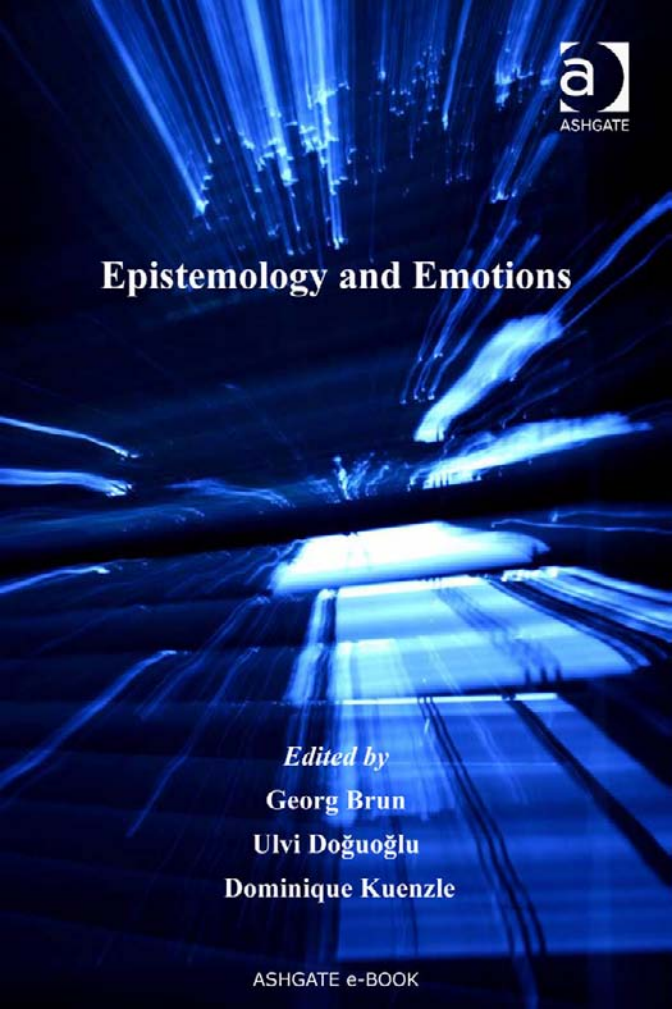
ASHGATE
Epistemology and Emotions
Edited by
Georg Brun
Ulvi Doğuoğlu
Dominique Kuenzle
ASHGATE e-BOOK

EPISTEMOLOGY AND EMOTIONS

Undoubtedly, emotions sometimes thwart our epistemic endeavours. But do they also contribute to epistemic success? The thesis that emotions 'skew the epistemic landscape', as Peter Goldie puts it in this volume, has long been discussed in epistemology. Recently, however, philosophers have called for a systematic reassessment of the epistemic relevance of emotions. The resulting debate at the interface between epistemology, theory of emotions and cognitive science examines emotions in a wide range of functions. These include motivating inquiry, establishing relevance, as well as providing access to facts, beliefs and non-propositional aspects of knowledge.

This volume is the first collection focusing on the claim that we cannot but account for emotions if we are to understand the processes and evaluations related to empirical knowledge. All essays are specifically written for this collection by leading researchers in this relatively new and developing field, bringing together work from backgrounds such as pragmatism and scepticism, cognitive theories of emotions and cognitive science, Cartesian epistemology and virtue epistemology.

Ashgate Epistemology and Mind Series

Series Editors:
Dr Bill Brewer, St Catherine's College, Oxford
Professor Stewart Cohen, Arizona State University, Tempe

The Ashgate Epistemology and Mind series presents contemporary international research at the intersection of metaphysics, epistemology, mind and language. Rooted in an awareness of the great epistemological and metaphysical issues dominating early modern philosophy, but focussed on contemporary, cutting edge research, the series aims to be inclusive and titles cross sub-disciplinary boundaries and integrate philosophy with psychology and other social sciences and the sciences. Offering research-led, yet intellectually accessible, stimulating new contributions to each topic for a wide international breadth of readers, each book breaks new ground in contemporary research to stimulate fresh debate in epistemology and philosophy of mind. This series of individually authored books is framed by a cluster of specially commissioned volumes of essay collections edited by leading scholars in the field. Particular focus areas for the series include: The A Priori; Other Minds; Psychological Explanation; Rationality and Norms of Reason; Empiricism; Contextualism; Direct Perception; Virtue Epistemology; Deductive Argument; Transcendental Arguments; Emotion and Expression; Privileged Access and First Person Authority.

Epistemology and Emotions

Edited by

GEORG BRUN
ETH Zürich, Switzerland

ULVI DOĞUOĞLU
University of Zürich, Switzerland

DOMINIQUE KUENZLE
University of Zürich, Switzerland

ASHGATE

© Georg Brun, Ulvi Doğuoğlu and Dominique Kuenzle 2008

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the publisher.

Georg Brun, Ulvi Doğuoğlu and Dominique Kuenzle have asserted their moral right under the Copyright, Designs and Patents Act, 1988, to be identified as the editors of this work.

Published by
Ashgate Publishing Limited
Gower House
Croft Road
Aldershot
Hampshire GU11 3HR
England

Ashgate Publishing Company
Suite 420
101 Cherry Street
Burlington, VT 05401-4405
USA

www.ashgate.com

British Library Cataloguing in Publication Data
Epistemology and emotions. – (Ashgate epistemology and mind series)
1. Knowledge, Theory of 2. Emotions (Philosophy)
I. Brun, Georg II. Doğuoğlu, Ulvi III. Kuenzle, Dominique
121

Library of Congress Cataloging-in-Publication Data
Epistemology and emotions / edited by Georg Brun, Ulvi Doğuoğlu, and Dominique Kuenzle.
p. cm. – (Ashgate epistemology and mind series)
ISBN 978-0-7546-6114-6 (hardcover : alk. paper)
1. Knowledge, Theory of. 2. Emotions. I. Brun, Georg, 1963– II. Doguoglu, Ulvi, 1968– III. Kuenzle, Dominique, 1973–

BD161.E6225 2008
121–dc22

2008004812

ISBN 978-0-7546-6114-6

Printed and bound in Great Britain by
MPG Books Ltd, Bodmin, Cornwall.

Contents

This page intentionally left blank

List of Contributors and Editors

Georg Brun. Senior Research Fellow, Environmental Philosophy Group, Institute for Environmental Decisions, ETH Zürich.

Ulvi Doğuoğlu. Senior Research Associate, University of Zürich

Daniel Dohrn. Lecturer, Department of Philosophy, Ludwig-Maximilians-University, Munich.

Sabine A. Döring. Professor of Philosophy, University of Tübingen.

Catherine Z. Elgin. Professor of the Philosophy of Education, Graduate School of Education, Harvard University.

Peter Goldie. Samuel Hall Chair in Philosophy at The University of Manchester.

Christopher Hookway. Professor of Philosophy, University of Sheffield.

Dominique Kuenzle. Research Assistant, Department of Philosophy, University of Zürich.

Ronald de Sousa. Professor Emeritus of Philosophy, University of Toronto; Fellow of the Royal Society of Canada.

Alessandra Tanesini. Reader in Philosophy, Cardiff University.

Paul Thagard. Professor of Philosophy, with cross appointment to Psychology and Computer Science, and Director of the Cognitive Science Program, at the University of Waterloo.

Markus Wild. Research Assistant, Department of Philosophy, Humboldt-University Berlin.

This page intentionally left blank

Guide to the Essays

The essays collected in this book are all concerned with the role of emotions in philosophical epistemology. In the introductory chapter, **Georg Brun** and **Dominique Kuenzle** offer a survey of theories of emotions and developments in epistemology in order to determine what epistemological relevance can, and has been, assigned to emotions.

Catherine Elgin and Christopher Hookway were among the first epistemologists to argue that emotions contribute to our epistemic success. In her contribution to this collection, **Catherine Z. Elgin** analyses epistemic functions of emotions, drawing attention to parallels and differences between emotions on the one hand, and perceptions and beliefs on the other. She argues that, like perceptions, emotions deliver representations of various kinds of facts and, like beliefs, emotions act as sources of salience.

According to **Christopher Hookway**, emotions can play a part in the acquisition of beliefs that are epistemically immediate in the sense that their justification does not depend on consciously accessible reasons. Focusing on doubt, Hookway shows how the distinction between 'felt' doubts and idle 'paper doubts' can serve internalist purposes, insofar as epistemic subjects ought to act on felt doubts only. The insight of externalism is that affective evaluations, such as felt doubt, need not be accessible to reflection.

Alessandra Tanesini invokes emotions to develop a version of fallibilism that is based on humility as an intellectual virtue. Several accounts of fallibilism are examined and rejected before intellectual humility is introduced and characterized partly by appeal to salience generating properties of emotions. The resulting account of fallibilism is then shown to meet the criteria of adequacy that rival theories fail to meet.

The following two chapters approach from a different angle, and in more detail, the analogy between emotions and perceptions, as well as epistemic immediacy. **Sabine A. Döring** analyses and utilizes the similarity of the epistemic deliverances of emotions and perceptions. She develops the view that differences in content and attitude make it possible that there can be conflicts between beliefs and emotions with respect to the question of how they represent the world, even though such a conflict does not amount to logical inconsistency.

Daniel Dohrn focuses on Hookway's work on epistemic immediacy. He argues that while epistemic immediacy may be relevant for knowledge acquisition, emotions cannot justify such knowledge. This argument is based on a preference for 'Cartesian' epistemologies stressing the relevance of reflective assessment, as opposed to epistemologies that emphasize the role of immediate affective valuations.

Markus Wild offers another critical assessment of emotions' role in epistemology. He claims that emotions should neither be conceived as advancing knowledge by tracking salience, nor should they be seen as regulating inquiry. The problems that

prevent emotions from performing these epistemic functions are not solved, but only replaced by new problems, if we invoke epistemic virtues. Wild ends by making room for a more modest version of affective epistemology.

Even if the epistemic yield of emotions is acknowledged in principle, we must not forget that emotions may still 'skew the epistemic landscape'. **Peter Goldie** diagnoses an overly optimistic trend in recent epistemological theorizing about the emotions, which he sees as fuelled by their role in the 'fast and frugal heuristics' of our bounded rationality. Starting with Hume's proposed corrections of aspects of moral emotions, Goldie draws attention to how emotions can systematically mislead us. He explains this by appeal to environmental mismatch; emotion-based heuristics have emerged in specific environments, some of which have changed.

Paul Thagard assimilates emotions to beliefs and desires by conceiving of all these states and processes as patterns of neural activities. He starts with an argument against propositional attitudes, which hinges on the claim that inference to the best explanation can never appeal to abstract objects. Thagard then presents an account of the neurophysiological interconnections between cognition and emotion that allows for emotions to frequently contribute to the growth of knowledge.

In the final chapter, **Ronald de Sousa** focuses on specifically epistemic feelings, which he classifies according to their object and the phase of knowledge acquisition in which they occur. A sample of experimental findings shows firstly how such feelings interact in various ways with other cognitive functions and secondly the important role they play in bridging different systems of mental processing.

Introduction

A New Role for Emotions in Epistemology?

Georg Brun and Dominique Kuenzle*

This chapter provides an overview of the issues involved in recent debates about the epistemological relevance of emotions. We first survey some key issues in epistemology and the theory of emotions that inform various assessments of emotions' potential significance in epistemology. We then distinguish five epistemic functions that have been claimed for emotions: motivational force, salience and relevance, access to facts and beliefs, non-propositional contributions to knowledge and understanding, and epistemic efficiency. We identify two core issues in the discussions about such epistemic functions of emotions: First, even though it is plausible that emotions are involved in epistemic processes, it may be doubted whether they really matter for the normative question of what counts as knowledge or justified belief. Second, some of the epistemic functions claimed for emotions in general may only be attributed to some specifically epistemic emotions, which have been present all along in traditional epistemology, albeit under different labels such as 'intuitions'.

Epistemic activities can be very emotional affairs. Curiosity, doubt, hope and fear trigger everyday cognitive activities as well as academic research, which in turn are sources of surprise, frustration and joy. Less intellectual emotions may also play their part when tireless scrutinizing is driven by jealousy, or when an experiment is too disgusting to occur to any researcher.

Nevertheless, emotions did not play a significant role in traditional epistemology and if they were paid any attention at all, they were mainly thought of as impairing cognition. Recently, however, epistemologists and emotion theorists have started to discuss the question of whether the epistemological standing of emotions needs to be reassessed. Are there epistemic functions that can be assigned to emotions? And which emotions are suitable candidates for these functions? These questions are at the centre of this collection of essays.

The significance some epistemologists have attributed to emotions over the last ten years or so can arguably be claimed to be new in the context of contemporary English speaking epistemology. Emotions entered epistemology discussions in the 1990s after having been reintroduced to ethics and moral philosophy some decades earlier. This development has been helped by the rediscovery of emotions in cognitive science (Antonio Damasio's *Descartes' Error* among the best-known

* We would like to thank Christoph Baumberger, Monika Betzler, Simone Dohle, Gertrude Hirsch Hadorn, Michael Roth and Alessandra Tanesini for helpful comments on earlier drafts.

examples[1]) and by epistemology becoming more closely associated with action theory and moral philosophy, as in virtue epistemology. But while the cognitive significance of emotions was quickly acknowledged and, under headings like 'emotional intelligence', made it to newspapers, general interest magazines and self-help books, most epistemologists have been less enthusiastic about emotions. For instance, in Blackwell's 1992 *Companion to Epistemology* (Dancy and Sosa 2001), 'emotion' is not even listed in the index (neither are related terms, such as 'feeling' or 'affect') and nothing significant can be found in Kluwer's 2004 *Handbook of Epistemology* (Niiniluoto et al. 2004). Over the last decade, however, many strands of research centring on the nature and function of emotions have led to important insights and adjustments, both within and outside of epistemology.

While this introduction focuses on recent research, one should not forget that the idea that emotions matter a great deal in epistemology has a longer history. An example can be found in a surprising passage of Moritz Schlick's *On the Foundation of Knowledge*, where he explicitly uses satisfaction, fulfilment and even joy as the criteria for successful validation of inductively achieved hypotheses:

> [We] pass an observational judgement that we expected, and have in doing so a sense of fulfilment, a wholly characteristic satisfaction; we are content. It is quite proper to say that the affirmations or observation statements have fulfilled their true mission, as soon as this peculiar satisfaction is obtained. … Once the prediction comes to pass, the aim of science is achieved: the joy in knowledge is joy in verification, the exaltation of having guessed correctly. … Are our predictions actually realized? In every single case of verification or falsification an 'affirmation' answers unambiguously with yes or no, with joy of fulfilment or disillusion. The affirmations are final. (Schlick [1934], 382–3)

Affirmations ('Konstatierungen') are a core element of Schlick's foundationalism and, it turns out, they have their 'true mission' in eliciting emotional responses. The immediate joy of seeing a hypothesis confirmed and the disappointment of experiencing it falsified make it possible for affirmations to provide the infallible foundation of knowledge and science. This basic function of emotions is echoed in Quine's observation that Goodman's 'new riddle of induction' is best dealt with by appealing to a feeling of simplicity (Quine 1960, 19). Recently, Christopher Hookway picked up on this remark (Hookway 2003a, 81; Hookway, this vol., p. 54), using it as a starting point for his own contribution to the current debate on emotions and epistemology. Goodman himself started to highlight the epistemological significance of the emotions in the 1960s (Goodman 1976, ch. VI.4). His arguments for adopting understanding, instead of knowledge, as the central epistemic goal were taken up by Catherine Elgin, who developed a comprehensive account of epistemology that gives emotions a prominent role (Goodman and Elgin 1988; Elgin 1996).

In what follows, we first sketch some traditional stances and more recent developments in epistemology (section 1) and the theory of emotions (section 2). On this basis, we will then (section 3) present a survey of various ways in which emotions recently have been claimed to be relevant to epistemology, followed by a

1 See Lazarus (1999) for a short historical account.

brief discussion of some possible objections to the proposed reappraisals of emotions in epistemology (section 4).

1. Background in Epistemology

Recent developments within philosophical epistemology have prepared the ground for attributing epistemic significance to emotions. These developments are best understood against the background of some core features of traditional epistemological thinking.

Features of Traditional Epistemology

Within the philosophical tradition, epistemology has tended to present itself not as an empirical, but as a normative discipline, often motivated by a wish to answer sceptical challenges. Philosophical epistemologies explore the grounds and validity of knowledge. While the question of *how we go about* acquiring and maintaining knowledge has countless aspects that call for empirical investigations, epistemology as traditionally understood attempts to tell us *what counts as* acquiring or having knowledge.

Accordingly, questions of the validity of epistemic claims (e.g. evaluating something as epistemically justified, attributing knowledge to somebody) are often contrasted with questions of their formation or genesis, and only the former are treated as epistemologically relevant. This view is often presented by recourse to the distinction between the 'context of discovery' and the 'context of justification'. The resulting picture with respect to the emotions is familiar enough. Research, actual processes of discovering and justifying, may well be driven by all sorts of emotions, such as curiosity or fear of dropping out of an academic career, but these emotions do not play any part in evaluating whether the results of research add to our knowledge. Emotions are important in the context of discovery as they influence the way researchers actually proceed. Nevertheless, they are irrelevant to the context of justification since the validity of the results is independent of such emotions. We will discuss this stance in more detail in section 4.

A considerable part of traditional epistemological theorizing includes a further assumption that contributes to a situation in which emotions were not perceived as epistemologically relevant. Clearly present in Descartes' *Meditations* (Descartes [1641]) and prevalent in traditional foundationalist epistemological projects, certainty or infallibility have been conceived as requirements of knowledge. This prioritizes deductive over inductive inferences and it leads to quests for infallible epistemic foundations and algorithms to choose between competing theories. On the face of it, emotions do not make promising candidates for such processes, since their cognitive output seems particularly fallible. The feeling of jealousy, for example, may occasionally help to discover facts that would otherwise go unnoticed; thus it may help acquire knowledge. But all too often it results in nothing but ill-founded suspicion (see Goldie, this vol.).

