions

10.1 Summary of the thesis

As a response to the overarching research question of this study, the preceding discussion can be summed up under a general theme of leveraging *affective neuroscience* as a means to understanding normativity, including the foundations of moral decision-making. I have argued that normativity, as a mental phenomenon, is quintessentially affective and emotional. A central theoretical outcome is that the meta-ethical perspective of *emotivism* is restored as a theoretical framework for understanding moral psychology, moral thought and talk, and moral action and decision-making. At the outset (Ch. 1), the research question was formulated as follows: *Is it possible to explain the phenomena of normativity based on the current state of scientific knowledge about emotions, and if this is possible, what would the resulting explanatory theory look like?*

An abridged answer to the two components of this question is as follows: (1) An explanation of the phenomena of normativity *does* indeed seem possible based on an empirically informed scientific account of primary emotions. These primary emotions are identifiable with corresponding neurochemical processes and neuroanatomical circuits in the human brain (as described by Panksepp, 1998), attached to a broader capacity for basic affective consciousness, and embedded in the social context by the sensory inputs it represents. (2) The resulting explanatory theory of morality looks like a specific version of the specific meta-ethical philosophical perspective that *emotivism* represents.

I have shown that the accumulated evidence presented by *affective neuroscience* exposes central empirical underpinnings of emotivism, pinning the theory down to facts about the human brain. Furthermore, I have pointed out that the relation between subjective mind

and objective brain is specified by *dual-aspect monism*. Incorporation of this philosophy-of-mind perspective is important in order to develop a holistic emotivist theory. By doing so, it can be shown more clearly that normativity cannot straightforwardly reduce to descriptive analysis, notably because normativity represents a dimension that is inherently subjective and experiential. At the same time, dual-aspect monism recognizes that the subjectivity of the mind is caused by objective facts about the human brain and nervous system, which is approachable by methods of scientific inquiry. This theoretical integration moreover allows emotivism to connect with the current state of knowledge concerning human consciousness (which is roughly as reported by Faw, 2006).

In the philosophy literature, the theoretical perspectives of meta-ethical emotivism and dual-aspect monism have been argued for independently, respectively in moral philosophy and in philosophy of mind. I have brought these perspectives in under the same theoretical umbrella and cemented this unified position by specifying relevant evidence about how the human brain and mind work. Affective neuroscience and affect theory establish beyond reasonable doubt that emotions are affective and based on specific brain circuitry. Human affect is hierarchically structured, flowing upward from affective consciousness (located in the upper brain stem), through the emotional seeking system, and then spreading out from there to six other distinct emotional systems (primary level emotions), which in turn activate emotions systems that are conditioned by learning and memory processes (secondary level emotions), before they finally reach the deliberative parts of the brain (tertiary level emotions in the frontal lobes of the cortex).

Crucially, affect provides a subjective sense of “good” and “bad”, at the most primordial level appearing as hedonic “pleasure” and “unpleasure”. I argue that it is this basic sense of pleasure and unpleasure that grows into the recognizable moral sense of “good” and “bad”, “right” and “wrong”, “virtuous” and “vicious”, etc., as affect progresses up the

emotional hierarchy. According, I propose the *normativity-as-affect* hypothesis (Ch. 9.4), which allows us to locate normativity neuroanatomically. Specifically, I propose an internalist view according to which normativity is co-located with the neurological structures of emotional affects and should be thus identified. I have shown that rival theories about the nature of normativity and where it comes from, have weaker scientific support.

This perspective gives directions for developing a specific version of emotivism. While the basic tenets of classical emotivism are correct, they need further specification. Since no holistic theory of this kind appears to have been attempted before, I have chosen to dub this distinct theoretical perspective as *neuroemotivism*. A central proposition of neuroemotivism is that we should see normativity as affect; which, as noted, I call “the normativity-as-affect hypothesis.” The normative content of moral reasoning and decision-making is, according to this view, affective.

Neuroemotivism provides a naturalistic connection for emotivism. Another central proposition is that there is an experiential aspect of affects that cannot be reduced to physical material naturalism; namely subjective *qualia*, or the notion that affects “feel like something” that escapes description. As Frijda (2007: 199) points out, the feelings we experience “cannot be measured from the outside.” At the same time, subjective feelings cannot be ignored in a holistic account of the human mind. Hence, neuroemotivism is compatible with the theoretical perspective of *dual-aspect monism* in philosophy of science. According to this perspective, the subjective mind fully depends on, and is caused by, the brain (monism), at the same time as it is recognized that the subjective mind has an experiential dimension that escapes material reduction (the dual aspect).