Apart from attempts to formulate and answer sceptical challenges or epistemic regress worries (e.g. by recourse to foundationalism), contemporary analytic epistemology has long been preoccupied with analysing key epistemic concepts such as (epistemic) justification and, above all, propositional knowledge. Analysis in terms of justified true belief has served as a promising starting point. Again, this model is unfavourable to emotions as long as the justification condition remains tied to inferential relations between beliefs. To say that subject *S* knows that *p* is to say that *S* truly believes that *p* and that this belief is epistemically justified. Whether *p* is justified depends primarily on its inferential relations to *S*'s beliefs.

Developments Within Epistemology

A range of more or less recent developments have shaped epistemology in favour of emotions. In this respect, debates about the justification condition of knowledge have been especially influential. According to an important proposal, what is needed for subject *S* to know that *p* is not that *S* be in a position to give reasons for *p*, but that *S* be in a position to rule out relevant alternatives to *p* (Dretske 1970; Goldman 1976). Attempts to determine what counts as epistemically relevant in any given situation soon opened the door to considerations previously thought of as alien to epistemology. Some difference between real, felt doubts and idle philosophical paper doubts, for example, could be used to establish that blind tasting Barolo is a relevant alternative to blind tasting Chianti, whereas being a brain in a vat is not. As we will describe in section 3, epistemic relevance and salience have become some of the most discussed functions of emotions within epistemology (see de Sousa 1987; Elgin 1996; Hookway 2003b).

In further attempts to remove the justification condition from the epistemic subject's cognitive control, it was argued that for a true proposition *p* to count as knowledge, it matters not so much whether the subject is in a position to give reasons for *p*. Instead, we better ask whether *p* was produced in an appropriate way, for instance by suitable causal chains (Goldman 1967), by reliable belief-forming mechanisms (Goldman 1976) or by properly functioning cognitive equipment (Plantinga 1986). To some extent, such moves towards 'externalist' theories of epistemic justification have blurred the traditional distinction between context of discovery and context of justification. After all, causal chains and the kinds of mechanism that produce or fix beliefs are features of the formation of knowledge.

But even though emotions are often part of processes of knowledge production, they did not immediately attract the externalists' attention. It was largely assumed that the function of the justification condition, whether spelt out in internalist or externalist terms, is to rule out beliefs that are merely accidentally true. Consequently, only those features of belief-forming processes that systematically contribute to the truth of their products were seen as normatively, and hence epistemologically, relevant. Insofar as emotions seem particularly fallible, they do not seem epistemologically relevant.

The case for emotions is strengthened once principled questions are raised with respect to counterexamples to various analyses of the concept of knowledge. As Hilary Kornblith has recently insisted, at the end of the day we are not interested

in our concepts of knowledge and epistemic justification, but in knowledge and justification themselves (Kornblith 2006, 12). Mark Kaplan (1985, 354) makes a similar point when he argues that unless it concerns the 'proper conduct of inquiry', analysis of the concept of knowledge is idle. The underlying view of epistemology is that it should primarily aim at 'understanding and advancing rational inquiry' (Kaplan 1985, 362).

Hookway (1990) suggests a similar shift away from characteristics of static belief systems to epistemic activities. This move is motivated by a pragmatist interpretation of sceptical challenges. Hookway thinks that sceptical challenges undermine the idea that we can simultaneously understand ourselves as participating in normatively regulated inquiries and as autonomous, responsible agents (Hookway 1990, 215). For him, this move towards practices and processes goes together with a widening of the epistemological focus from propositional knowledge to epistemic evaluations in general, as has also been argued for by Goodman and Elgin (1988; Elgin 1996).

Various considerations and claims have been put forward to defend such a development. Knowledge may just be ill chosen as the goal of epistemic activities. It may be too hard to achieve (especially if the tripartite analysis is correct) or it may impose inapplicable standards to our inquiries. Furthermore, knowledge, as it is typically discussed in epistemology, is restricted to propositions. But propositional knowledge may rest upon non-propositional elements, such as categories, concepts and methods. Or it may even be better approached in terms of knowing how to do certain things, such as conducting inquiries or revising one's beliefs.

Elgin (1996) argues that inquiry is better seen in terms of striving for understanding than in terms of knowledge acquisition. If analysis of epistemic processes is not restricted to their propositional results, but includes non-propositional components of understanding, then values, rules, categories and methods may be epistemically evaluated along with judgements or assertions (Elgin 1996, 122). Related considerations have led to a weakening of the truth-requirement in epistemic evaluations to acceptability (Goodman and Elgin 1988), tenability (Elgin 1996) or 'enough' truth (Elgin 2004) to make room for approximations and even fictions that contribute to understanding (Elgin 1996, esp. 122–7; see the discussion in Wild, this vol.).

Such moves towards epistemic processes and activities highlight aspects and properties of epistemic agents that have previously been neglected. Emotional states can be seen as part of this new, broader picture, which is also influenced by philosophers of science such as Kuhn or Feyerabend, who argue for broadly construed scientific rationality to replace the logical empiricists' ideal of logical procedures in matters of theory choice. One way of spelling out such a notion of rationality appeals to scientists' decision making. Harold Brown, for example, models the rational scientist on Aristotle's man of practical wisdom, who is

> … a model of the maker of crucial scientific decisions which cannot be made by appeal to an algorithm, and I offer the making of these decisions as a model of rational thought. It is the trained scientist who must make these decisions, and it is the scientists, not the rules they wield, that provide the locus of scientific rationality. (Brown 1977, 149)

Ernest Sosa (1980; 1985), James Montmarquet (1993) and Linda Zagzebski (1996; see also DePaul and Zagzebski 2003) undercut the distinction between internalist and externalist theories of epistemic justification by adopting the notion of virtue from ethical theory and focusing on epistemic or intellectual virtues.[2] This amounts to a reversal of direction of epistemological analysis. In the traditional order, epistemic evaluations of propositions, sentences or mental states were analysed first, and epistemic agents, acts and processes were then accounted for in terms of these analyses. Virtue epistemologists, however, start with normative properties of epistemic agents. Emotions come to play parts within such a strategy by contributing to the analysis of the epistemically relevant virtues or character traits (cf. the critique of Wild, this vol.).

Lorraine Code's (1984; 1987) and Alvin Goldman's (1986; 1999) social epistemology further widens the scope from individual epistemic agents to processes within epistemic communities, while feminist epistemologists and philosophers of science examine whether, and in what ways, the gender of epistemic agents may be epistemically significant. The epistemic agent's emotional involvement is one aspect of the gender differences that are discussed in this context (Jaggar 1989; Diamond 1991).

Two more developments inside and outside philosophy should be mentioned as having led to a surge of interest in emotions. In philosophy, Quine (1969) and others have initiated the project of naturalizing epistemology by assimilating it to psychology and cognitive science. Similar considerations fuelled hopes for a naturalized theory of mental states and concepts, which have contributed to the tendency to assign philosophical significance to empirical research on belief-forming processes and belief-revision (see Fodor 1984; Fodor 1987; Nisbett and Ross 1980; Goldie, this vol.). At roughly the same time, philosophical theories of emotions that emphasize their cognitive significance have been revived (e.g. Kenny [1963]), while the cognitive revolution in psychology started to give emotions a central place (see Lazarus 1999). Some twenty years later, the time was ripe for fusing these trends. Ronald de Sousa (1987) and Damasio ([1994]) combined cognitivist and naturalist aspects with great effect.

In summary, we can identify the following developments within philosophical epistemology that invite epistemological discussions of emotions: relevant alternative accounts and externalism about epistemic justification; calls for a theory of epistemic agents and practices, paradigmatically as opposed to conceptual analysis; opening the focus from propositional knowledge to epistemic evaluations in general; criticism of narrowly construed epistemic rationality within the philosophy of science; the recent prominence of virtue epistemology, discussions of social and feminist epistemology; and finally, the rise of cognitive science and naturalized epistemology.

2 Hookway (this vol.) appeals to emotions to combine key internalist and externalist commitments; Alessandra Tanesini (this vol.) introduces the concept of intellectual modesty to account for fallibilism about knowledge.

2. The Landscape of Emotions

Any exploration of possible epistemic functions of emotions presupposes some understanding of the variety and nature of emotion phenomena. In this section, we highlight a few general points that help to structure discussions of emotions and to avoid some sources of misunderstanding.

The Variety of Emotions

The first thing worth noting about the concept of emotion is the number and variety of emotion terms. Empirical research suggests that about 300 colloquial terms referring to emotions can be found in the English language (Plutchik 2003, 64–8). These range from 'anger' and 'anxiety' to 'indifference' and 'interest' as well as to 'self-respect', 'shame' and 'surprise'. Also, philosophers have compiled various systematically organized 'dictionaries' of emotions. Descartes' list in *The Passions of the Soul* (Descartes [1649], §§53–67) is a prominent example. More recently, Robert Solomon presented an *Emotional register – Who's who among the passions* with descriptions covering about fifty emotion terms (Solomon 1993a, ch. 8). Any such list faces the problem of what terms exactly deserve to be included, and the apparent diversity of phenomena classified as emotions raises the question of whether all emotions may be claimed to have epistemic relevance or only those of a certain type.

Attempts at structuring the universe of emotions include taxonomies along dimensions such as intensity (Plutchik 1980, 157–60), backward-looking and forward-looking, positive and negative (Lyons 1980, 89–91; Gordon 1987, 25–32; Prinz 2004b, ch. 7), as well as outer and inner direction (Solomon 1993a, ch. 8). Other ways of organizing the realm of emotions draw on designating certain emotions as 'basic', though this is an ambiguous attribute of emotions (see Ortony and Turner 1990; Plutchik 2003, ch. 4). One influential idea is to take some emotions as basic in the sense of 'elementary' and explain the others as derivations, mixtures or compounds thereof, in analogy to primary and secondary colours or chemical elements and compounds. There is a long history of attempts at such a reduction including some well-known philosophical proposals such as Descartes' list of *passions primitives*: wonder, hatred, joy, desire, love, sadness (Descartes [1649], §69). In cognitive science, Robert Plutchik's account (Plutchik 1980) is a paradigm of such a position. Alternatives to elementary emotions include the view that the emotions form a multidimensional spectrum which is structured by components that are not emotions themselves (Ortony and Turner 1990). Equally common is the use of 'basic' in the sense of 'pan-cultural'. Empirical investigations in psychology suggest that some emotions can be found in all cultures and have expressions that can be cross-culturally recognized (e.g. Ekman 1999b). This has been treated as evidence for their being relatively basic from a biological, specifically evolutionary perspective. In particular, these findings have been put forward as a challenge to the rival stance that emotions are social constructions (cf. Prinz 2004c). A classical list

of pan-cultural emotions are the 'big six': anger, fear, happiness, sadness, disgust, surprise (Ekman et al. 1969).[3]

Notable Distinctions

One reason why 'emotion' covers such a remarkable diversity of phenomena is that this term itself is used with a range of different meanings. In modern philosophical terminology, 'emotion', the older terms 'passion' and 'affect', as well as related adjectives are used in a great variety of ways, sometimes with contrasting meanings, sometimes as synonyms.[4] The same holds for everyday language, which additionally tends to use 'feeling' interchangeably with 'emotion'. In theoretical writings, there is a discernible tendency to distinguish between emotions, feelings and moods. Furthermore, 'affective' tends to be used in a broad sense, including but not confined to emotions and feelings, but covering, for instance, moods as well (cf. Davidson et al. 2003, xiii; Griffiths 2004b, 240–43).[5] Nevertheless, these are trends, not rules. One always has to be prepared to find divergent uses of 'emotion', 'affective' and the like, as well as distinctions drawn differently from what is suggested here.

Non-English usage of 'emotion' is an additional source of confusion because superficially similar terminology may cover up differences in meaning (see Cassin 2004). In German, for example, 'Emotion' has fairly recently been adopted from English and French and has started to replace more traditional terms such as 'Affekt' or 'Gemütsbewegung'. Often, but by no means always, it is used as a synonym for 'Gefühl', which in turn is not only the standard translation for 'feeling' but also used to cover emotions.

Two distinctions are particularly useful in any discussion of the epistemic relevance of emotions, since they help to avoid some misunderstandings and confusions that are caused by the variety of terminologies in use.[6] First, if an emotion is ascribed to somebody, what does the emotion term refer to? This question calls for distinguishing dispositions, processes, episodes and states. The second distinction concerns contrasting uses of 'emotion', 'feeling' and 'mood', where feelings can be seen as an aspect of emotions among others.

To begin with, there are *dispositional* and *non-dispositional* uses of emotion terms (Lyons 1980, 53–6). When we say

(1) Toby has a fear of flying.

3 In the wake of this research, 'basic' has sometimes been used meaning 'being a product of evolution'. Paul Ekman (1994) maintains that all emotions are basic in this sense. Furthermore, he now claims that it is actually *families* of emotions that are basic (Ekman 1999a).

4 Indices of recently published philosophy handbooks suggest that 'emotion' is in the process of replacing 'passion' and especially 'affect' (but not 'affective'). In psychology, 'emotion' and 'affect' are both widely used, though with various meanings as well (cf. Schwarz and Clore 2007, 385–6; Plutchik 2003, 62–3).

5 Hookway (this vol.), for example, uses 'affective' in this broad sense.

6 Ryle (1949, ch. 4) discusses many such sources of potential misunderstandings.

we may want to say that it is a characteristic of Toby that he fears flying; that is, we ascribe to him a disposition *to* an emotion. If this is the case '*S* has *e*' can be interpreted as a shorthand for '*S* is disposed to have *e*' or '*S* is the *e* type of person'.[7] Such a reading of (1) is appropriate in a context like:

(2) Don't even ask him to join you on this trip. He has a fear of flying.

Alternatively, we may use (1) to talk about an emotion actually affecting Toby (sometimes called an 'occurrent' emotion):

(3) The trip to Hawaii was a nightmare for Toby because, suddenly, he had a fear of flying.

In this case, '*S* has *e*' means the same as '*S* is affected by *e*'. The difference between these two usages makes it possible to say something like

(4) Toby's fear of flying has saved him from actually experiencing his fear of flying.

without contradicting oneself.

Additionally, this latter use of emotion terms can be further differentiated. Expressions that refer to an emotion affecting somebody can in fact refer to a great variety of emotional or emotion related phenomena. There are two relevant dimensions here. On the one hand, emotions have the character of a process. They develop over time, showing a pattern of changing features (Frijda 1993, 382; Goldie 2000, 12–14). This renders emotion terms applicable to anything from long term processes lasting for hours or months to episodes of short duration, in the limiting case even states with almost no discernible pattern of evolution (cf. Solomon 2003, 2). For example:

(5) The trip to Hawaii was a nightmare for Toby, for he started having his fear of flying the very day I suggested the trip to him.

(6) Everything went well, until Toby suddenly had another rush of his fear of flying in the middle of our flight.

On the other hand, having an emotion can include a great many elements or aspects such as feelings, behaviour, bodily conditions and dispositions, including dispositions to certain emotions:

(7) Irritated by his fear of flying, Toby was always on the verge of getting angry with the cabin crew.

---

7 For further differentiation between dispositions and character traits see Goldie (2000, ch. 6).

Many expressions can be used for referring to an emotion as a whole as well as for picking out one, or a few, of all aspects of an emotion. Often the aspect referred to is a feeling:

(8) As soon as we started, Toby's fear of flying got even more intense.

Behind this distinction between emotions and feelings lurk important problems concerning the nature of emotions. For there are theories of emotions which claim that emotions essentially *are* feelings, while others reject such an identification or insist on them being conceptually or factually independent. 'Feeling' and the verb 'to feel' are themselves used in a wide variety of ways in everyday language. Sometimes they are obviously closely related to emotions, as in 'I feel angry at him'. For other uses, the relations to emotions are less obvious, as in 'I feel like having a bath', 'I feel hungry', 'I cannot feel the vibrations you are talking about', 'I feel you should not interfere' and so on (cf. Alston 1967, 483; Kenny [1963], 36–7). In philosophy, the majority of writers use 'feeling' to refer to some quality of consciousness; that is, to some state of awareness, to be described, if possible, in phenomenological terms, similar to the qualia of perception.[8] On-going disputes concern the question of what it is that is sensed in a feeling, the relation of feelings to bodily conditions and to behaviour, as well as the questions of whether feelings are accessible by introspection alone and whether they are intentional (cf. Goldie 2000, ch. 3).

Finally, there is 'mood', a third term which is used in close connection with emotions:

(9) Ever since Toby flew to Hawaii, he has been a bit down.

A widely accepted psychological definition holds that moods are diffuse, global, low intensity emotions of longer duration (Oatley et al. 2006, 30). Against this it has been quite convincingly argued that the most salient difference between emotions and moods is not their duration or intensity, but the fact that moods do not have a specific intentional object (e.g. Frijda 1993; Goldie 2000, 143–51; Prinz 2004b, 182–8; for sceptical remarks see Plutchik 2003, 63). Emotions, so the argument goes, can be described as intentional affective states, oriented towards rather specific objects. The target of an outburst of anger, for instance, may be a certain word one believes to have overheard. Moods, on the other hand, either lack intentionality altogether or they are non-specifically oriented towards, for example, 'everything', 'nothing' or 'the world'. In short, moods are emotions without specific objects, the difference being gradual rather than categorical (Goldie 2000, 17).

To sum up, as emotion terminology is anything but uniform, statements about emotions are often ambiguous. They admit of disposition-to- and affected-by-readings, or of being interpreted as referring to emotional processes of different 'sizes' or to some aspect of an emotion, such as a feeling or a bodily condition.

8 In his contribution to this volume (p. 186), de Sousa draws another contrast between 'emotion' and 'feeling'. 'Emotion' is reserved for phenomena on a personal level, whereas 'feeling' includes subpersonal phenomena as well.

A rather restrictive interpretation of 'cognitive' requires that theories of emotions count as cognitive only if they attribute propositional content to emotions. On such accounts, emotions are typically assimilated to beliefs or normative judgements. Claims along these lines can be found in the theories of, for example, Solomon (1993a, ch. 5.3; but cf. Solomon 2003 and 2007, ch. 18), William Lyons (1980, 71–7) and Martha Nussbaum (2001, ch. I.1).

More liberal interpretations of 'cognitive' additionally include theories that construe emotions on the model of perception. Such an analysis of emotions has already been present in de Sousa's influential study *The Rationality of Emotion* (1987; cf. de Sousa 2004). It is currently defended by several authors, Sabine Döring (this vol.), Elgin (1996; this vol.) and Prinz (2004b).[10] As a minimal definition of cognitive theories of emotions, one may count all theories as cognitive which include the claim that emotions are intentional (Döring 2003, 225–6; this vol.). According to this criterion, strictly behaviouristic theories and theories that reduce emotions to feelings are not cognitive.

There are two ways in which an emotion can be said to be intentional or oriented towards an object (Kenny [1963], 131–5). On the one hand, an emotion has a 'formal' object such as *being dangerous* in the case of fear[11] or *being disgusting* in the case of disgust. In de Sousa's characterization, the formal object of an attitude is 'that which gives the trivial answer to the question *Why do you hold this attitude?*' (de Sousa 2007,327; cf. de Sousa 1987, 121–3). More specifically, the formal object of an emotion can be defined as the property *x* must have, or the norm *x* must comply with, if the emotion in question is to be appropriate or at least intelligible with respect to *x*. On the other hand, a specific emotional episode is oriented to something particular, the 'material' or 'particular' object, which at least seemingly fits the formal object. The formal object of surprise, for example, is *being unexpected*, while the particular object of your surprise may be a long lost key found in one of your shoes. Intentionality in this sense of formal and particular objects is not just intentionality in the sense that emotions relate to some bodily conditions (Goldie 2004, 93). To say that disgust presents the body as being in a state of disgust is to miss the point of this emotion, which rather consists in presenting some object or situation as disgusting. For this reason, accounts which simply identify emotions with awareness of bodily changes do not count as intentional.[12]

As an alternative to a general cognitive theory of emotions, one may defend the more restricted view that certain emotions are specifically cognitive in one of the senses explained. This issue will be discussed in the final section.

---

10 The terminological muddle is illustrated by the fact that Prinz (2004b, ch. 2) defends his 'perceptual' theory as non-cognitive. He relies on his explication of cognition as organismic control: 'I propose that we call a state cognitive just in case it includes representations that are under the control of structures in executive systems, which, in mammals, are found in the prefrontal cortex' (Prinz 2004b, 47).

11 Those who argue that all emotions have response-dependent properties as their formal objects associate fear with *being frightening* (cf. Salmela 2006, 386).

12 This is often raised as an objection to accounts along the lines of James and Lange. (See Prinz 2004a, 54–6 for a brief rejoinder.)

A second debate revolves around the charge that philosophical theories of emotions rely on the invalid assumption that it is possible to develop a uniform theory of *the* emotions. Several writers, notably Amélie Rorty and Paul Griffiths, have claimed that the emotions do not constitute a natural kind (Rorty 1980b; Rorty 2004; Griffiths 1997; Griffiths 2004a; Griffiths 2004b). Rorty argues that there is no clear distinction between emotions and other mental states such as motives, moods and attitudes. Consequently, she emphasizes that philosophical accounts of emotions should be integrated into a comprehensive framework of a philosophy of mind (see also Solomon 2004a, 84). Griffiths claims that the diversity of phenomena called 'emotions' does not allow for a unified scientific account; that is, the category of emotions cannot be used to reliably derive the inductive generalizations that biology, neuroscience and psychology need for explaining the mechanisms underlying emotions. Even what appears to name a single type of emotion, for example 'anger', may collect diverse phenomena, ranging from an instinct-like 'affect program' to a voluntarily adopted strategic behaviour.[13]

Many philosophers have resisted this analysis, objecting that the emotions do have a lot in common (even if not with respect to biological mechanisms), that the taxonomies of the vernacular should be respected, and that Griffiths's arguments rely on too narrow a conception of what theories of emotion should aim at explaining (see e.g. Prinz 2004b, ch. 4; Roberts 2003, ch. 1.4). At any rate, even if philosophical analysis starts with vernacular concepts, it would be a misunderstanding to conclude that it is thereby confined to simply accepting these concepts. Philosophy, no less than the sciences, relies on the method of explication, which aims less at finding extensional equivalents than at replacing vernacular concepts for the sake of precision, simplicity and fruitful theories (cf. Carnap 1962, §§2–3). For the question about epistemological significance of emotions, we may draw the consequence that it could be a serious strategic error to presuppose that all emotions can be treated the same in this respect. Rather, one should be prepared to find that certain emotions fulfil some epistemic functions – perhaps only under certain circumstances – whereas other emotions are unsuitable for these functions, or are altogether irrelevant from an epistemological point of view.

3. Emotions in Epistemology

Emotion's Bad Reputation in Epistemology

Most epistemologists have not given positive accounts of emotions. This is part of an attitude that holds that 'reason should be the master of passion'.[14] Its roots can be traced back to ancient Greece, where Democritus, for example, stated: 'Medicine cures diseases of the body, wisdom frees the soul from emotions' (Diels 1951–52, 68 B 31; transl. in Sorabji 2000, 2). Such maxims have served as a guide not only for

13 Griffiths (e.g. 2004b, 234) argues that his view that the emotions do not form a natural kind does not imply that they do not fall under an univocal concept.

14 Cf. '[Our passions] are *Good Servants,* but *Bad Masters,* and Subminister to the Best, and Worst of Purposes, at once' (L'Estrange [1699], 38).

practical decision making, but also for cognitive activities. The distinction on which they depend remains in place when they are turned upside down, as most famously in Hume's declaration that reason is and ought to be the slave of the passions (Hume [1739/40], II.iii.3, 415).

In any case, the metaphor of master and slave is more ambiguous than one may first think. If reason is the master then the emotions are servants. Servants are generally kept because they do something useful, they sometimes have abilities their masters lack, and many a master would be rather lost without them. The metaphor of master and servant, together with the metaphorical mind vs. heart categorization of practical decisions, is embedded in a pre-theoretic cultural tradition of treating reasons and emotions as opposing one another.[15] In the philosophical tradition, the contrast between reason and emotion is closely related to various doctrines about the different faculties of the soul. However, irrespective of their position in various theories of the mind, and notwithstanding Hume's famous dictum, the reputation of emotions in epistemology tended to be unfavourable throughout the history of philosophy. There are a number of reasons for this traditionally prevalent negative assessment of emotions.

To start with, emotions have long been recognized as threats to rational and epistemic decision making. They can impair processes of knowledge acquisition or the assessment of knowledge claims. The most straightforward version of such a view, usually attributed to the Stoics, holds that emotions simply are misguided judgements (see Sorabji 2000, 55). Independently from such an identification, emotions have been charged with distorting perception, as well as leading to wishful thinking and self-deception. Explanations of such phenomena often rely on tying emotions to the will or to desires. Emotions are then criticized for being a means by which will or desire can 'take over' reason or perception, or disrupt a rational process. To guard oneself against such influences, emotions either have to be mastered (as recommended, for example, by the Stoics) or one has to strive for having the 'right' emotions (as Aristotle argued). In this volume (p. 159), Peter Goldie takes a closer look at how emotions can 'skew the epistemic landscape', especially when triggered in environments that are different from those they have evolved in (see also Wild, this vol.).

A second concern is that emotions do not contribute to knowledge because they are too subjective or private to be relevant to what should ultimately be the objective truth of beliefs, independently of how exactly 'objective' and 'truth' are understood. If, for example, two people spot an animal and one of them believes it to be a wolf while the other sees a dog, there is a fact that decides who is right. However, if somebody experiences fear when coming across a dog, then this emotion indicates the presence of something frightening, but this is so *only for that person*. It neither follows that the animal should be experienced similarly by other people, nor that they should consider it to be frightening.

15 The issue has been related to western culture's more general dualism of mind–body, culture–nature and the like (e.g. White 1993). However, similar oppositions can also be found in Asian thinking (see Marks and Ames 1995).

Thirdly, even if it is commonly conceded that there are emotions which are obviously linked to cognition by motivating and regulating cognitive activities, this motivational force is often dismissed as epistemologically irrelevant. Wonder, which comes first on Descartes' list of basic emotions (Descartes [1649], §69), is a case in point; it may motivate us to acquire a belief, but does not enter into epistemic evaluation of the belief. Typically, such arguments hinge on a distinction between the validity of beliefs and theories, and the history of their formation, which is thought to be epistemologically irrelevant. This issue will be discussed in section 4.

Similar considerations are brought to bear on those emotions which involve an evaluation of a propositional content with respect to some cognitive standard. A feeling that something is the case and similar emotions are not sources of knowledge, because their affective attitudinal aspect as such does not contribute to the justification of the embedded belief. Feeling that something is the case fares no better than acts of guessing or instances of clairvoyance (see Dohrn, this vol.).

The Trend Towards a Rehabilitation

Several points have been instrumental in the recent reassessment of emotions as candidates for epistemic functions. First of all, emotions can be and often are evaluated as rational or appropriate. There is currently a discussion on whether the appropriateness of emotions may be interpreted as emotional *truth* (see e.g. de Sousa 2002; Salmela 2006). Either way, this undermines the view that they necessarily misguide or distort cognition. Rather, one would expect that they only do so if they are irrational or inappropriate. Instead of dismissing emotions as intrinsically interfering with knowledge acquisition, we should specify the conditions under which they contribute to knowledge. Secondly, there are reasons to doubt the claim that the privacy and alleged subjectivity of emotions render them epistemically useless (Goldie 2004, 94–5; Solomon 2007, 150–58). In this respect the analogy between emotion and perception plays an important role (see de Sousa 1987, 145–58; Deonna 2006).

Finally, the renewed interest in the emotions has led to a whole range of philosophical theories of the emotions. Most of these theories hold that emotions include a cognitive element, which means that they are directed towards the world and can be evaluated accordingly. Such cognitive functions call for epistemological analysis, which in turn may lead to an epistemological reassessment of emotions.

Despite the recent trend to give emotions a more favourable place in epistemology, theories vary enormously with respect to the actual functions they assign to emotions, with respect to the kinds of emotions they give such functions to and with respect to the epistemological consequences they draw. Some conceive of themselves as still compatible with traditional positions in epistemology, invoking emotions, for example, to underwrite relevant alternative accounts of reliability. Others depart from traditional conceptions of epistemology and opt for emotions' cognitive significance in the context of a more or less drastic redesign of epistemology (Elgin, this vol.). In what follows, we look at the most frequently mentioned candidates for epistemologically relevant features and functions of emotions. These are motivational

force, salience and relevance, epistemic access to facts and beliefs, non-propositional contributions to knowledge and understanding, and epistemic efficiency.

Motivational Force

That emotions motivate cognitive activities can hardly be doubted. There is an abundance of anecdotal evidence of researchers describing themselves as motivated by emotions when they tell their stories outside the academic journals (cf. the case-study in Thagard 2002). Examples of motivating emotions include surprise, interest, doubt and puzzlement sparking inquiry, pride in standards of research, frustration and disappointment with the results achieved.[16] It has been argued that precisely emotions' disruptive character, so often treated as evidence for their supposed irrationality, makes them important, perhaps even indispensable for cognition. Emotions kick in when we are cognitively challenged, when our knowledge seems false, inadequate, irrelevant or not useful. Emotions are mechanisms that make us learn something (Oatley 1999, 274–5).[17]

Emotions can motivate not only further research within some accepted framework and according to shared and accepted standards, but also critical reflection on such frameworks or standards. In this way, the disappointing outcome of some experiment may not just motivate the scientist to repeat the experiment or to redesign it, but also to doubt the reliability of his instruments or to envisage theoretical revisions. In some cases, researchers may also be led to question the standards determining what counts as, for example, an established result or a reliable replication. Hookway argues that doubt, especially if assimilated to anxiety, can be counted as an emotion that motivates critical reflection of the reliability of results and standards of epistemic evaluation (Hookway 1998; Hookway 2000; Hookway, this vol.; Dohrn, this vol.).

While there is no doubt that inquirers sometimes are motivated by emotions, the question is whether the inquirer's motivation is epistemologically relevant. Considerations relying on a distinction between contexts of discovery and justification suggest a negative answer. However, one way of arguing to the contrary hinges on a shift of epistemological attention from the pursued result to the pursuit itself; that is, from knowledge and/or true belief to epistemic activities and cognitive agents. As mentioned in section 1, both Elgin (1996) and Hookway (1990; 2000) defend such a move. Consequently, justification of beliefs may be conceived as dependent on the history of their acquisition (Elgin 1996, 121–2). And since beliefs have to be evaluated in relation to cognitive actions, their evaluation may also depend on the identity and properties of desires, goals, mechanisms, motivations and virtues. Because emotions

16 The motivational function of emotions must not be conflated with the view that there is some kind of basic affective orientation in life, which motivates or orients scientific as well as philosophical inquiry. This view, put forward by Heidegger, turns not on emotions but on dispositions to feelings or moods; in Heidegger's terminology, 'Gestimmtheit' ('mood', literally 'being tuned') or 'Befindlichkeit' ('state of mind', better 'affectedness').

17 A different point is, that emotions have a huge impact on how effectively we learn and on the ways in which we can later use what we have learned. Neurological studies have turned out ample evidence for this long standing didactic truism (cf. LeDoux 1998).

with motivating force can themselves be normatively assessed in many ways ('How unreasonable to be disappointed and continue to do all these nightly experiments.'), appropriate motivating emotions could become available for constitutive accounts of justified beliefs. Including motivating emotions in justifications of beliefs is not far from, and is sometimes seen as related to, the account of justified beliefs that some virtue epistemologists tend to give (cf. Fairweather 2001).

Salience and Relevance

Emotions' potential function as a source of salience and relevance has been emphasized by de Sousa in his influential thesis that 'emotions are determinate patterns of salience among objects of attention, lines of inquiry, and inferential strategies' (de Sousa [1979], 137; de Sousa, this vol.; Lance and Tanesini 2004; Hookway, this vol.). This thesis has been applied to a well-known problem in decision theory (de Sousa 1987, 190–96).[18] In the standard model of rational choice, an ideal agent faces a combinatorial explosion because for each of the unlimited number of possible actions open to her at a given time there is an unlimited number of consequences to be taken into account. Real agents with limited resources to spend on a decision must therefore limit the number of actions and consequences to be considered if they ever want to reach a decision. The standard methods of rational choice cannot be used for accomplishing this reduction, because they would only reintroduce the combinatorial problem for every consequence of every action when the agent must decide on whether she should include this action in her reasoning about the decision she set out to take in the first place. According to de Sousa's argument, emotions, by functioning as sources of salience, effect the necessary narrowing down of the number of actions and consequences. Guilt, for example, may be an emotion that helps selecting strategies of social interaction by drawing attention to possible dangers of non-cooperative behaviour (Ketelaar and Todd 2001, 200–203).

As a first approximation, for emotions to be sources of salience means that they establish a focus on certain aspects of a situation, they act as 'spotlights' (Peters 2006, 458). However, as Elgin has emphasized, emotions establish salience in highly complex ways that are not limited to simply putting some properties of a situation into the foreground. An emotion is 'a frame of mind or pattern of attention that synchronizes feelings, attitudes, actions, and circumstances' (Elgin 1996, 148). Like beliefs, emotions cannot be reduced to an attitude towards a proposition or situation, but comprise attitudes to other situations, commitments to categories being appropriate for classifying aspects of the actual and alternative situations, acceptance of standards for the evaluation of and dispositions to act, believe or feel in such situations (Elgin 1996, 153; Elgin, this vol.). An unnerved neighbour, for example, may perceive a child's crying as nothing but loud and piercing, whereas the child's dismayed parents hear signs of some specific kind of distress, say pain, drawing their attention to possible causes and ways of bringing relief (see Elgin 1996, 153–4).

18 The issue is discussed under labels such as 'the (philosopher's) frame problem' or 'the search problem' and it is debated how the various ways of spelling out the problem are related to each other (cf. Evans 2002).

Just as with motivation, salience and relevance come into focus once we see epistemology as primarily dealing with cognitive activities, as opposed to properties of belief systems. As sources of salience and relevance, emotions are themselves evaluations. To consider something to be relevant or salient with respect to some goal and context of inquiry is to evaluate it; namely as something that ought to be considered further. Moreover, emotions fulfilling this function can be evaluated, for we can wrongly find something salient or relevant.

Additionally, salience can (but need not) be seen as an instance of epistemic immediacy. We often struggle to find out why we find something salient, and whether and why we are right to do so. Sometimes, at least, this happens because our epistemic evaluation of salience is either not governed by rules or governed by rules that we cannot articulate (Hookway, this vol.; see the remarks on Goodman's grue paradox p. 2).

Finally, relevance has been taken seriously by epistemologists who support relevant alternative accounts of knowledge. According to such theories, we would deny that a subject knows that she is looking at, say, a robin if she could not visually distinguish this situation from an alternative in which she is looking at some other bird (e.g. Dretske 1970; Goldman 1976; Goldman 1986). Some possible situations however, including sceptically threatening brains-in-vats scenarios, need not be ruled out – they are irrelevant to assessing the knowledge claim in question (e.g. Goldman 1976, 775). The strategy thus requires a distinction between relevant and irrelevant alternative scenarios, which may be established by emotions, for example by some difference between real, felt doubts and mere philosophical paper doubts. We may just need confidence in our ability to focus on relevant alternatives (see Hookway 2003b, 190–91).

Epistemic Access to Facts and Beliefs

A third way of attributing epistemological significance to emotions is based on the claim that emotions are an additional source of knowledge (alongside reason, perception, intuition, testimony). This thought can be fleshed out in two ways, although they are not always easy to keep apart in the literature. The strong version holds that emotions provide epistemic access to otherwise inaccessible facts. Weaker varieties claim that while emotions can be sources of true beliefs, they are not indispensable.