The perspective of neuroemotivism is my main theoretical contribution. It brings together evidence and insights that have previously seemed unrelated and disciplinarily far

apart. This is apparent when we consider the fact that emotivism seems to have gone unnoticed by neuroscientists. Prominent neuroscientist like Antonio Damasio and Jaak Panksepp refer to classical philosophers such as Aristotle, Spinoza, Hume, and Kant, as well as to recent developments in philosophy of science, but appear less aware of developments in modern moral philosophy, and notably emotivism. By bringing these intellectual endeavors closer together and creating a theoretical unity between them, my hope is that philosophers, scientists, and practitioners will be equipped to deploy the concept of *normativity* in a new and meaningful way.

Hence, I argue that scientists, philosophers, and practitioners do not need to reject the notion of normativity, dismissing it as “normativism”, for example, in the manner Stephen Turner ends up doing (discussed in Ch. 1.2 and 1.3). Instead, neuroemotivism provides a scientifically informed and useful way of conceptualizing normativity. This conceptualization locates normativity *internally* in biologically based affective brain structures, rather than in the external social context which is the position that Alasdair MacIntyre and several other prominent philosophers have been inclined to (see Ch. 1.2), or in the “fabric of the universe” as Henry Sidgwick and Gerald Moore and their contemporary followers claim (see Ch. 1.3).

I have devoted one chapter (Ch. 7) to argue *for* meta-ethical relativism, but at the same time against facile normative extrapolations represented by specific versions of normative relativism. I have devoted another chapter (Ch. 8) to argue *against* objective values and *externalist* meta-ethical accounts of ethics. Together these chapters provide *philosophical* reasons in support of locating normativity internally in subjective experience. My main aim, however, is to present a more detailed picture of how normativity can be explained in terms of affective emotions straddling different layers of consciousness.

While normativity has almost exclusively been an interest among philosophers, I argue that it is now in principle accessible to scientists because it is possible to make sense of the term in relation to science. I meet *Anscombe's challenge*, as specified in the introduction (Ch. 1.3) empirically through *the location argument* (Ch. 4) and theoretically through *neuroemotivism* (Ch. 9.4). Practitioners and professionals who deal with normative issues can benefit from this by understanding how normativity relates to science, and especially to human affect and emotions, adding to their body of expertise. This has relevance for any practitioner or professionals implicitly or explicitly involved in applied ethics (e.g., business professionals, medical professionals, decision-makers), or who otherwise deal with affective-normative issues (e.g., psychologists, psychotherapists, psychiatrists).

The affective revolution in neuroscience provides evidence that necessitates reformulating theories of emotions, consciousness, and in turn, social psychology, decision-making and all practical fields that depend on understanding the moral nature of social human beings. The paradigmatic changes that take place, therefore, have wide-ranging theoretical implications. They are also likely to have many indirect practical implications, because some ways of thinking that bear on practical decisions turn out to be untenable. For example, it has now become apparent that one should *embrace* rather than strive to eradicate the role of affective emotions in moral decision-making (Ch. 6), although the particular role of emotions requires careful attention.

It should be pointed out that a cost of the theoretical perspective I have presented is that we cannot help ourselves to objective justification in ethical theory and moral practice. Conversely, professionals, practitioners, and decision-makers involved with normative questions have a stronger and more reliable case for justifying their decisions subjectively and inter-subjectively in place of objective justification. I call this type of justification “justification light” (Ch. 9.5). This perspective on justification can be seen as a de-

mystification and a humanizing of normative ethics, which unites us with our biological nature (see Ch. 5).

As a byproduct of my account of justification, some of the normative issues concerning moral responsibility become more tractable. For example, it may now seem plausible that moral responsibility can be parsed out to individual decision-makers in proportion to their involvement in a given decision. This avoids the problems represented by more nebulous notions of responsibility, such as ‘collective responsibility’ and ‘corporate social responsibility’, by which individual responsibility is rendered opaque. At the same time, it is recognized that actual individual responsibility is constrained by the limited freedom of will of the agent, as shown by studies in experimental psychology and neuroscience. Moral responsibility is exposed as a psychologically complex notion resulting from exchanges between individual biology and socio-cultural factors. There are limitations as to the extent to which it is possible for a given person to act as a free agent. The precise degree of free will and agency, however, remain inadequately specified by empirical science. As a result, ascribing moral responsibility seems viable and indeed important as a social practice but should be done with care and moderation, and the realization that such ascriptions are fallible.

10.2 Main conclusions

(1) *The sources of normativity*, at the foundational level, are internal to human beings and biologically based. They can be traced down neuroanatomically to specific emotional neural circuits in the brain. Each of these neural circuits function by means of sophisticated neurochemical and electrical communication, and is located above the spinal cord and below the neocortex, and includes several subcortical nuclei. This is further specified in chapter 4.

(2) *The experience of normativity* is an affective emotional experience that is inherently subjective and affectively conscious. Emotional affect is necessary for valence, values, and valuation; more generally for perceiving matters from a normative perspective. Affect is always present in our lived lives and cannot be switched off at will. Affect is necessary for experience, motivation, and our sense of self. This is elaborated on in chapter 5.

(3) *Foundational normativity*, conceived of as psychological and a biologically based phenomenon, is a premise for such activities as proposing, defending, endorsing, accepting, creating, and co-constructing morality in inter-subjective social space. These inter-subjective social processes, more generally the social embeddedness of morality, make human morality characteristically relational and social. However, this sociality depends on neuropsychological processes that take place within the mind and brain of individual persons. Foundational normativity and first-order personal morality is a premise for second-order socially constructed morality. This topic is discussed in chapter 9.

(4) Normativity starts out as affective *implicit normativity* before it can be transformed into the *explicit normativity* that enters awareness as reasons and intentions in the minds of decision-makers. This important distinction is pointed out in chapter 6.3. I take it that we are under the influence of implicit normativity insofar as we attain rudimentary affective consciousness (see Ch. 5), and only sometimes guided by explicit normativity. I propose that implicit normativity is below the radar of awareness but is at the same time affectively experienced.

(5) *Dual-aspect monism* is the best theoretical approximation to reconcile the internal subjective reality that we experience with the notion of a physical external objective reality. Dual-aspect monism suggests that the subjective mind and the physical brain are two different aspects of the same thing. At present, empirical evidence appears insufficient to establish

scientifically a more detailed explanation of the mind-brain relation and the exact emergence of subjective consciousness. I have shown that emotivism is supported by dual-aspect monism, which is discussed in more detail in chapter 3.

(6) All *emotions* have an affective core that is subcortically based in the emotional neural circuits described above (and in Ch. 4). Three layers of emotions can be distinguished neuroanatomically; at the bottom there is “raw affect” or “primary emotions”, subsequently secondary emotions associated with learning and conditioning, and finally cognitively conglomerated tertiary emotions that feature in explicit reasoning and conscious awareness. These layers of emotions appear to be structured in a nested hierarchy, where lower emotions levels are components of higher-level emotions. There are at least seven basic or primary emotions.

(7) *Emotional circuits* have been identified with some degree of neuroanatomical and neurochemical specificity. Presently, at least seven distinct emotional circuits are identified well within margins of scientific confidence. These are accounted for in chapter 4.

(8) *Normative content* is necessary for making moral decisions and engaging in moral reasoning. Normative content is supplied by affective emotional brain structures, and gives meaning to mental object representations of the external world and from imagination.

(9) *Normative structure* is superimposed on normative content by means of sophisticated cognitive capabilities. Cognitive capabilities allow for normative content to be projected onto perceived objects in the external world. They also allow for extending normative content temporally into the future or back to memories of past events. Cognitive capabilities are furthermore important for thinking in terms of specific desires and giving normative structure to them.

(10) *Affectively based desires* are psychologically structured into a hierarchy with two layers. At the bottom there are first-order desires; at the top, second-order desires. Second-order desires are desires about first-order desires. While relatively stable over time, second-order desires give rise to subjective values.

(11) Normative content together with normative structure superimposed by cognitive processes readily provide normative guidance for decision-making and deliberation whenever not obstructed from doing so. Obstructions to affectively based decision-making can partly cripple moral decision-making.

(12) *Meta-ethical relativism* should be accepted because it accurately portrays the nature of normativity and moral values as well as their inter-personal role. Normativity at the foundational level should be seen as relative to persons and their motivational sets of values and desires.