As Elgin observes (1996, 164–5; this vol.), emotions provide epistemic access to certain response-dependent properties that are directly tied to emotions. The properties of being amusing, depressing or disgusting may serve as examples. Response-dependent properties are typically, but not exclusively, part of discussions of secondary qualities. Their defining feature is that whether a given object has the property in question must be decided with reference to our responses to that object. Some of the properties that lend themselves to analysis as response-dependent are dependent on differing emotional responses. There is a sense in which for something to *be* disgusting, for example, is for us (or at least some of us) to respond to it by finding it disgusting. It is highly plausible that, in the right circumstances, emotions can disclose such response-dependent properties.

Moreover, emotions are said to provide access to facts more generally (e.g. Goldie 2004, 94–9). Typically, emotions are not only reactions to stimuli, but are intertwined with beliefs that may relate to many aspects of a given situation. They are, generally speaking, 'sensitive to information' (Elgin 1996, 156). The result is that emotions provide complex patterns of attitudes, feelings, expectations and dispositions which correlate to complex nets of features of actual and possible situations. This is the basis for 'exploiting' emotions as cues for facts which are in some way or other related to the occurrence of the respective emotion.

While these points support the claim that emotions can provide access to certain facts, it is not obvious that they also support the strong claim that there are facts that are epistemically accessible only through emotions. As an example, we may look at response-dependent properties such as *amusing* and *disgusting*, which are plausible candidates for the strong claim. For the sake of the argument, we may put aside questions related to the criteria of correct application of such predicates. We can, for example, simply think of a new predicate 'minimally disgusting' defined as applying to anything at least one person finds disgusting. If Jacques, for example, finds immature cheese disgusting, then the proposition that immature cheese is minimally disgusting can be justified by appeal to his emotion, and indeed it must be justified by appeal to somebody's emotion. However, this does not amount to a proof of the strong claim. Although the fact that something is minimally disgusting is by definition constituted by somebody's emotional response, this does not imply that one cannot know about it without having certain emotions. Perhaps Jacques's disgust can be reliably read off his facial expressions, which reliably express his emotion.

A more convincing case for the strong claim may be made with reference to emotions that provide epistemic access to one's own propositional attitudes and commitments (perhaps even beliefs, if we allow for subjects' having beliefs without being aware of them). By 'behaving' emotionally the way one happens to do, one can make discoveries about the beliefs, expectations or standards one implicitly endorses. This may often be the only way to find out about a certain commitment. However, such discoveries presuppose that one knows about one's emotions. The fact that somebody's behaviour disappoints me can be my sole clue for discovering that I had certain expectations towards this person's behaviour, which in turn can imply that I am committed to certain standards of behaviour or to beliefs about the person or situation (cf. Elgin 1996, 159–61; Elster 1996, 1393–4).

The view that emotions are sources of beliefs is particularly attractive in the context of cognitive theories of emotions, which hold that emotions embed beliefs or other propositional contents. A paradigmatic example is Aristotle's analysis of anger as 'a desire accompanied by pain, for a conspicuous revenge for a conspicuous slight at the hands of men who have no call to slight oneself or one's friends' (Aristotle Rhetoric, 1378a31–33). On this account, being angry presupposes believing quite a few things: that a slight occurred, that it was unjustified and so on. Although belief-centred cognitive theories have proved difficult to defend as general theories of emotions, it is still possible to defend the view that some emotions are essentially tied to types of propositional content. The question is whether such emotionally accessed beliefs are not of just the same epistemological interest that beliefs in general are.

Emotionally 'embedded' propositions can be true or false, and believing them can be justified or unjustified, just like any other propositional content or belief.

Non-propositional Contributions to Knowledge and Understanding

As mentioned in section 1, there have been attempts to widen traditional epistemology's focus from propositional knowledge to a broader conception of understanding that includes non-propositional contributions, skills and methods. Emotions have been claimed to contribute significantly to the formation of categories and cognitive organization, as well as standards of inquiry and warrant.

The central aspect of cognitive organization is a system of categories or concepts used for classification. Any epistemological theory that deals with propositions and concepts will acknowledge that the content of propositional attitudes depends, among other things, on conceptual content. Hence the content of Anna's belief that the stone she bought is obsidian depends on the concept *obsidian*, as used by herself or by the person that attributes this belief to her. The involvement of categories in belief is also a crucial point in theoretical debates, such as the one between epistemological foundationalism and coherentism (e.g. BonJour 1985, ch. 2.3). Concepts and categories can themselves be evaluated. It is, for example, not difficult to think of stupid or fruitless ways of dividing the animal kingdom, as exemplified in Borges' story 'The Analytical Language of John Wilkins' (Borges [1942]). Consequently, one may claim that categories are themselves subject to justification in relation to some given epistemic end or general epistemic considerations. Elgin, for example, holds that for categories to be justified is for them to 'fit' into a constellation of 'tenable commitments to promote tenable ends'; that is, to be part of a consistent and systematic system of beliefs, commitments, standards and methods that is tethered in pre-theoretic commitments (Elgin 1996, 104–5).

Similarly to the view that salience ought to be understood in affective terms, it may be argued that the application of new categories, the abandonment of pointless categories, the re-activation of dormant categories and the revision of existing categories are best tied to emotions. There are two ways in which emotions may affect classifications (Elgin 1996, 161–9). Emotional responses constitutive of response-dependent properties (e.g., amusing, interesting, boring, disgusting) are examples of emotions that straightforwardly affect conceptual organization. In this case, emotions we accept as appropriate determine what is covered by certain concepts. In a second family of cases, appropriate responses are not criteria for the applicability of a given predicate, but criteria for the appropriateness of a category, as is obviously the case with certain moral categories. If two situations strike us in the same way as being, say, outrageous, this can count as evidence against a proposed system of classes and its standards if that would compel us to classify them as opposites from a moral point of view (e.g. the same behaviour is good or bad just because it is a man's or a woman's behaviour respectively); and similarly the other way around.

Hookway applies the idea that emotions can be invoked in order to justify preferences for some categories or concepts to Goodman's 'grue' paradox (Goodman 1983, 74). Quine suggests that we ought to prefer the green-hypothesis, because we somehow feel that this is simpler (Quine 1960, 19). The crucial claim is that such a

choice is epistemically immediate insofar as it would be mistaken to put epistemic subjects under the obligation to justify such a choice by, for example, articulating the rules they follow; that is, the principles of induction and projection they rely upon. If one accepts that simplicity is, at least sometimes, just felt, then such a feeling of simplicity also exemplifies how emotions may be said to embed epistemic standards that may not be applicable in other, rule-based ways. More precisely, emotions can reflect evaluations that rely on standards which are not directly accessible to reflection. Consequently, attempts to uncover such standards must rely on methods that explore our emotions for evaluative patterns, such as thought experiments. Hookway concludes: 'We can formulate our evaluative standards only as a result of a search for an explanation of our habits of evaluation' (Hookway 2002, 253).

At least with respect to certain standards, such as validity of basic patterns of inference, one may want to go one step further and claim that they must be reconstructed as principles that explain patterns of emotional reactions. Although this does not imply that emotions justify these standards, it amounts to claiming that some standards are accessible via emotions only (de Sousa 1998).

In line with Quine's naturalistic stance, such claims may get support from a more biologist point of view. Paul Thagard, for example, has presented a theory of coherence which includes emotional aspects in such a way that one can interpret it as the hypothesis that judgements about coherence are connected to emotional reactions on a neurological level (Thagard 2000, 211–13; Thagard 2002, 245–7; Thagard, this vol.). This boils down to the claim that there is a causal connection between coherence and some emotions, for conceptual and emotional coherence are associated on a biological level. If we accept coherence as a crucial standard for assessing theories, we can argue that such causal connections warrant the claim that aesthetic emotions sometimes are relevant to the evaluation of a theory. They can be interpreted as indicators of (in)coherence and, if conscious, as perceptions of (in)coherence.

Epistemic Efficiency

There are two ways of claiming epistemological significance for emotion's contributions to cognitive efficiency. The first and weaker claim is that emotions make it easier to perform things that could also be done in their absence. This claim is hardly controversial, but even if it is true that emotions are merely heuristic devices that are in principle dispensable, it is unclear whether we have to take emotions seriously when engaging in responsible epistemic evaluation (Hookway 2003a, 80).

According to the stronger and more controversial claim, there are important cognitive functions which humans cannot perform successfully at all without relying on the efficiency-enhancing quality of emotions. A candidate for the stronger claim is de Sousa's view that emotions make rational deliberation humanly possible by selecting relevant information (p. 17). Furthermore, emotions could also be seen as bringing sceptical challenges to a halt. They determine the point at which the demand for yet another justification can be rejected, thereby blocking the threatening regress. Such 'shallow reflection'-accounts (Hookway 2003a, 82) do not only claim that emotions enhance epistemic efficiency, but also that without the help of emotions

'excessive reflection' would block any inquiry right at the start (see Dohrn, this vol. for a critical discussion).

Research in the context of recent discussions of emotional intelligence and emotions' cognitive functions has aimed at identifying evidence for claims along these lines. Our ability to take rational decisions, in particular, seems to be severely limited without emotions (e.g. Damasio [1994]). These findings primarily relate to practical decisions, but it seems plausible that they also hold for theoretical decisions, insofar as they, too, involve decisions to act in certain ways. For example, the question of whether a certain experimental design can be considered valid is related to the decision whether the experiment can be carried out as designed or should be redesigned (see Goldie 2004, 98–9).

4. Two Focus Questions

As one would expect, the proposed positive contributions of emotions to epistemology are contested for various reasons. In the following, we concentrate on two issues. Firstly, the epistemological significance of emotions can perhaps be denied by invoking the distinction between the context of discovery and the context of justification. Secondly, we may suspect that epistemological relevance cannot be claimed for emotions in general, but only for a subset of specifically epistemic emotions, and that some of the mental states epistemologists have recently invoked as emotions are not really emotions. We think that these issues implicitly shape a great deal of the discussion on epistemology and emotions.

A Normative Perspective on Emotions

As pointed out in section 1, the field of epistemology has often been delineated with reference to a distinction between context of discovery and context of justification.[19] On this basis, one may argue that emotions are just one of many epistemologically irrelevant aspects of epistemic agents, practices, processes and states. They may be important factors in actual processes of belief acquisition and revision, partly determining what beliefs we arrive at and how much confidence we have in them, but they do not play any role when it comes to epistemically assessing beliefs or cognitive processes. Neither can we reasonably answer the question whether a belief counts as knowledge by appealing to emotions, nor will the answer to this question depend on whether emotions contributed to our getting to consider this belief as a candidate for knowledge. In short, epistemologists deal with the question of what counts as knowledge; it is not their job to find out, for example, about how efficient

19 Introduction of the terms 'context of discovery' and 'context of justification' is usually attributed to Reichenbach (1938; in German, Reichenbach introduced a distinction between 'process of discovery' and 'process of justification' in Reichenbach 1935). However, the distinction was common ground for the logical empiricists, and there are a range of historical precursors, sometimes under different labels, such as 'genesis' vs. 'validity' or 'quid facti' vs. 'quid juris' in Kant's *Critique of Pure Reason* (Kant [1781/87], A84/B116; see Hoyningen-Huene 1987, 502–3 for historical details).

various processes of belief formation are or about what motivations tend to advance or hinder inquiry. Concerns along these lines are probably at work in many positions in the debate on the potential epistemological significance of emotions. Within this volume, they are explicitly addressed in the papers of Tanesini and Thagard.

However, it is not at all clear what exactly the objection is, for the distinction between the context of discovery and the context of justification is not as plain as it might seem. As Paul Hoyningen-Huene has pointed out, it has been associated with at least five different senses, which often have been conflated (Hoyningen-Huene 1987, 504–6; Hoyningen-Huene 2006, 119–23). Firstly, the distinction can be taken to distinguish between two different types of historical processes, typically assumed to be temporally distinct. Or, secondly, between actual historical processes of discovery and considerations relevant to their justification; that is, between the factual and the normative. Thirdly, between empirical study of a discovery, which may involve historical, psychological and sociological research, and analysis or critical testing of justification by logical means. Fourthly, between academic disciplines with respect to their methodology; specifically, between epistemology and philosophy of science on the one hand, history, psychology and sociology on the other. Finally, between two types of questions.

According to Hoyningen-Huene's analysis, some of these distinctions face serious difficulties (especially the first one), but at the heart of the second to fifth version of the distinction we can identify a difference between factual and normative perspectives and questions (Hoyningen-Huene 1987, 511). When we examine cognitive agents, processes and products, we can either ask what exactly they are, or we can ask whether they are correct, good, rational, justified, etc. We can aim at accurate descriptions or at epistemic (or other) evaluations.

This way of understanding the distinction has the advantage that different questions can have the same answer, different perspectives can single out different aspects of the same practice, process or state (Hoyningen-Huene 2006, 129). Specifically, it would be fallacious to think that describing epistemic activities and agents by appeal to emotions precludes the latter from being relevant to the epistemic evaluation of the former. There is no reason why emotions should not feature in answers to both normative and empirical questions. Epistemic states as well as processes, agents and practices can all be examined from both factual and normative perspectives. Of course, we may look at processes of belief formation when epistemically evaluating beliefs or theories. In fact, this is just what causal, reliabilist and virtue theorists of epistemic justification do. For a belief to be epistemically justified is, according to those views, for it to be caused in the right way, to be produced by mechanisms with certain properties or to be formed by applying certain virtues. Clearly, such accounts take a normative perspective, for they attempt to spell out the conditions that determine whether something can be evaluated as epistemically justified.

Nonetheless, it is possible to appeal to the context distinction as a way of stressing that epistemological questions are evaluative, not descriptive. But of course, advocates of emotions in epistemology may defend the normative character of epistemology as well. Elgin, for example, is explicitly committed to doing so (Elgin 1996, 5–6). Consequently, the real disagreement concerns the question whether emotions matter at all in evaluating knowledge claims or only in various descriptions of cognitive

agents and activities. But defending the normative significance of emotions is not enough. For emotions to be epistemologically significant, they must not only be significant from a normative perspective, but from a specifically epistemic (as opposed to moral, practical, aesthetic) perspective. Critiques of emotions in epistemology may admit that emotions play an important role in living up to certain norms, but argue that these norms are epistemically irrelevant by pointing out that they are independent of epistemic ends, which are typically identified as truth and truth-conducive justification. Such a charge against, for example, the claim that cognitive efficiency is epistemically significant can be countered in various ways. One may defend that more efficiently arrived at beliefs are more likely to be true, that efficiency is suitably related to some other epistemic end or maybe that efficiency itself is an epistemic end.

Epistemic Emotions

The second issue concerns the nature of some of the states or dispositions epistemologists appeal to under the heading 'emotions'. It is not always clear that arguments for epistemological reappraisal of emotions really amount to anything like a general defence of the epistemological standing of emotions. To begin with, various authors have claimed that there are emotions which are tied to specifically epistemic contexts and thus have specifically epistemic character. Surprise is a paradigmatic case. Israel Scheffler ([1977]), for example, defines as 'cognitive emotions' those emotions which presuppose a claim that concerns the nature of the subject's cognitions. Especially interesting are cases where such a claim is epistemologically relevant to the cognitions it concerns. Surprise, for example, involves the 'supposition that what has happened conflicts with prior expectation' (Scheffler [1977], 12). But Scheffler also mentions joy of verification and disappointment or joy of falsification, echoing the quote from Schlick in the introduction to this paper. Such emotions are epistemologically significant since they embody epistemic standards. In a Peircean vein Scheffler then goes on to point out that a certain receptivity to surprise is an attitude which plays an important epistemic function shielding the inquirer from the 'epistemic apathy' (Scheffler [1977], 13) of radical scepticism and radical credulity, as well as from the inertia of dogmatism. We cannot hope to improve epistemically without allowing ourselves to be surprised (see Hookway, this vol., on doubt and Tanesini, this vol., on intellectual modesty).

On this base, one might object that even if epistemic emotions merit a place in epistemology, this may not be the case for emotions in general. Perhaps epistemic emotions are epistemologically relevant not insofar as they are emotions, but simply because they are epistemic. In addition, such an objection implicitly suggests that the whole case for the epistemological significance of emotions boils down to the truism that the epistemic is epistemologically significant.

However, neither Scheffler nor the other advocates of emotions rest their case just on arguments that are restricted to some narrowly defined class of epistemic emotions. In this volume, de Sousa, for example, gives an overview of epistemic feelings based on a double classification according to their object and the phase of inquiry in which they can occur. This covers a wide range of phenomena from

curiosity to certainty. But his account is not restricted to such epistemic feelings. For he also argues that emotions in general have epistemologically relevant aspects and consequently we should think of the emotions as epistemologically relevant in various degrees. Hookway, too, despite sometimes speaking of specifically epistemic emotions, often draws parallels between epistemic evaluations and emotional responses in general (Hookway 1998; Hookway, this vol.).

If we examine what states, dispositions and feelings have been invoked by epistemologists, we find not only paradigmatic emotions such as fear (de Sousa, this vol.; Elgin, this vol.) and anger (Thagard, this vol.), but also less straightforward examples like experience of beauty in the context of theory evaluation (Thagard 2002), de Sousa's (this vol.) feeling of knowing or Hookway's (2003a; this vol.) recourse to Quine's feeling of simplicity. Even if we grant epistemic value to those latter phenomena, we may still wonder whether they are emotions.[20] Or, to put it provocatively: have we started to call 'emotions' what has always been recognized as epistemologically relevant, albeit under different labels such as 'intuition'? Would not Descartes' 'mentis inspectio clara & distincta' (Descartes [1641], II.12, p. 31), translated as 'clear and distinct mental scrutiny' (Descartes 2002, 21), be a perfect example of what nowadays many call an 'epistemic emotion'? To some extent, such qualms can be dismissed as idle terminological questions about the use of the word 'emotion'. But this only highlights how important it is for epistemologists to discuss the question of which mental states and processes contribute to what aspects of epistemic practice and evaluation.

References

Alston, William P. 1967. 'Emotion and Feeling'. In: Edwards, Paul (ed.). *The Encyclopedia of Philosophy*. New York: Macmillan. Vol. 2. 479–86.

Aristotle. 1985. *The Complete Works of Aristotle. The Revised Oxford Translation*. Ed. Jonathan Barnes. Princeton: Princeton University Press.

Bless, Herbert; Fiedler, Klaus; Strack, Fritz. 2004. *Social Cognition. How Individuals Construct Social Reality*. Hove: Psychology Press.

BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge, Mass.: Harvard University Press.

Borges, Jorge L. [1942] 1964. 'The Analytical Language of John Wilkins'. In: *Other Inquisitions, 1937–1952*. Transl. Ruth L.C. Simms. Austin: University of Texas Press. 101–5.

Brown, Harold I. 1977. *Perception, Theory and Commitment. The New Philosophy of Science*. Chicago: Precedent Publishing.

Carnap, Rudolf. 1962. *Logical Foundations of Probability*. 2nd edition. Chicago: University of Chicago Press.

20 In psychology, cognitive and metacognitive 'emotions' (e.g. surprise, feelings of familiarity and accessibility of information; see Schwarz and Clore 2007) are sometimes termed 'nonaffective' (e.g. Bless et al. 2004) or 'nonemotional' (e.g. Stepper and Strack 1993) feelings and distinguished from emotions proper.

Carruthers, Peter; Stich, Stephen; Siegal, Michael (eds). 2002. *The Cognitive Basis of Science*. Cambridge: Cambridge University Press.

Cassin, Barbara (ed.). 2004. *Vocabulaire européen des philosophies. Dictionnaire des intraduisibles*. Paris: Editions du Seuil.

Code, Lorraine. 1984. 'Towards a "Responsibilist" Epistemology'. *Philosophy and Phenomenological Research* 45: 29–50.

Code, Lorraine. 1987. *Epistemic Responsibility*. Hanover: University Press of New England and Brown University Press.

Dalgleish, Tim; Power, Mick J. (eds). 1999. *Handbook of Cognition and Emotion*. Chichester: Wiley.

Damasio, Antonio R. [1994] 2000. *Descartes' Error. Emotion, Reason, and the Human Brain*. New York: Harper Collins.

Dancy, Jonathan; Ernest Sosa (eds). 2001. *A Companion to Epistemology*. Oxford: Blackwell.

Davidson, Richard J.; Scherer, Klaus R.; Goldsmith, H. Hill (eds). 2003. *Handbook of Affective Sciences*. New York: Oxford University Press.

Deonna, Julien A. 2006. 'Emotion, Perception and Perspective'. *Dialectica* 60: 29–46.

DePaul, Michael R.; Zagzebski, Linda (eds). 2003. *Intellectual Virtue. Perspectives from Ethics and Epistemology*. Oxford: Clarendon Press.

Descartes, René. [1641] 1996. *Meditationes de Prima Philosophia*. In: Adam, Charles; Tannery, Paul (eds). *Oeuvres de Descartes*. Paris: Vrin. Vol. 7.

Descartes, René. [1649] 1996. *Les passions de l'âme*. In: Adam, Charles; Tannery, Paul (eds). *Oeuvres de Descartes*. Paris: Vrin. Vol. 11. 293–497.

Descartes, René. 2002. *Meditations on First Philosophy. With Selections from the Objections and Replies*. Transl. and ed. John Cottingham. Cambridge: Cambridge University Press.

de Sousa, Ronald. [1979] 1980. 'The Rationality of Emotion'. In: Rorty 1980a, 127–51.

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.

de Sousa, Ronald. 1998. 'Modeling Rationality. A Normative or Descriptive Task?' In: Danielson, Peter A. (ed.). *Modeling Rationality, Morality and Evolution*. New York: Oxford University Press. 119–34.

de Sousa, Ronald. 2002. 'Emotional Truth'. *Proceedings of the Aristotelian Society*, *Supplementary Volume* 76: 247–63.

de Sousa, Ronald. 2003. 'Emotion'. In: Zalta, Edward N. (ed.). *The Stanford Encyclopedia of Philosophy* Spring 2003. http://plato.stanford.edu/archives/spr2003/entries/emotion/.

de Sousa, Ronald. 2004. 'Emotions. What I Know, What I'd Like to Think I Know, and What I'd Like to Think'. In: Solomon 2004b, 61–75.

de Sousa, Ronald. 2007. 'Truth, Authenticity, and Rationality'. *Dialectica* 61: 323–45.

Diamond, Cora. 1991. 'Knowing Tornadoes and Other Things'. *New Literary History* 22: 1001–15.

Diels, Hermann (ed.). 1951–52. *Die Fragmente der Vorsokratiker. Griechisch und Deutsch*. Ed. Walther Kranz. 6th edition. Berlin-Grunewald: Weidmannsche Verlagsbuchhandlung.

Döring, Sabine A. 2003. 'Explaining Action by Emotion'. *Philosophical Quarterly* 53: 214–30.

Dretske, Fred. 1970. 'Epistemic Operators'. *Journal of Philosophy* 67, 1007–23.

Ekman, Paul. 1977. 'Biological and Cultural Contributions to Body and Facial Movement'. In: Blacking, John (ed.). *The Anthropology of the Body*. London: Academic Press. 39–84.

Ekman, Paul. 1994. 'All Emotions Are Basic'. In: Ekman, Paul; Davidson, Richard J. (eds). *The Nature of Emotion. Fundamental Questions*. Oxford: Oxford University Press. 15–19.

Ekman, Paul. 1999a. 'Basic Emotions'. In: Dalgleish and Power 1999, 45–60.

Ekman, Paul. 1999b. 'Facial Expressions'. In: Dalgleish and Power 1999, 301–20.

Ekman, Paul; Sorenson, E. Richard; Friesen, Wallace V. 1969. 'Pan-Cultural Elements in Facial Displays of Emotion'. *Science* 164: 86–8.

Elgin, Catherine Z. 1996. *Considered Judgment*. Princeton: Princeton University Press.

Elgin, Catherine Z. 2004. 'True Enough'. In: Sosa, Ernest; Villanueva, Enrique (eds). *Epistemology.* (*Philosophical Issues* 14.) Malden: Blackwell. 113–31.

Elster, Jon. 1996. 'Rationality and the Emotions'. *Economic Journal* 106: 1386–97.

Evans, Dylan. 2002. 'The Search Hypothesis of Emotion'. *British Journal for the Philosophy of Science* 53: 497–509. [Reprinted in: Evans, Dylan; Cruse, Pierre (eds). 2004. *Emotion, Evolution, and Rationality*. Oxford: Oxford University Press. 179–91.]

Fairweather, Abrol. 2001. 'Epistemic Motivation'. In: Fairweather, Abrol; Zagzebski, Linda (eds). *Virtue Epistemology. Essays on Epistemic Virtue and Responsibility*. Oxford: Oxford University Press. 63–81.

Fodor, Jerry. 1984. 'Observation Reconsidered'. *Philosophy of Science* 51: 23–42.

Fodor, Jerry. 1987. *Psychosemantics*. Cambridge, Mass.: MIT Press.

Frijda, Nico H. 1993. 'Moods, Emotion Episodes, and Emotions'. In: Lewis and Haviland 1993, 381–403.

Goldie, Peter. 2000. *The Emotions. A Philosophical Exploration*. Oxford: Clarendon Press.

Goldie, Peter. 2004. 'Emotion, Feeling, and Knowledge of the World'. In: Solomon 2004b, 91–106.

Goldman, Alvin. 1967. 'A Causal Theory of Knowing'. *Journal of Philosophy* 64: 357–72.

Goldman, Alvin. 1976. 'Discrimination and Perceptual Knowledge'. *Journal of Philosophy* 73: 771–91.

Goldman, Alvin. 1986. *Epistemology and Cognition*. Cambridge, Mass.: Harvard University Press.

Goldman, Alvin. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.

Goodman, Nelson. 1976. *Languages of Art. An Approach to a Theory of Symbols*. 2nd edition. Indianapolis: Hackett.

Goodman, Nelson. 1983. *Fact, Fiction, and Forecast*. 4th edition. Cambridge, Mass.: Harvard University Press.

Goodman, Nelson; Elgin, Catherine Z. 1988. *Reconceptions in Philosophy and Other Arts and Sciences*. London: Routledge.

Gordon, Robert M. 1987. *The Structure of Emotions. Investigations in Cognitive Philosophy*. Cambridge: Cambridge University Press.

Griffiths, Paul E. 1997. *What Emotions Really Are. The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Griffiths, Paul E. 2004a. 'Emotions as Natural and Normative Kinds'. *Philosophy of Science* 71: 901–11.

Griffiths, Paul E. 2004b. 'Is Emotion a Natural Kind?' In: Solomon 2004b, 233–49.

Hookway, Christopher. 1990. *Skepticism*. London: Routledge.

Hookway, Christopher. 1998. 'Doubt. Affective States and the Regulation of Inquiry'. In: Misak, Cheryl (ed.). *Pragmatism*. *Canadian Journal of Philosophy*, suppl. vol. 24: 203–25.

Hookway, Christopher. 2000. 'Regulating Inquiry. Virtue, Doubt and Sentiment'. In: Axtell, Guy (ed.). *Knowledge, Belief, and Character. Readings in Virtue Epistemology*. Lanham: Rowman and Littlefield. 149–60.

Hookway, Christopher. 2002. 'Emotions and Epistemic Evaluations'. In: Carruthers et al. 2002, 251–62.

Hookway, Christopher. 2003a. 'Affective States and Epistemic Immediacy'. *Metaphilosophy* 34: 78–96.

Hookway, Christopher. 2003b. 'How to Be a Virtue Epistemologist'. In: DePaul and Zagzebski 2003, 183–202.

Hoyningen-Huene, Paul. 1987. 'Context of Discovery and Context of Justification'. *Studies in History and Philosophy of Science* 18: 501–15.

Hoyningen-Huene, Paul. 2006. 'Context of Discovery versus Context of Justification and Thomas Kuhn'. In: Schickore, Jutta; Steinle, Friedrich (eds). *Revisiting Discovery and Justification. Historical and Philosophical Perspectives on the Context Distinction*. Dordrecht: Springer. 119–31.

Hume, David. [1739/40] 1978. *A Treatise of Human Nature*. Ed. Lewis A. Selby-Bigge and Peter H. Nidditch. Oxford: Oxford University Press.

Jaggar, Alison. 1989. 'Love and Knowledge. Emotion in Feminist Epistemology'. In: Garry, Ann; Pearsall, Marilyn (eds). *Women, Knowledge, and Reality*. Boston: Unwin Hyman.

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

James, William. [1890] 1950. *The Principles of Psychology*. Vol. 2. New York: Dover.

Kant, Immanuel. [1781/87] 1976. *Kritik der reinen Vernunft*. Hamburg: Meiner.

Kaplan, Mark. 1985. 'It's Not What You Know That Counts'. *Journal of Philosophy* 82: 350–63.

Kenny, Anthony. [1963] 2003. *Action, Emotion and Will*. New edition. London: Routledge.

Ketelaar, Timothy; Todd, Peter M. 2001. 'Framing Our Thoughts. Ecological Rationality as Evolutionary Psychology's Answer to the Frame Problem'.

In: Holcomb III, Harmon R. (ed.). *Conceptual Challenges in Evolutionary Psychology. Innovative Research Strategies*. Dordrecht: Kluwer. 179–211.

Konstan, David. 2006. *The Emotions of the Ancient Greeks. Studies in Aristotle and Classical Literature*. Toronto: University of Toronto Press.

Kornblith, Hilary. 2006. 'Appeals to Intuition and the Ambitions of Epistemology'. In: Hetherington, Stephen (ed.). *Epistemology Futures*. Oxford: Clarendon Press. 10–25.

Lance, Mark; Tanesini, Alessandra. 2004. 'Emotion and Rationality'. In: Ezcurdia, Maite; Stainton, Robert; Viger, Christopher (eds). *New Essays in the Philosophy of Language and Mind. Canadian Journal of Philosophy*, suppl. vol. 30: 275–95.

Lazarus, Richard S. 1999. 'The Cognition–Emotion Debate. A Bit of History'. In: Dalgleish and Power 1999, 3–19.

LeDoux, Joseph E. 1998. *The Emotional Brain. The Mysterious Underpinnings of Emotional Life*. London: Weidenfeld and Nicolson.

L'Estrange, Roger. [1699] 1972. *Fables, of Aesop and Other Eminent Mythologists. With Morals and Reflections*. Westmead: Gregg. (Reprint of the 3rd edition, London, 1699.)

Lewis, Michael; Haviland, Jeannette M. (eds). 1993. *Handbook of Emotions*. New York: Guilford Press.

Lyons, William. 1980. *Emotion*. Cambridge: Cambridge University Press.

Lyons, William. 1999. 'The Philosophy of Cognition and Emotion'. In: Dalgleish and Power 1999, 21–44.

Marks, Joel; Ames, Roger T. (eds). 1995. *Emotions in Asian Thought. A Dialogue in Comparative Philosophy*. Albany: State University of New York Press.

Montmarquet, James. 1993. *Epistemic Virtue and Doxastic Responsibility*. Lanham: Rowman and Littlefield.

Niiniluoto, Ilkka; Sintonen, Mattik; Woleński, Jan (eds). 2004. *Handbook of Epistemology*. Dordrecht: Kluwer.

Nisbett, Richard E.; Ross, Lee D. 1980. *Human Inference. Strategies and Shortcomings of Social Judgment*. Englewood Cliffs: Prentice-Hall.

Nussbaum, Martha C. 2001. *Upheavals of Thought. The Intelligence of Emotions*. Cambridge: Cambridge University Press.

Oatley, Keith. 1999. 'Emotions'. In: Wilson, Robert A.; Keil, Frank C. (eds). *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, Mass.: MIT Press. 273–5.

Oatley, Keith; Keltner, Dacher; Jenkins, Jennifer M. 2006. *Understanding Emotions*. 2nd edition. Malden: Blackwell.

Ortony, Andrew; Turner, Terence J. 1990. 'What's Basic about Basic Emotions?' *Psychological Review* 97: 315–31.

Peters, Ellen. 2006. 'The Functions of Affect in the Construction of Preference'. In: Lichtenstein, Sarah; Slovic, Paul (eds). *The Construction of Preference*. Cambridge: Cambridge University Press. 454–63.

Plantinga, Alvin. 1986. 'Epistemic Justification'. *Noûs* 20: 10–19.

Plutchik, Robert. 1980. *Emotion. A Psychoevolutionary Synthesis*. New York: Harper and Row.

Plutchik, Robert. 2003. *Emotions and Life. Perspectives from Psychology, Biology, and Evolution*. Washington, DC: American Psychological Association.

Prinz, Jesse J. 2004a. 'Embodied Emotions'. In: Solomon 2004b, 44–58.
Prinz, Jesse J. 2004b. *Gut Reactions. A Perceptual Theory of Emotion*. New York: Oxford University Press.
Prinz, Jesse J. 2004c. 'Which Emotions Are Basic?' In: Evans, Dylan; Cruse, Pierre (eds). *Emotion, Evolution, and Rationality*. Oxford: Oxford University Press. 69–87.
Quine, Willard V. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.
Quine, Willard V. 1969. 'Epistemology Naturalized'. In: *Ontological Relativity and Other Essays*. New York: Columbia University Press. 69–90.
Reichenbach, Hans. 1935. 'Zur Induktionsmaschine'. *Erkenntnis* 5: 172–3.
Reichenbach, Hans. 1938. *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press.
Roberts, Robert C. 2003. 'Emotions. An Essay in Aid of Moral Psychology'. Cambridge: Cambridge University Press.
Rorty, Amélie O. (ed.). 1980a. *Explaining Emotions*. Berkeley: University of California Press.
Rorty, Amélie O. 1980b. 'Introduction'. In: Rorty 1980a, 1–8.
Rorty, Amélie O. 2004. 'Enough Already with "Theories of the Emotions"'. In: Solomon 2004b, 269–78.
Ryle, Gilbert. 1949. *The Concept of Mind*. London: Hutchinson's University Library.
Salmela, Mikko. 2006. 'True Emotions'. *Philosophical Quarterly* 56: 382–405.
Scheffler, Israel. [1977] 1991. 'In Praise of the Cognitive Emotions'. In: Scheffler, Israel. *In Praise of the Cognitive Emotions and Other Essays in the Philosophy of Education*. New York: Routledge. 3–17.
Schlick, Moritz. [1934] 1979. 'On the Foundation of Knowledge'. In: Mulder, Henk L.; van de Velde-Schlick, Barbara (eds). *Moritz Schlick. Philosophical Papers.* Vol. 2. 1925–1936. Dordrecht: Reidel. 370–87. (German original in *Erkenntnis* 4: 79–99.)
Schwarz, Norbert; Clore, Gerald L. 2007. 'Feelings and Phenomenal Experiences'. In: Kruglanski, Arie W.; Higgins, E. Tory (eds). *Social Psychology. A Handbook of Basic Principles*. 2nd edition. New York: Guilford Press. 385–407.
Solomon, Robert C. 1993a. *The Passions. Emotions and the Meaning of Life*. 2nd edition. Indianapolis: Hackett.
Solomon, Robert C. 1993b. 'The Philosophy of Emotion'. In: Lewis and Haviland 1993, 3–15.
Solomon, Robert C. 2003. 'Emotions, Thoughts and Feelings. What Is a "Cognitive Theory" of the Emotions and Does It Neglect Affectivity?' In: Hatzimoysis, Anthony (ed.). *Philosophy and the Emotions*. (*Royal Institute of Philosophy Supplement* 52). Cambridge: Cambridge University Press. 1–18.
Solomon, Robert C. 2004a. 'Emotions, Thoughts, and Feelings. Emotions as Engagements with the World'. In: Solomon 2004b, 76–88.
Solomon, Robert C. (ed.). 2004b. *Thinking about Feeling. Contemporary Philosophers on Emotions*. Oxford: Oxford University Press.
Solomon, Robert C. 2007. *True to Our Feelings. What Our Emotions Are Really Telling Us*. Oxford: Oxford University Press.