(13) Four distinctive versions of *normative relativism* can be identified. I have termed these versions ‘normative cynicism’, ‘normative syncretism’, ‘normative nihilism’, and ‘normative collectivism’, respectively. It is argued that each of these normative perspectives is problematic and should be rejected. The two first versions exhibit an unstable normative structure that leads to circular or arbitrary argumentation. The third version, in effect denies the subjective normative content of morality, and should be rejected because it does not provide a basis for morality. Finally, the fourth version of relativism places morality within supra-individual social structures and thereby severs the connection to personal morality based in subjective affective experience and emotions. Hence, it is for example unable to provide an account of personal moral responsibility.

(14) The notion of objective values should be rejected. Although belief in objective normative values is relatively widespread, there seems to be no reasonable way of

substantiating such beliefs (see Ch. 8). Consequently, they should be treated as untenable. From a practical point of view, it seems that belief in objective values is not needed in order to establish an adequate theory of morality, and that moreover they represent a potential threat to affectively based subjective values as a basis for normativity.

(15) *Emotivism* is an epistemological philosophical perspective that explains ethics or morality in terms of emotions. Given the strong analogies between normativity and emotions surrounding motivational function and subjective experience, it seems plausible to view normativity in terms of emotions; the proposition I call the “normativity-as-affect hypothesis”. Accepting this hypothesis allows for a relatively detailed and biologically based account of normativity, which draws on recent advance in affective neuroscience. This is the version of emotivism I have termed *neuroemotivism*. Besides offering a better understanding the nature of normativity, it also allows a *pro tanto* identification of the sources of normativity mapped in terms of the current neuroanatomy of affective emotions. Affective neuroscience represents the scientific underpinnings of emotivism, enabling it to provide a scientifically corroborated and empirically adequate explanatory account of moral thought, talk, and action.

(16) Complete and objective *moral justification* is likely to be an unattainable ideal. Moral theories that require meeting such ideals should therefore be rejected. This does not imply that it is nonsensical to have moral ideals to stretch toward in moral practice. Neither does it imply that there is no sense in which we can talk about moral justification. I propose that there is an inter-subjective sense of moral justification where we justify our moral actions and decisions towards each other; subject to subject. The recognition of this mode of justification is of practical importance to personal ethics as well as professional ethics, business ethics included.

(17) It is argued that the endorsement of meta-ethical relativism, emotivism, and a basis of normativity in subjective experience is *in no way committed to egoism*. This mistaken conclusion can be referred to as the ‘egoism fallacy’. In its most basic form, the egoism fallacy involves erroneously equating the normative content of morality with what is beneficial for the person whose contentment this is.

10.3 Challenges for moral theory and practice

In this section, I raise some controversial issues that are relevant for practical decision-makers and professionals when they themselves face moral issues or provide advice for others. Herein I offer some thoughts about how practitioners and professionals ought to incorporate ethical theory and ethical principles into their practices.

Applied ethics, or the application of ethical theory to practical moral problems, typically emphasizes the importance of *giving reasons* and *justifications* for one’s moral decisions and evaluations. This is unquestionably important for social and communicative purposes in *socially dense* contexts, such as we typically find in the organizational context and the professional context. Both of these contexts are highly *relational*.

Sometimes the practice of giving reasons and justifications is conceived of as the quality stamp that separates genuine “ethics” from mere “morality”; being legitimate, theoretical, reflectively approached ethical correctness as opposed to the malleability and thoughtlessness of practice and subjectivity. I wish to draw attention to the danger, assuming that that is what it is, of reducing the handling of moral issues to the provision of justification. In particular, I propose that we should make an effort to understand the psychology behind moral deliberation, and that we should *challenge* our emotions by *engaging* them in a direct manner.

There are several problematic issues attached to the practice of ethical justification; all of which are mentioned in previous chapters. These include: (1) the lack of a single ethical theory that is uncontroversial, (2) the problem of insufficient epistemic access to and knowledge about our own actual motivation in practical cases of moral decision-making and evaluation, (3) the fact that the provision of ethical justification implies an element of normative relativism (specifically normative collectivism; see Ch. 7.2.6) insofar as it involves conforming to prevailing standards and norms that regulate the particular social practices of justification. These standards and norms are culturally relative, since they are established by the current way of doing things in time and place.