Sorabji, Richard. 2000. *Emotion and Peace of Mind. From Stoic Agitation to Christian Temptation. The Gifford Lectures*. Oxford: Oxford University Press.

Sosa, Ernest. 1980. 'The Raft and the Pyramid. Coherence versus Foundations in the Theory of Knowledge'. *Midwest Studies in Philosophy* 5: 3–25.

Sosa, Ernest. 1985. 'Knowledge and Intellectual Virtue'. *Monist* 68: 224–45.

Stepper, Sabine; Strack, Fritz. 1993. 'Proprioceptive Determinants of Emotional and Nonemotional Feelings'. *Journal of Personality and Social Psychology* 64: 211–20.

Thagard, Paul. 2000. *Coherence in Thought and Action*. Cambridge, Mass.: MIT Press.

Thagard, Paul. 2002. 'The Passionate Scientist. Emotion in Scientific Cognition'. In: Carruthers et al. 2002, 235–50.

White, Geoffrey M. 1993. 'Emotions Inside Out. The Anthropology of Affect'. In: Lewis and Haviland 1993, 29–39.

Zagzebski, Linda. 1996. *Virtues of the Mind. An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.

This page intentionally left blank

Chapter One

Emotion and Understanding

Catherine Z. Elgin*

Emotions share important cognitive functions with perceptions and beliefs. Like perceptions, they afford epistemic access to a range of response-dependent properties, such as being admirable or contemptible, and provide evidence of response-independent properties that trigger them. Fear is evidence of danger; trust is evidence of reliability. Like beliefs, emotions provide orientations that render particular facets of things salient. In the grip of an emotion, we notice things we would otherwise miss. The variability and volatility of emotional deliverances might seem to undermine their claim to epistemic standing. I argue that variability and volatility can be epistemic assets, keying the subject to multiple, quickly changing features of things. Emotions, like other modes of epistemic access, are subject to refinement to increase their epistemic yield. The arts provide opportunities for such refinement.

1. The Claim to Epistemic Standing

'Reason is and ought only to be the slave of the passions,' Hume maintains ([1739/40], 415). Others reverse the relationship, contending that to reason well requires subduing, dominating or discounting the passions. Either way, reason and passion are antithetical. To be under the sway of emotion is to be irrational. To be rational is to be cool, calm, and deliberate; that is, to be unaffected by emotion. Let us call this the standard view. I think it is a mistake. Rather than being opposed to reason, I suggest, emotion is a facet of reason. It is an avenue of epistemic access, hence a contributor to the advancement of understanding.

Emotional deliverances are representations conveyed through emotional channels. A representation of frogs as dangerous that presents itself via fear of frogs is a deliverance of that fear. If the standard view is correct, emotional deliverances are at best epistemically inert, there being no reason to trust them. At worst, they are deleterious; there is reason to distrust them. Either way, insofar as our goals are cognitive, if we cannot subdue or silence the passions, we should ignore or discount their deliverances.

In contending that the standard view is mistaken, I do not mean that every emotional deliverance, as it stands, is epistemically acceptable. Rather, I believe that emotions provide resources that serve epistemic ends. But just as natural resources like iron ore need to be processed to yield material we value, so do emotional deliverances.

* I am grateful to Amélie Rorty, Sabine Döring, and members of the University of Zürich Workshop on Epistemology and Emotions for comments on earlier drafts of this paper.

We need to know how to recognize the epistemically valuable insights that emotions afford and how to use them effectively. The goal of this chapter is twofold: first, to show that emotions afford such resources; and second, to say something about how to refine the raw materials and increase their epistemic yield.

Elsewhere I have argued that an understanding is a system of cognitive commitments in reflective equilibrium. The individual commitments that comprise such a system must be reasonable in light of one another, and the system as a whole must be at least as reasonable as any available alternative in light of relevant antecedent commitments (Elgin 1996). A system's equilibrium derives from the mutual supportiveness of the components. Its answering to antecedent commitments at least as well as available alternatives insures that the equilibrium is one that on reflection we can accept. Not all the cognitive commitments that comprise such a system are truth bearers: perspectives, vocabularies, methods and standards are as integral to an understanding as beliefs. Nor need the truth bearers in a tenable system always be true. A rudimentary science that contains only rough approximations can provide some understanding of its subject. That is why it is worth taking seriously. Finally, understanding is holistic. No commitment is fully tenable in isolation.

A systematic, interconnected network of utterly untenable commitments would not yield an understanding of the phenomena it bears on. Coherence alone does not suffice for tenability. The tie to antecedent considerations is crucial. Unless some considerations have some initial tenability prior to systematization, warrant cannot be generated. But this is not our situation.

At any given time, a person has a cluster of cognitive commitments. These include her beliefs, perceptual takings, and emotional deliverances. They also include the methods, perspectives, and sources she tends to trust, as well as the epistemic priorities and weighting factors she endorses. All such commitments are initially tenable. Their being held gives them a slight claim on epistemic allegiance. But initial tenability is not full tenability. No commitment, however firmly held, is fully warranted in isolation. Epistemic warrant accrues through systematization – the development of an integrated network of mutually supportive cognitive commitments.

Systematization requires revising as well as conjoining and augmenting, for a person's initially tenable commitments can, and often do, clash. They are apt to be incompatible, non-cotenable or implausible in light of one another. So to arrive at an acceptable system, we need to revise and/or reject some of our initial commitments and to adopt others that we previously did not hold. A tenable system thus need not incorporate all the antecedent commitments it answers to. But if it does not, it should show why the commitments it rejects seemed reasonable when they did. It might, for example, reveal that they owed their previous plausibility to limited evidence or unsophisticated models. An individual commitment is warranted by its place in a tenable system of thought.

The crucial point is this: To say that emotional deliverances are initially tenable is to assign them only a weak and precarious epistemological status. They have that status because an agent, in the grip of an emotion, has a tendency to credit its deliverances. To experience a frisson as a fear of frogs, is to take that frisson to embed the idea that frogs are dangerous.

To credit emotional deliverances with initial tenability might seem trifling. The epistemic status of emotions turns not on whether their deliverances are initially tenable, but on whether they are fully tenable. If such deliverances are immediately and decisively overridden, then the fact that they start out with some measure of initial tenability seems insignificant. This concern has merit. For there is a clash. The standard view that emotional deliverances are unreliable is in tension with the opinions emotions embed. But the standard view is not unfounded. Emotions at least sometimes distort or derail reason. The challenge facing me is to show that the best way to alleviate the tension is to reject or revise the standard view. This does not require saying that all emotions under all circumstances are epistemically estimable, or that the deliverances of those that are must be taken at face value. It requires only showing that systems that integrate emotional deliverances are sometimes more tenable than rivals that exclude them.

Although widespread, the conviction that emotions are epistemically inert is a bit odd. Emotions are not spontaneous upwellings of arbitrary feelings. They are reactions to events. So if we can correlate emotional reactions with the events that trigger them, we can use those reactions as sources of information about the environment. Experiencing emotion *a* would be an indication that circumstance *b* obtains. The standard view considers such correlations inherently unreliable. To be sure, the existence of a causal connection between individual emotion and event pairs does not show otherwise. What we need is evidence that emotional reactions are reliably correlated with events. If an arbitrary subject were as likely to feel joy, dismay, revulsion or amusement, regardless of the trigger, then from the occurrence of a particular emotion, nothing would follow about the nature of its source. The question then is whether suitably reliable correlations can be found.

Our views about the informativeness of emotions waver. In the grip of an emotion, one normally believes its deliverances. When I am frightened, I think that the situation is dangerous. When I am infatuated, I think that my beloved is wonderful. Nor, at the time, do I consider the connection between my occurrent emotions and beliefs accidental. I am frightened, I believe, *because* the situation is dangerous. I adore him, I believe, *because* he is wonderful. In cooler moments, I may think differently. I recognize that many of my fears have proven unwarranted. I concede that I have not been drawn unerringly to wonderful men. Such failures might persuade us that suitably reliable correlations are not to be had. Then whatever we think or feel in the heat of the moment, it might be wise to defer to our cooler judgement that emotional deliverances are not trustworthy sources of information. Still, we go too quickly, I think, if we dismiss them.

One reason is biological (de Sousa 1987). The limbic system, the seat of emotion, is a product of evolution. Although evolution yields some by-products that lack survival value, the fact that a trait or complex of traits is a product of evolution is evidence that it promotes fitness. Emotions might promote fitness by being indicators of significant features of the environment. If fear is triggered by danger and is a sufficiently reliable indicator of danger, then developing and exercising the capacity for fear would be conducive to survival. If having their father present enhances the survival prospects of young humans, then a mother's eliciting the love of someone she can live with promotes her reproductive success.

Such evolutionary arguments are weak. The evolutionary lineage of emotions is compatible with their having no survival value at all. They could be 'free riders' on genes whose contribution to fitness lies elsewhere. At most, the evolutionary argument shows that there is some presumption that the capacity to experience emotions is adaptive. It does not show that the capacity to experience *each* emotion is adaptive. Nor does it show that if that capacity is adaptive, its adaptiveness lies in a correlation between emotions and circumstances. Perhaps the deliverances of emotions, like the contents of dreams, are independent of the circumstances in which they occur. The biological contribution of emotions might then lie in something like their expending excess neural energy or in making and breaking covalent bonds in the amygdala, not in anything that affords epistemic access to the circumstances in which they occur. Nevertheless, it is not unreasonable to think that fear promotes fitness by sensitizing to danger or that affection between parents promotes fitness by fostering living conditions that increase the likelihood that offspring will survive. The fact that emotions evolved affords some presumption that their deliverances vary with circumstances. Given that presumption, it is not unreasonable to think that they are keyed to circumstances in such a way that if we can discover the key, we can use the occurrence of particular emotions as sources of information about the circumstances in which they occur. All the evolutionary argument contends is that it is not unreasonable to suppose that emotions are capable of affording epistemic access to their objects. Weak, though it is, the argument is strong enough to do that.

2. Response Dependence

Some properties – such as being contemptible, or admirable, or amusing, or depressing – dovetail with emotions in such a way that without those emotions, the properties would not exist. These properties are response-dependent (Johnston 1989). They are genuine properties of the objects that possess them, but they owe their identities to responses they evoke. If contemptibility is the property it is because of the contempt it is apt to evoke, it would be astonishing if feelings of contempt did not afford epistemic access to contemptibility. This is not, of course, to say that my contempt for Karl demonstrates that he is contemptible. My contempt could be misplaced. But if the property of being contemptible depends for its identity on evoking feelings of contempt, then at least some evokings of contempt and perhaps other related emotions should be evidence of contemptibility. Indeed, emotional responses seem to afford our most direct access to such properties.

The foregoing considerations suggest parallels to perception. Perception is both triggered by and indicative of aspects of the environment. Perceptual systems evolved and endure because their deliverances promote fitness. Being able to see, or hear, or smell a predator, like feeling instinctively afraid of it, enhances an animal's prospects of evading it. Perception manifestly affords epistemic access to useful information about the environment.

Secondary qualities, such as colours, tones, and odours, are response-dependent. They are known by perception, and perhaps cannot be fully known without perception (Jackson 1982). Still, mistakes are possible. Something's looking blue to me does

not establish that it is blue. An isolated colour perception is not epistemically acceptable on its own. But something's looking blue is ordinarily evidence that it is blue. If the analogy holds, emotional deliverances are indicators, but not always accurate indicators of aspects of their objects. Just as my experiencing something as blue is evidence, but not conclusive evidence, that it is blue, my being frightened of something is evidence, but not conclusive evidence, that it is dangerous.

According to a familiar criterion for colours, something is blue just in case it would look blue to a normal observer under normal conditions. Conditions often are normal and most of us are normal colour perceivers. So in the absence of contraindications, one reasonably takes it that something's looking blue to her is reason to believe that it is blue. But the criterion for secondary qualities is not always the response of the normal perceiver. In assessing whether a bassoon is in tune, the criterion is not how it would sound to a normal perceiver, but how it would sound to a perceiver with perfect pitch. Unless I believe myself to have perfect pitch, I assign considerably less weight to my judgement that a bassoon is in tune than I do to my judgement that the bassoonist is wearing a blue shirt. If the analogy with emotions holds, the criterion for some cases could be what a normal subject would feel in normal circumstances (with both occurrences of 'normal' spelt out in an informative way), while in other cases it could be what a suitably sensitive subject would feel in specified circumstances. Like perceptions, emotions might key to circumstances in any of several different ways.

The capacity to distinguish among secondary qualities is neither uniform nor fixed. One perceiver might consider two colours or tones identical, while another readily distinguishes between them. A single person's capacity to discriminate may vary over time and with circumstances. Colours that look indistinguishable against one background are often easily distinguished against another. Moreover, perceptual abilities are subject to refinement. We can develop an ability to distinguish between properties that we previously could not tell apart. At a wine tasting, for example, one learns to taste the difference between wines that initially tasted the same.

Although secondary qualities are more tightly linked to perception than other perceptible properties, the same variability, fallibility and sensitivity to circumstances is characteristic of perception in general. Some people can identify the makes of American cars on sight; others cannot. Some can distinguish between deciduous trees in full leaf, but not when the limbs are bare. Skiers learn to differentiate qualities of snow on the basis of the way the snow looks and feels, even though originally snow seemed to them to be just one, undifferentiated sort of stuff.

Despite this variability, fallibility and sensitivity to circumstances, we readily concede that perceptual deliverances afford epistemic access to their objects. So the fact that emotional deliverances are similarly variable, fallible, and sensitive does not disqualify them from epistemic standing. Rather, the epistemic yield of emotions, like the epistemic yield of perceptions depends not on taking all deliverances at face value, but on a sophisticated understanding of when and to what extent they are trustworthy.

Our practice of assessing the appropriateness of various emotions reflects such an understanding. Phobias are irrational fears. Our ability to distinguish between phobias and other fears shows that we have a lien on which fears are rational. It is

hard to see what the basis for the distinction would be, if it were not that rational fears are ones that align with genuine dangers. To recognize a fear as rational is to recognize that the belief that its object is dangerous is at least prima facie rational as well. We also make more local assessments, for example, when we charge people with overreacting. We say that Harry is insanely jealous, indicating that his jealousy is far greater than it should be. This is quite different from saying that he has no basis for jealousy. We say that Fred is excessively angry, indicating that some lesser level of anger would be appropriate. We think that Sam should be proud of himself, indicating that were he to feel pride, his doing so would be reasonable. Our propensity for charging people with over- or underreacting indicates that we take ourselves to be in a position to tell when people's emotions are apt. That is, we take ourselves to be able to reliably correlate emotions with circumstances. The robustness of our practice is evidence that we are right.

3. Is Instability a Problem?

Despite these parallels to perception, one might argue, the deliverances of emotions are untrustworthy, because emotions are far more volatile than perceptions. To investigate this charge, we need to examine two questions: (1) whether emotions are volatile and (2) whether volatility is epistemically incapacitating.

As Hume notes, not all passions are volatile (Hume [1739/40], 276). Calm passions, such as fondness for a lifelong friend, are steady, enduring, and unexciting. Nor need more violent passions be volatile. Some people carry a grudge for decades, without in the least moderating their attitude toward its object. Although some emotions may be volatile, volatility is not an invariable property of emotions. If volatility is an epistemic disqualifier then, only some emotions and their deliverances are disqualified.

But it is not clear that volatility should be a disqualifier. The question is whether emotions are, and can be known to be, so tied to circumstances that from the occurrence of an emotion we can glean information about the circumstances. There is no obvious reason why a volatile emotion should not be so tied. Suppose, for example, that fury is a volatile emotion, which normally arises only when exceedingly objectionable events occur and which ebbs quickly when objectionability wanes. In that case, if objectionability quickly waxes and wanes, fury's volatility is simply a responsiveness to rapidly changing circumstances. This would be an epistemic asset rather than a defect. Evidently, neither the violence nor the volatility of an emotion undermines its epistemic qualifications. To do that would require showing that the emotions vary independently of circumstances – that, for example, fury arises or endures regardless of the objectionability of the object.

This might be the case. Emotional reactions seem to vary considerably from one person to another. The same object can anger one person, mildly irritate a second, sadden a third and amuse a fourth. That being so, one might think, an emotional deliverance could not possibly afford reliable information about its object. Such wide variation might suggest that emotional deliverances are entirely subjective.

Pretty much anything, it seems, could trigger any emotion; so from the occurrence of an emotion, nothing can be gleaned about the nature of the object.

Again it is worth considering the analogy with perception. Similarly situated observers viewing the same scene may see different things because of differences in background knowledge and interests. A paediatrician sees a case of measles, when a parent just sees her child's rash. A linguist hears a Texas accent, while a reporter hears a political speech, and the audience hears a call to arms. A critical grandmother sees crude, ill-mannered behaviour in her granddaughter, whereas a doting grandmother sees the same behaviour as refreshingly spontaneous (Murdoch 1970, 17–18). If the rash is a case of measles, both the parent and the paediatrician are right. If the political speech is a call to arms issued in a Texas accent, all three auditors are right. If the granddaughter's behaviour lies at the intersection of 'crude and ill mannered' and 'refreshingly spontaneous', both grandmothers are right. The mere fact that different observers perceive the same situation differently does not show that any of them is wrong, or that perception is not reliable.

Many events are both infuriating and depressing, so the fact that some people experience anger and others sadness in response to the same event does not show that the responses do not track features of the world. But, we are apt to think, if it is infuriating, it is *not* amusing. So the fact that the same event infuriates some people while amusing others seems strong evidence of the subjectivity, hence untrustworthiness, of emotional deliverances. This need not be so. It might be evidence that emotions are perspectival. A Red Sox fan sees the game winning play while a Yankees fan sees the game losing play. They are looking at exactly the same play. And they are discerning properties that the play genuinely has. The play that wins the game for one team loses it for the other. The spectators, given their diverging allegiances, simply orient themselves to different aspects of the play. One group see it in terms of its consequences for the Yankees; the other, in terms of its consequences for the Red Sox. Analogously, the subject who is amused or elated by the event that others find sad or infuriating may have a different perspective on it, one that discloses aspects of the event that the others are insensitive to. That the Red Sox fan is happy and the Yankees fan is sad is no surprise. From the Red Sox fan's point of view, the outcome is pleasing; from the Yankees fan's point of view, it is displeasing. A single situation can have multiple emotion-sensitive aspects. The diversity of responses can be due to the fact that different respondents are sensitive to different aspects.