It is of little help to attempt to abate this in the manner of Kelsen (1967), by appealing to a conception of fundamental law or “Grundnorm”, or to Kant’s famous law formulations of the “categorical imperative.” As argued, none of these proposals turn out to be self-satisfactory, and I have argued, more generally, that externalism about normativity is speculative as well as doubtful (Ch. 1 and Ch. 8).

Looking at human morality from the vantage point of human biology, psychology and evolution, permits a radically different picture. First of all, we have a working model of what normativity boils down to. I suggest, as I have consistently argued, that normativity should be seen in terms of affective emotions. From this picture, moreover, it seems evident that ethical theory and reflective thought can sometimes clutter otherwise fine-tuned affective moral sensibilities. For example, there are moral cases where, in Bernard Williams’ famous phrase, taking appropriate action is obstructed by “one thought too many” (Williams, 1981: 18); should I just go ahead and save my child from the sinking ferry, or does impartiality require me to toss a coin about saving my child versus saving the child of a stranger?

At this point, we have the option of proceeding deeper into rationalistic lines of moral reasoning (just as we might have learnt in a course on applied ethics). Which normative ethical theory should be applied? What criteria justify the choice of ethical framework? What specific ethical principles should my justification appeal to in that theory? How should conflicts between ethical principles be resolved in a rational manner? None of these exercises actually elucidates moral issue unless they somehow manage to bring our moral feelings to bear on the particular issue. Our aim as moral agents should not, I contend, be simply to find a solution to the moral problem at hand, but a satisfactory solution; and by satisfactory I mean a solution that first and foremost engages our *affective subjectivity*; our emotional systems and or affective consciousness.

Examples such as those Bernard Williams points to, show that cognitive thought sometimes can divert attention and resources away from what matters in regard to appropriately addressing a moral issues. However, the problems are arguably more serious and entrenched than this. Not only can cognition divert attention away from dealing with the moral issue at hand, but cognitive processes directly *repress*, *suppress* and *inhibit* affect. Berlin (2011) points out that there is explicit effortful *suppression* as well as automatized implicit *repression* of the affect. One potential effect of taking ourselves as object, as we do in cognitive moral reasoning, is that we *alienate* ourselves from our subjective selves.

As pointed out in the preceding chapters (see e.g., Ch. 2.7.3 and Ch. 6.3.3), it is well documented that cognition inhibits affect. These suppressive and repressive effects are likely to be exacerbated if we continuously strive to make ethics as explicit as possible and in line with the various criteria of rationality and ethical theory. I am suggesting that, paradoxically, we could, in part due to our own efforts and schooling, be in danger of losing touch with our affective human nature.

Being out of touch with our affects is a matter that should not be taken lightly. Attached to our affective nature are feelings of *empathy* for fellow human beings, as well as biologically built-in humanity stemming from central moral emotions such as *care*. If we lose or significantly deplete these capacities, will we not thereby lose some of the most important premises of human morality and moral responsible action?

I am not suggesting that cognition and our sophisticated capability for taking ourselves as objects are unimportant. What I am suggesting is that moral agents need something in *addition* to this, and something that partly counter-balances the effects of the cognitive reasoning approach.

The 19th century Norwegian playwright, Henrik Ibsen, presented a well-known metaphor for the human psyche and the search for the self at its core by likening it to an onion that can be peeled layer by layer (Ibsen, 1993/1867), expressed by the character *Peer Gynt*. Ibsen's psychological insight deftly pinpoints the difficulty of finding oneself, one's inner moral self, by peeling from the outside view of oneself as an object. I believe we should welcome Bergson's (1935) recommendation that we should move away from strictly rationalistic approaches to morality and focus more on understanding the biological premises of human morality; that we should emphasize "open morality" over "closed morality."

According to Bergson's outlook, *open morality* is dynamic, experiential, and driven by internal aspiration, whereas *closed morality* is static, structural, and driven by external pressure. He associates open morality with the morality of a person, and closed morality with the morality of the group. However, it is not just the normative tension between the individual and group that is the issue here, but also the psychological normative tension *within* the individual. Bergson (1935: 82) makes it clear that "[A]ll morality, be it pressure or aspirations, is in essence biological."