Such variability would show, not that emotions are epistemically inert, but that reactions require calibration along at least three dimensions: the perspective the subject adopts, how sensitive a subject is, and which emotions dominate in the subject. Maasai tribesmen inure themselves to pain. Middle class Americans do not. Hence a Maasai's wince evinces approximately the same intensity of pain as an American's groan. So a physician who is attempting to determine the level of her patient's pain needs to know whether her patient is exhibiting American or Maasai pain behaviour. If analogously, a highly emotional person's elation is approximately equivalent to a stolid person's mild pleasure as evidence of joyousness, then to figure out how joyous an event is, we need to factor in the responsiveness of the responder. We also need to know about the subject's responsiveness to different

emotion-triggering properties: If a person encounters something that is both funny and sad, is she likely to be amused, saddened or both? Attempting to address the issue at this level of generality is of course much too crude. For emotional responses are highly sensitive to context and history. We need to know a good deal about the respondent, including her recent history and her relationships to and attitudes regarding the object. Her response to someone's acute embarrassment might depend on whether that person is someone she loves or loathes. If the former, she is apt to feel pity or sympathy. If the latter, she may feel schadenfreude.

Even if emotional deliverances are less reliable than perceptual deliverances, it does not follow that they are epistemically inert. In neither case are deliverances creditable in isolation. Perceptions that cohere with other things we believe or endorse easily integrate into a tenable system of thought. Those that do not require more support in order to be tenable. Often that support is available, so the fact that a perceptual deliverance is suspect does not utterly disqualify it. Moreover, we learn which perceptions are trustworthy, and in what circumstances. We do not, for example, trust our colour perceptions at dusk or our judgements about delicate distinctions in flavour immediately after eating a jalapeño pepper. So if emotional deliverances are less reliable than perceptual deliverances, they need more collateral support in order to be tenable. For they start out with less initial tenability. But to have less initial tenability is not to have none. The very fact that they present themselves as indicators of how things stand gives them some degree of initial tenability.

Moreover, no more than perceptual deliverances are emotional deliverances on a par. As we learn more about the world and our emotional attunement to it, we can assign different degrees of initial tenability to different deliverances. Having learnt that peripheral vision is not as accurate as focal vision, we assign less weight to what we see out of the corner of the eye than we do to what we see at the centre of the visual field. Having learnt that I am prone to overreacting to offences against my child, I take my feelings of outrage as evidence not of the outrageousness of the situation, but of its having some, perhaps small, measure of objectionability.

To make the case that emotions afford epistemic access to aspects of their objects, we need not establish correlations for every emotion. It suffices if we can identify some recognizable subset whose members are reliable indicators. Their deliverances are epistemically estimable even if the deliverances of other emotions were not. Moreover, the correlations can be different for different subjects and/or against different backgrounds. The upshot is this: Emotional deliverances, like perceptual deliverances, afford epistemic access to their objects, but not in every case. Nor does a given emotion have the same correlate, regardless of subject or circumstances. So rather than saying that emotional deliverance *d* indicates the presence of property *p*, it may be better to say that in circumstances *c*, subject *a*'s emotional deliverance *d* indicates property *p*.

The payoff so far may seem slight. I have been speaking of emotional deliverances as though each deliverance takes the form of a particular judgement: 'that is dangerous'; 'this is delightful'; 'that is repulsive', and so on. I mentioned that there are response-dependent properties of things, where the responses on which they depend are emotional. Although our emotions are not utterly reliable indicators of the presence of such properties, we can often tell which emotional reactions

reflect the presence of emotional response-dependent properties. So under certain recognizable circumstances, an emotional reaction affords epistemic access to such properties.

4. Complexities of Response Dependence

Even if this were the whole story, it would not be trivial. A response-dependent property is not just a brute propensity to trigger a given response. The property is identified by reference to its ability to trigger a response, but the property is not just the bald capacity to trigger that response. In fact, response-dependent properties are often highly complex. Something's being blue, for example, involves its having a complex capacity to reflect and absorb light. Moreover, in perceiving something as blue, we do more than just react differentially to an instantaneous stimulus. The perception orients us to the past and the future. We count the current deliverance as of a new colour if it differs from its immediate predecessor. We stand ready to judge immediately subsequent colour perceptions as the same as the current one only if they are as of blue. We are surprised or unsurprised, depending on whether the item that presents itself as blue is the sort of thing we would expect to see and the sort of thing we would expect to be blue, and so on. We license ourselves to draw inferences about ourselves and the situation that would be ungrounded had our current perceptual experience been different.

Emotional response-dependent properties are also complex. A situation is blameworthy only if it is unfortunate and a subject is somehow responsible for its occurrence. Remorse is an emotion triggered by a sense of blameworthiness. It may be felt as a stab of distress, but it is more than that. It orients its subject to the past and the future in ways that other stabs of distress, such as feelings of regret, do not. To the past, because it embeds a feeling that something she did or refrained from doing contributed to the misfortune. To the future, because her sense of blameworthiness colours her feelings about her obligations and opportunities, and her sense of herself as a moral agent with on-going relations to other moral agents. Indeed, as stabs of distress, regret and remorse may feel exactly the same. Then their difference lies in the orientations that they supply. A destructive tsunami is terrible, hence highly regrettable. It is not blameworthy, since no one causes it and no one can prevent it. Tragic though it is, no one should feel remorse over its occurrence.

Response-dependent properties vary in that some are more normative than others. Something is visible only if it can be seen. Being visible seems to be a dispositional property having little if any normativity. Being blue involves a norm: something is blue only if it would be perceived as blue by a normal perceiver under normal circumstances. Being in tune involves an ideal. Something is in tune only if it would be perceived as in tune by someone with perfect pitch. Emotionally response-dependent properties show similar variability. Something is enjoyable only if it is capable of being enjoyed. Something is indecent only if it would offend a normal subject in normal circumstances. Something is admirable only if it ought to be admired and despicable only if it ought to be despised. Perhaps the norms to which normal agents respond can be identified purely sociologically. But the conditions

for the application of the more evaluative emotional response-dependent concepts cannot. Such concepts are what Bernard Williams calls 'thick concepts'. They fuse descriptive and evaluative elements (Williams 1985, 129).

Trust is the emotion that tracks the response-dependent property of trustworthiness. Truthfulness is (at least a major component of) trustworthiness of an informant. Williams's account of truthfulness affords insights into trustworthiness and trust, and indirectly into the complexity of other thick emotional concepts.

Williams explicates the thick concept of truthfulness in terms of accuracy and sincerity (Williams 2002). Trusting an informant, he maintains, is taking her utterance to be accurate and sincere. Both accuracy and sincerity require explication. It might seem that someone is sincere just in case she says what she believes. Then she is not lying. But a person can be untruthful without lying, for truthfulness is a matter of conveying what one takes to be true, not of stating what one takes to be true. One can convey false beliefs by expressing misleading truths. So, perhaps truthfulness requires that one neither lie nor mislead. Williams considers this too simple. For truthfulness is a virtue. Since not all lies, evasions, obfuscations and omissions are morally objectionable, an agent is not morally obliged to refrain from conveying all untruths. One aspect of the virtue of truthfulness is a sensitivity to when truthfulness is called for. Some information is private. Your interlocutor may have no right to know the details of your love life or your trade secrets or your investment strategy. If refusing to answer unduly intrusive questions is unfeasible, Williams maintains, evasion is permissible. Your interlocutor should recognize that in relentlessly pursuing such questions, she may transgress the boundaries within which truthfulness is required. Adversarial exchanges may involve an obligation to tell the truth, but no obligation to be entirely open. Parties to such exchanges know this, so they expect omissions and carefully crafted assertions. The obligation to be truthful is thus fine-tuned to cultural circumstances. Where and when evasions are permissible depend on aspects of the culture, including what is considered private, what exchanges are adversarial, and who deserves to know what. If Williams is right, then feeling that someone is sincere requires attunement to circumstances. We need to be sensitive to the sort of exchange we are in, the sort of information our interlocutor considers private, whether she considers our interest intrusive, and if so, whether she takes us to belong to a class whose members are entitled to such intrusive information.

Sincerity does not insure accuracy. Saying what one believes does not guarantee that one utters a truth. Being reflective, we monitor our truth seeking efforts in order to improve our accuracy. As we learn more about things we learn more about what methods of inquiry yield reliable results. We devise and refine our methodologies and control our belief formation accordingly. The methodologies are public and are institutionally grounded. But they are not just an additional layer of assessment perched atop antecedently formulated beliefs. For investigators who belong to an intellectual community committed to the values that its methodologies embody internalize those values and form their own beliefs accordingly. To some extent, this is because of the education they receive. They come to understand that the methodologies of their intellectual community are effective means for forming beliefs that are, and that the community will recognize as, accurate.

To trust an informant is to believe that her utterance is accurate and sincere. The utterance could be true by being a lucky guess. But when we trust an informant, it is not because we consider her a lucky guesser, but because we consider her a responsible epistemic agent. So the feeling of trust is keyed to norms of accuracy. We also think she is sincere. This, as we saw, involves a culturally variable sensitivity to matters of privacy. In trusting her, we take her to be respecting what we take, and take her to take, to be appropriate norms.

I went into this example in some detail in order to highlight exactly how cognitively complicated and evaluatively loaded emotions can be. Trust is an emotion. Because a feeling of trust can be experienced at an instant, we are apt to overlook how richly textured its conditions are, how much we had to learn and internalize in order to be in a position for the deliverance to be a deliverance of that emotion.

The same is true of perceptions. We see mountains and football games, microscopes and DVD players, ballots and Torah scrolls.[1] We also see sleazy actions and generous ones, brilliant manoeuvres and clumsy ones. Obviously we need to know a lot to be able to see such things. But we do know a lot, and given that we have the requisite knowledge, we directly and immediately see a vast array of different sorts of things. Emotions, I suggest, function similarly. In suitable circumstances, trustworthiness affords epistemic access to honesty and intellectual integrity. A speaker's honesty and intellectual integrity are evidence of her reliability. So feeling that an informant is trustworthy is evidence that she is reliable, that her report is true, and that, as she says, the Krebs cycle occurs in mitochondria. The epistemic creditability of an emotional deliverance thus figures in our grounds for believing a matter of scientific fact.

5. Salience

Epistemologists often proceed as though information is hard to come by. But actually our problem is the opposite. We are prey to massive information overload. Inputs flood our sense organs. Infinitely many obvious consequences follow from every belief. To know, understand, perceive or discern anything requires overlooking a lot. The question is: what should be overlooked?

Some aspects of a situation are salient, others are not. Salient aspects command attention, overshadowing other epistemically accessible factors. If a factor is salient it is, or at least presents itself as being, presumptively significant or relevant (Hookway 2000, 67–70). Emotions are sources of salience. They fix patterns of attention, highlighting certain features of a domain and obscuring others. Compare the following scenarios:

> Walking along a familiar street late at night, I suddenly hear footsteps approaching from behind. I am afraid. My fear does not just deliver the information that I feel I am in danger. It also orients me to my surroundings, highlighting aspects I otherwise overlook. I notice that aside from my pursuer, there is no one around, that the shops are all closed and barred, that no cars are on the road. I realize that the cobblestone sidewalk is uneven,

1 The example of the Torah Scrolls is Sydney Morgenbesser's.

> so it would be hard to run, that my high heeled shoes hobble me, that I am so out of shape that neither fight nor flight is an attractive option.

Alternatively:

> Walking along a familiar street late at night, I suddenly hear footsteps approaching from behind. Recognizing them as those of a friendly, local dog walker, I relax. Feeling serene, I notice the warm, gentle breeze, the slight scent of flowers, the attractive displays in the shop windows, the brightness of the moon, and the absence of crowds on the normally bustling street, with its charming cobblestone sidewalks. It is a lovely night for a leisurely stroll.

All of the individual facts adduced in both scenarios obtain. Under the sway of two emotions, I notice different ones. The effect of the one is not just to downplay the significance of the facts that the other makes salient. Although those facts are in principle epistemically accessible, some of them simply do not register. (If I think I am about to be mugged, I am hardly going to notice gentle breezes!) My emotion may do more than merely highlight discrete facts, like the scent of flowers or the absence of police officers. It may enable me to discern a pattern in what I would otherwise take as separate facts. The isolation, lack of light, uneven sidewalks and so on make the street dangerous for pedestrians late at night. By wearing shoes in which I can barely walk, rarely exercising, and strolling down a dangerous street, I put myself at risk. Even though the individual facts were readily known, the orientation the emotion supplies may be crucial to discerning the pattern they constitute (Elgin 1996, 149–56).

Emotions direct attention. They reveal certain aspects of a domain as worthy of notice. Fear is a frame of mind that prompts me to look at a situation in ways that would reveal evidence of danger and opportunities to avoid it. Childlike enthusiasm prompts me to look at the situation in terms of opportunities for fun. Such emotions may enable us to draw distinctions we otherwise would lack the resources to draw. For they supply focused incentives to refine our sensibilities. I have two students who regularly turn their work in late. One irritates me; the other worries me. If I disregard my emotions, their behaviour seems the same. But the fact that they occasion different emotions is evidence that what initially looks like one phenomenon is actually two. I dimly sense that the students are displaying different academic difficulties. My having different responses does not demonstrate that this is so. But it does provide grounds for suspicion. Hence it puts me in a position to seek further evidence and gives me an incentive to do so.

In this respect emotion is like belief. A belief is a propositional attitude, a propensity to take things to be as the belief content says that they are. Ordinarily, philosophers concentrate on the propositional element. They construe the belief that *p* as an attitude towards the proposition that *p*, a feeling that *p* is so. But a belief is not just an attitude towards a proposition. It is an attitude towards the world. The belief that it is raining orients me towards the world in such a way that I reach for an umbrella and cancel the picnic. I mentally and perhaps physically prepare myself to get splashed. I revise my intention to water the lawn. I look for a break in the clouds, listen to the raindrops on the roof, and so on. The belief is not just an

internal affirmation of a propositional content. It involves expectations, states of perceptual readiness, patterns of salience, and dispositions to utter, accept, reject, and investigate matters that I otherwise would have no inclination to consider. The belief that it is raining frames my interpretation of events. A wet spot on the floor intimates a leaky roof, not a spilt drink. Worries about flooding become prominent; worries about drought recede. Believing that *p*, is not just an affirmative orientation to the content that *p*. It is a complex orientation to those aspects of the world to which the question whether or not *p* is relevant.

By rendering previously ignored features and previously unknown patterns salient, emotions provide new insights into a domain. They enrich our cognitive capacities by sensitizing us to likenesses and differences, patterns and discrepancies that we would otherwise overlook. This is so whether the emotion is well founded or not. Even though fear of flying is unwarranted, people who suffer from it often end up knowing a lot more than other travellers about airline safety precautions and their limits. Their fear leads them to pay careful attention to information (about life vests and emergency exits) that other travellers routinely ignore and to seek out additional information (about metal fatigue and engine maintenance) that most travellers have no interest in. Even irrational emotions can be cognitively fruitful. But, one wants to say, insofar as the emotion is irrational, the insights it yields are slightly askew – not in the sense that they are not genuine insights, but in the sense that they do not bear on the subject's situation in the way that she thinks they do. The cognitive predicament is rather like that of someone who, seeking information about London, Ontario, inadvertently looks up information about London, England. Granted, she learns a lot, but not what she wants or needs to know. If an emotion is well founded, however, it properly attunes its subject to her situation. The factors it highlights are relevant and significant. The emotion then is not just a source of information, but a source of information that in the circumstances is worth having.

The more refined the sensibilities, the greater their yield. Consider a subject who can draw no distinctions along the scale of fears. Either he is afraid or he is not. If he is afraid, he is attuned to dangers in his domain and to opportunities to avoid them. If he is not afraid, he is largely insensitive to danger. His fear is undiscriminating. It sensitizes him to all potential dangers, regardless of their cause, their probability or their magnitude. When afraid, he is constantly on guard, for threats of any magnitude could come from any direction at any time. Another subject has the capacity to feel a range of fears from trepidation through terror. Terror, when he feels it, overrides all other concerns. He sees everything through the lens that the prospect of immanent, significant peril provides. But he rarely feels terror. Trepidation is more frequent. It, however, simply makes potential dangers slightly more salient than they otherwise would be. He is aware of them, but they colour but do not dominate his experience. For the most part, he can focus on other things. Clearly the second subject's cognitive situation is preferable to the first subject's.

To be cognitively well placed, it is not enough to have a large number of justified true beliefs. If we are lost in the woods, it would be beneficial to have justified true beliefs about how to get food and water, how to avoid predators and other dangers, how to find our way home. If my companion's overall doxastic system is richer than mine, because he knows a lot more than I do about French symbolist poetry, then

in our current predicament, he is no better off than I am. For his knowledge about poetry is irrelevant. Evidently, to be cognitively well situated, we do not just need justified true beliefs, we need relevant justified true beliefs. But even having an extensive array of relevant justified true beliefs may not be enough. If my extensive justified true beliefs about survival in the woods do not include beliefs about the prevalence of rattlesnakes, the evidence of their presence, and ways to recognize and avoid them, I may be in trouble. If my doxastic system does include such beliefs, but they are not salient, I may also be in trouble. Certain considerations should be salient in certain circumstances. And I am cognitively badly off if they are not.