The *professional role*, because it is artificial in the sense that it moves us away from our ordinary way of living and our ordinary identities, is likely to be particularly heavily leaning toward cognitive mental processing, both the implicit and explicit kind. The professional role therefore is likely to widen the distance between “ourselves as objects” and “ourselves as subject” or what Damasio (1999: 17) calls our “core self”. The professional role becomes yet another layer of seeing ourselves as objects, and of sifting information from the external context.

A caricature may help us see which direction the process of professionalization may lead. Ishiguro (1989) wrote a novel, *The remains of the day*, about a classical British butler. Mr. Stevens, the butler in this story, epitomizes the professional ethos. He holds dignity as a supreme professional value, and observes the professional principle of absolute loyalty, regardless of whatever social and moral implications this might have. He is always self-controlled and never expresses emotion, holding an ideal of cold rational thought. He is always self-effacing, never self-assertive. The point, of course, is that he has lost sight of his real affective self, and that he is in danger of *becoming* that professional which he *enacts*.

What could the practical upshot of all this be? What measures could professionals, decision-makers, advisors, and educators take in order bring out the affective-normative content of their subjective self in appropriate ways? I think that, in order to promote a humane morality, it is essential to emphasize, as Slovic (2007: 79) aptly formulates it, “the capacity to experience *affect*, the positive and negative feelings that combine with reasoned analysis to guide our judgments, decisions, and actions.” One way to try to achieve this is by engaging our creative capacity for imagination and emotional visualization of expected consequence that may result from relevant courses of action.

Slovic (2007) emphasizes that our affective capacities has clear shortcomings. While we for example are good at empathizing with one and even two others, our ability to empathize with greater numbers of people seems severely limited. The important point is that precisely *because* of these shortcomings we need to deploy our affective capacities to the fullest extent, and by combining our cognitive capabilities and our affective capacities in a fertile symbiosis. Thereby, we should be as well equipped as possible, within our biological and psychological limitations, to address the moral problems that face others and us.

10.4 Limitations

Firstly, I should mention that am aware of the fact that the theoretical view I have assembled and developed into a unified theory, *neuroemotivism*, needs further development. I have merely stipulated what this theoretical perspective looks like and what it seems to imply practically and theoretically. Fortunately, its components are richly discussed in the corresponding literature. However, neuroemotivism needs to be explored practically as well as theoretically; it needs to be developed and specified further, and it needs to be rigorously tested and discussed in the scientific literature as a unified theory. One of the philosophical areas that need further scientific and philosophical scrutiny is the normativity-as-affect hypothesis.

Secondly, I am left humbled by the evident fact that there is a lot more to learn and many avenues of scientific inquiry to explore (see the following section). In writing this manuscript as a coherent piece, I had to make many delimiting choices. My aim through and through has been holistic in the sense that I have refrained from closing disciplinary doors. Nevertheless, the “whole” of the foundations of normativity is too big to cover within the space of a few hundred pages, especially considering the wealth of empirical detail with which the scientific aspect of normativity in principle could be specified, including physics

and chemistry. I have had to be selective, and certain disciplines have therefore been emphasized over others. Guided by the particular angle of my research questions and the hypothesis that emotions is a fruitful avenue to explore normativity, I have focused mainly on three disciplines; namely neuroscience, psychology, and philosophy. Notably, this leaves out important disciplines such as psychiatry, biology, sociology, economics, and cultural anthropology. Furthermore, I have had to limit my discussion to specific areas within the broad disciplines I draw upon.

Thirdly, my own expertise within my focal disciplinary anchors is limited. By formal education, I am neither a psychologist, nor a professional philosopher, nor a neuroscientist. While I feel that philosophy is somewhat liberal and liberating as long as one has a sufficient broad and deep grasp of the relevant discussions and seminal works, I have relied on citations and extensive literature referencing in order to make accurate statements concerning the heavily empirical and fact-oriented literature of neuroscience. Psychology, in this respect, falls somewhere between philosophy and neuroscience, but I found that once central factual points in neuroscience were pinned down, it facilitated reading and assessment of the psychology literature. Undoubtedly, there remains much for me to learn in all three disciplines.

Affective neuroscience, as I have argued, constitutes a major paradigmatic shift. The process of substituting and modifying mental models will undoubtedly take time even within neuroscience. Subsequently, or in concert, scientific mental models will also have to change in various affected domains across the social sciences. Finally, it will trickle down to the ‘applied’ sciences. My aim was in many ways to support this process, but I recognize that my contribution can only amount to a small fraction of what is to come.

10.5 Suggestions for future research

There are many avenues to proceed from the endpoint of my study, and a vast number of questions are left to be resolved. One avenue that I find especially compelling is to develop *neuroemotivism* and make a more detailed statement of this theoretical perspective. While the key components of this theory seem to be in place, its conceptual attributes need to be clarified and discussed in more detail, not least in relation to various relevant theoretical areas that have to be further explored. In extension of such a theory-development pursuit, the practical and theoretical implications of neuroemotivism also need to be explored more extensively and stated more explicitly.

One research option is to specify the fuller details and implications within one of the disciplinary perspectives that have been presented. Another option is, reversely, to go in the direction of integrated science, by integrating additional disciplines. There are a number of good candidates concerning the latter option. Some additional disciplinary perspectives are already in development along the scientific and theoretical lines presented, but could not be included due to necessity of limiting the scope of the thesis. Additional disciplinary perspectives include psychiatry, psychoanalysis, and especially affective neuropsychology, which has been independently developed by respectively Mark Solms and Georg Northoff, both of whom I have cited several places (e.g., Northoff, 2011). In addition, broad social science disciplines such as law, economics, sociology, and political science appear to be interesting areas.

One thing is to expand integration; another important task is to investigate the fuller implications for each sub-discipline and for the integrated scientific understanding. At the foundational level, science itself seems to offer a range of interesting questions. Philosophy, including (amongst others) moral philosophy, philosophy of science, philosophy of mind, and

philosophy of action, also appear to be areas of investigation and development. A particularly interesting topic is questions regarding the freedom of will. This was mentioned in the thesis, but I made no attempt at a full exposition. This would be an excellent area to explore, and the understanding of human nature as fundamentally affective would be an important background for researching the set of questions this area raises.

Chapter 10 References

- Bergson, H. (1977) *The Two Sources of Morality and Religion*. (R. A. Audra and C. Brereton, Trans.). Paris: University of Notre Dame Press. (Original work published in 1935).
- Berlin, H. (2011) The neural basis of the dynamic unconscious. *Neuropsychoanalysis* **13** (1): 5-31.
- Damasio, A. R. (1999) *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.
- Faw, B. (2006) Are we studying consciousness yet? *Journal of Consciousness Studies* **13** (4): 94–112.
- Frijda, N. H. (2007) *The law of emotions*. Mahwah: Lawrence Erlbaum Associates.
- Ibsen, H. (1993) *Peer Gynt: A Dramatic Poem*. J. Northam (trans.). New York: Oxford University Press. (Originally published in 1867).
- Ishiguro, K. (1989) *The remains of the day*. London: Faber and Faber.
- Kelsen, H. (1967) *The pure theory of law*. Berkeley: California University Press.
- Northoff, G. (2011) *Neuropsychoanalysis in practice: Brain, self and objects*. Oxford: Oxford university press.
- Panksepp, J. (1998) *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.
- Slovic, P. (2007) ‘If I look at the mass I will never act’: Psychic numbing and genocide. *Judgment and Decision Making* **2** (2): 79–95.

Solms, M. and Panksepp, J. (2012) The “id” knows more than the “ego” admits: Neuropsychoanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Sciences* **2**: 147–175.

Williams, B. (1981) Persons, character, and morality. In B. Williams (Ed.) *Moral Luck: Philosophical papers 1973-1980* (pp. 1-19). Cambridge: Cambridge University Press.

iii. List of tables and figures

Figures:

Fig. 1 – Outline of the main argumentative structure	p. 46
Fig. 2 – Lazarus’s cognitive appraisal theory of emotion	p. 81
Fig. 3 – Zajonc’s affective theory of emotion	p. 82
Fig. 4 – LeDoux’s fast route	p. 92
Fig. 5 – LeDoux’s slow route	p. 93
Fig. 6 – The James-Lange theory of emotion	p. 109
Fig. 7 – The Cannon-Bard theory of emotion	p. 111
Fig. 8 – The Schachter-Singer theory of emotion	p. 113
Fig. 9 – A taxonomy of theoretical positions within relativism	p. 295

Tables:

Table 1 – Theoretically and empirically supported basic emotions	p. 120-121
Table 2 – Affect vs cognition	p. 126

Fin