The critical question then is what determines what should be salient. The answer, it seems, is that something should be salient in given circumstances, if one would want or should want to have epistemic access to it in those circumstances. 'Would want' is fairly easily explicated in dispositional terms. If you had epistemic access to *p*, would you consider your current epistemic situation for that reason improved? (Roughly, the things we would want epistemic access to are the things for which, were we later to find out about them, we would say, 'I wish I had known!') What we should want depends on the situation. Given that situation, would we be epistemically better off with respect to it if we had the information? An orientation or a frame of reference not only renders certain facts salient, but renders certain dimensions salient. So it can reveal worrisome gaps in our doxastic system. It can reveal what it is that we need to know. That is, if the frame is apt. By providing orientations, emotions are sources of salience. Emotions of experts are sources of what should be salient. For experts are apt to be properly tuned to the situation. So the emotions of the experienced, knowledgeable hiker – his fear, concern, complacency, and so on – are likely to cue him to relevant, significant features of the situation, providing him with epistemic access to useful information.

I have identified several ways in which emotions function cognitively. They provide epistemic access to emotional response-dependent properties. They provide epistemic access to other properties that provoke emotional responses. They fix frames of reference that render salient factors that the agent would otherwise overlook. Some of these factors may be ones that should be salient.

6. Expertise

Although I think this is all true, it is hardly enough. Ideally, we would like to do at least two more things: (1) identify when emotions are reliable indicators, and (2) improve their yield. In fact we can, and regularly do, do both.

We often recognize when people's emotional reactions are not to be trusted. When we appreciate that Joan is overwrought and that Sam is deliriously happy, we conclude that her verdicts are apt to be excessively negative and his excessively positive. We realize that we should take neither at face value. We also recognize expertise. We take a critic's enjoyment of a performance as much stronger evidence of its excellence than the enjoyment displayed by the soloist's doting mother. We consider the distrust exhibited by someone we regard as a good judge of character to be more telling than the distrust of the well-known cynic. If someone is an expert

in an area, it is likely that the dimensions along which he assesses are the right dimensions and that the assessments are reasonably accurate. An expert likes, admires, enjoys, fears, and worries about the right things in his area of expertise. He knows what deserves to be liked, admired, enjoyed, feared, and worried about and reacts accordingly. Our judgements about such matters are far from infallible. But they are sufficiently better than chance that rather than rejecting them out of hand, or trusting our luck, it is worth considering how we could improve their yield.

At least two strategies are effective. The first is to refine our sensibilities. By attending to and reflecting on our emotional responses, the situations that trigger them, and the orientations they give rise to, and by assessing the opinions they generate, we have resources for developing more nuanced and more accurate responses. As I learn that my fear of frogs is unfounded, I gradually cease to fear frogs. In consequence, my fear better attunes me to danger, since at least this one misattunement has been eliminated. When I learn to distinguish between regret and remorse, my emotions better track my sense of responsibility. To some extent, refinement of the sensibilities is a matter of learning to tell different feelings apart. To a greater extent it is a matter of learning to tell the orientations, deliverances and triggers apart.

Another way to improve the cognitive yield of emotions is through calibration. Some measuring instruments are more sensitive than others. A big change in the pointer position on one voltmeter is equivalent to a small change in the pointer position on another. To know how to read a meter requires knowing how sensitive it is. The same holds for emotional deliverances. A highly emotional subject will experience a big response to a stimulus that triggers only a slight change in a stolid subject. To read the magnitude of the stimulus off the emotional response requires knowing how sensitive the respondent is. This too is something we can discover.

Although we gain some understanding of things through knowledge of the emotional responses of others, much of what each of us gleans from emotions comes from her own case. Self-knowledge enables us to access the information our emotions embed. If we can identify our emotions, assess our level of expertise, and recognize how sensitive we are, we can profit cognitively from their deliverances. Reflective self-awareness pays epistemic dividends.

Suppose I am right. Having emotions affords some measure of epistemic access to the environment. Self-monitoring – attending to our responses and our responses to our responses, assessing appropriateness, learning to discern subtleties that align with circumstances, and so forth – enables us to increase their epistemic yield. Still, even in the interests of advancing knowledge, there are some emotions that we would go far out of our way to avoid. For the circumstances that trigger them are simply terrible.

We might simply concede the point. Experience is limited. If epistemic access to a domain requires experience of a sort you never have, you never gain that epistemic access to the domain. Some emotions are so painful that the benefits are not worth the cost. Some may be so painful that the pain swamps any benefits that could conceivably accrue. So perhaps there are available insights we should willingly forgo.

Instead of writing such emotions off completely, and forgoing the insights they embed, I suggest, we can gain access to them and their deliverances through the arts.

One reason we engage with works of art that elicit negative emotions is that they enable us to experience such emotions in muted forms and to explore, off line, the perspectives and insights they yield. If you are lucky, you will never directly experience the horror of coming to realize that you have actually murdered your father and married your mother. You can, however, experience an attenuated form of that horror through *Oedipus Rex*. Imagining your way into Oedipus' horror, adopting the perspective it provides, and seeing how the world looks from that perspective enriches your life. For you gain the ability to see and feel and discern and respond in ways you previously could not. If emotions afford epistemic access to things, and the arts refine, extend, heighten, and provide opportunities to experience emotions, the arts contribute significantly to cognition. And if we ask why we enjoy tragedies, horror films, and other art forms that elicit negative emotions, the answer is, at least in part, because we enjoy expanding and exercising our abilities, as engagement with such art forms enables us to do.

7. Conclusion

I have argued that emotion should be recognized as a facet of reason. Emotional deliverances are in some respects like perceptual deliverances and in some respects like beliefs. Like perceptual deliverances, they typically are responses to environmental triggers. If properly interpreted they provide information about the items that trigger them. Like beliefs, they fix patterns of attention, rendering salient and thereby affording epistemic access to factors that the agent might otherwise overlook. Like both beliefs and perceptions, emotions are educable. We can assess them for reasonableness and can gradually realign them if we find them unreasonable. And like both beliefs and perceptual deliverances, emotional deliverances are fallible. A deliverance, whatever its source, is tenable only if it is integrable into a system of cognitive commitments in reflective equilibrium. No matter how compelling an emotional response may seem, it is not epistemically acceptable on its own. But as a strand interwoven into a tenable system, it may strengthen, deepen, and modulate a person's understanding of herself and the world.

References

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.

Elgin, Catherine Z. 1996. *Considered Judgment*. Princeton: Princeton University Press.

Hookway, Christopher. 2000. 'Epistemic Norms and Theoretical Deliberation'. In: Dancy, Jonathan (ed.). *Normativity*. Oxford: Blackwell. 60–77.

Hume, David. [1739/40] 1967. *A Treatise of Human Nature*. Oxford: Clarendon Press.

Jackson, Frank. 1982. 'Epiphenomenal Qualia'. *Philosophical Quarterly* 32: 127–36.

Johnston, Mark. 1989. 'Dispositional Theories of Value'. *Proceedings of the Aristotelian Society, Supplementary Volume* 63: 139–74.

Murdoch, Iris. 1970. *The Sovereignty of the Good*. London: Routledge and Kegan Paul.

Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. Cambridge, Mass.: Harvard University Press.

Williams, Bernard. 2002. *Truth and Truthfulness*. Princeton: Princeton University Press.

This page intentionally left blank

Chapter Two

Epistemic Immediacy, Doubt and Anxiety: On a Role for Affective States in Epistemic Evaluation

Christopher Hookway

This chapter argues that emotions and other affective states have an important role in our practice of epistemic evaluation not least in enabling us to make sense of a kind of epistemic immediacy in the context of an epistemology that is largely externalist. Affective evaluations can enable us to understand both how reflection can be shallow and also how the limits of reflection cannot limit our rationality. After a general discussion of these issues, the paper explores how this applies to the case of rational 'real' or 'felt' doubt.

1. Introduction

This chapter is concerned with how far emotions and other affective states have a role in epistemic evaluation, in our ability to inquire effectively, to form beliefs rationally and to obtain knowledge. The position I want to defend involves three elements:

1. Affective states (sentiments and emotional responses) have an essential role in our practice of epistemic evaluation.
2. The rationality of our beliefs (and our epistemic rationality more generally) depends upon the rationality of (some of) our affective responses.
3. The regulation of our affective responses forms an important part of the regulation of our beliefs and inquiries.

Such views form part of a broadly pragmatist approach to issues in epistemology and, indeed, all three of the classical pragmatists (Peirce, James and Dewey) explicitly defended these three views. They can also contribute to our making sense of some important and controversial themes in contemporary epistemology. Since the use of the word 'pragmatism' can often stir misleading connotations, I should emphasize from the beginning that I do not take it to be immediately evident that pragmatism is a form of irrealism or anti-realism. The considerations I shall make use of may well be compatible with holding that the role of affective states in epistemic evaluation (1) contributes to the objectivity of our beliefs, and, indeed, (2) that affective evaluations guide us in obtaining beliefs that are *true*.

Some qualifications should be made here. It is very easy to find cases where our emotional responses can function as obstacles to cognitive success. Pride in your achievements can blind you to possible objections to your views; anger can distract the inquiry from being as careful as it should; jealousy of a rival can prevent our properly acknowledging his achievements; and so on. Moreover, we cannot reasonably aspire to the kind of control over our activities that puts us in a position to *guarantee* that they are not being subverted by our emotional responses. In reacting to the assumption, once common, that this is the *only* kind of impact that emotional states can have upon cognition, we cannot claim that their effects are always beneficial (see Goldie, this vol.). What we can claim is that, in many circumstances, emotional responses can make a positive contribution to our cognitive success, and that much of the time it is rational for us to *trust* our affective responses as a guide in epistemic matters. We need to be alert for signs that we have been misled, and we need confidence that we have been well trained, that, generally speaking our responses are reliable. But much of the time it is rational to hope that we are not being misled; and trusting these affective responses is often the only way to avoid taking on board an unreasonable burden of reflective self-questioning. We can be behaving irrationally while unable to realize that this is our position; but it does not follow from this that when our affective responses are not misleading us, our trust in them is irrational. Lack of criticism from others, growing self-knowledge, experience of cognitive success, together with other factors can reassure us (fallibly) that our trust is not misplaced.

The word 'emotion' can be a philosophical minefield, apparently unifying a range of disparate phenomena (Griffiths 1997). My concern here is with a variety of phenomena that are both evaluative and affective in their character. Many manifestations of what any theorist would agree were emotions would be included in this class of phenomena, as may a variety of manifestations to which the term 'emotional' does not apply very happily. My concern is with the role affective responses to our cognitive position can have, for good or ill (often for good) in directing our attention in dealing with epistemic matters and in explaining many of our cognitive successes.

Section 2 introduces some issues about epistemic immediacy and notes their relevance to some recent epistemological debates. After indicating some of the epistemic evaluations that may call for this kind of immediacy in section 3, I suggest (section 4) that giving a role to emotions and other affective states provides a promising way of making sense of this kind of epistemic immediacy. Section 5 returns to some examples – and in this context I turn to some characteristic claims about *doubt* from the pragmatist tradition.

2. Immediacy

When philosophers talk about 'immediate knowledge' or 'immediate' justification, they can have some very different things in mind. Two are particularly relevant to my concerns here. In describing knowledge as 'immediate' we are denying that it is 'mediated', that our knowledge somehow depends upon some 'intermediary'.

But there are different kinds of mediation, many kinds of intermediaries, and distinguishing these will lead to some different ways of understanding 'immediate knowledge'.

One way in which our knowledge can be 'mediated' is when it is the product of conscious reasoning: I can defend it by showing that it serves as the conclusion of a good argument whose premises I reasonably accept; or, I can defend it by offering *reasons* in its support. This suggests a way of understanding what it is for knowledge to be immediate: I am rational to hold it, but my appropriate confidence in my right to the proposition in question does not depend upon my being able to offer reasons in its defence or in my being able to offer a convincing argument for it. Describing the knowledge as *immediate*, in such cases, is partly a phenomenological claim or introspective report. But it is important for epistemology because it influences how the knowledge interacts with our practice of reasoning, of offering reasons and considering arguments. When knowledge is immediate, it is available for *use* in distinctive ways.

In the light of the above, we might be tempted to say that something serves as immediate knowledge when it is a 'premise which is not itself a conclusion' or when the propriety of its acceptance does not depend upon the mass of other knowledge that we possess. Our confidence in the proposition does not depend upon any other information at all. It is easy to see that this would be a mistake. In characterizing immediate knowledge above, I referred to *conscious* reasoning, and the fact that something serves as a first premise for conscious reasoning is compatible with its resulting from a process of unconscious reasoning (or other processes). And background knowledge may be indispensable to the justification a belief possesses, even if it is not available as a reason to be offered in defence of the belief (or as a premise in an argument which can be used to support it). When I ask myself why I should believe some proposition, there may be nothing for me to say – beyond, with Peirce, 'I cannot think otherwise'. It does not follow from this that my belief is not held in place by a mass of information that I already possess, but whose role in supporting my belief is not reflectively available to me.

So the notion of immediate justification that we shall be employing in this chapter is: A belief is immediately justified when its justification does not depend upon the believer being able to offer reasons or arguments in its support, and the believer would recognize no reason to provide such support.

And, as we have seen, this is compatible with the belief being held in place by other information we possess. This seems paradoxical only in the context of a *strongly* internalist epistemology which holds that information can only be even *relevant* to whether a belief is justified if it is available to the believer and usable by him as a tool for defending the belief. But it is likely that someone who holds to this sort of strong internalist position is likely to find it difficult to allow that there is any immediate knowledge at all. If immediate knowledge has to be knowledge which is not shaped or sustained by background information *at all*, it would have to be restricted to knowledge of the sort of bare sensory input that is often referred to as the *given*. And it is well-known that any philosophical position that defends immediate knowledge of the given is likely to be extremely problematic (Sellars 1963, 139ff).

The notion of *immediate knowledge* that I have articulated above is an externalist one. This is because there are considerations which are relevant to whether my belief is indeed immediately justified which have at least one of two significant properties: I cannot be reflectively aware of these considerations; and I may not be reflectively aware of their relevance to the judgements that I am making. One major theme of this chapter is that an externalist in epistemology *ought* to recognize that there is immediate knowledge and *ought* to understand it in the way described above. Unless such immediacy is available to the externalist epistemologist, he will find it difficult to explain how we can exercise reflective rational control over our beliefs and inquiries and be confident in our ability to do so. Self-control is an indispensable part of epistemic rationality, but for the externalist it must be legitimate that reflection can be *shallow*, that there can be limits to how self-control can work. Typically, we exercise self-control in the cognitive realm by raising questions about our beliefs: we can ask what reasons we have to hold them and, when we have identified those reasons, we can ask why we should accept them and why we should treat them as relevant to the belief we began by considering. Such self-control can seem problematic for an externalist: we can make demands for justification and question the justification we possess for any and all of our beliefs; but it is a simple consequence of externalism that we will not have the answers to all (or even many) of these questions. Reflection must be shallow because that explains why, in many cases, we do not always need to raise questions about what our justification is. We can be confident in our ability to control our cognitive activities only if the questions come to an end before the answers dry up. A promising way of explaining how we can recognize that there are limits to how far our reflections can extend without, as a result, feeling alienated from our cognitive achievements, is to allow that the judgements that are immediate have an *affective* character. When we make such judgements, they are accompanied by – or partly constituted by – an evaluation of them that has an affective or emotional character. This is the sort of view that I propose to defend below.

A nice example of a philosopher whose position has the sort of character that I have just described is Quine, in a passage I have used before when writing on this topic. Quine is writing about the norms that govern our inductive reasoning:

> The sifting of evidence would seem … to be a strangely passive affair, apart from the effort to intercept helpful stimuli: we just try to be as responsive as possible to the ensuing interplay of chain stimulations. What conscious policy does one follow, then, when not simply passive towards this inter-animation of sentences? Consciously the quest seems to be for the simplest story. Yet this supposed quality of simplicity is more easily sensed than described. Perhaps our vaunted sense of simplicity, or of likeliest explanation, is in many cases just a feeling of conviction attaching to the blind resultant of the interplay of chain stimulations in their various strengths. (Quine 1960, 19)

Although the topic is not mentioned, we can read this passage as an acknowledgement of the challenge to the Carnapian project of inductive logic found in Goodman's 'new riddle' of induction. That all observed emeralds are green (and are thus also grue) gives us a reason to believe that all emeralds are green but does not provide us with a reason to believe that all emeralds are grue. We sense ('feel')

that the 'green' hypothesis is simpler than the 'grue' one, and this reflects facts about the structure of our 'web of belief': but reflection does not make transparent to us just what factors we are sensitive to, and, according to Quine, we cannot identify formal principles of inductive reasoning that we are following. And the claim that the 'green' hypothesis is 'simpler' is best seen as an expression of the fact that we find it more plausible, and not as the identification of the deep fact that explains its plausibility. In view of our background knowledge and the evidence we have collected, the inference is plausible and compelling: we find it compelling and we are right to do so. But how and why this is the case is not something that is transparent to us; it is not something that we can bring to full reflective consciousness. If we want to describe the phenomenology of such an inference, of the sense that it is compelling, we might say that it possesses a kind of 'immediacy'.

If Quine is right about the phenomenology of inductive reasoning, the kind of immediacy that I have described has an important role in our epistemic practices and our inferential practices will be legitimate only if we are warranted in making judgements that are immediate in this sense. These are judgements which involve positive evaluations of propositions, which somehow reflect our background knowledge, but whose immediacy is manifested in the fact that we see no need to raise questions about why or how we make the judgements we do: we do not see the need to (and cannot) bring forward adequate reasons in defence of these propositions. Indeed, our knowledge of how our beliefs are formed and of why they are reasonable is very limited. When asked why 'green' is projectable, then we can only insist that 'there is nothing else for us to think': the fact that our knowledge here is *immediate* reflects the paucity of our epistemic self-knowledge. We talk of immediacy when our cognitive attitude is fully appropriate but where, perhaps, we lack access to how it was made and, thus, cannot offer reasons in its support.

Although Quine does not develop the point, the passage quoted itself suggests that these evaluations are affective. When we use an induction, its goodness is 'more easily sensed than described' and this may involve no more than 'a feeling of conviction attaching to the blind resultant of the interplay of chain stimulations in their various strengths.' Note some features of this:

1. We respond to a *feeling* of conviction.
2. Our response to evidence is 'strangely *passive*'.
3. Our passive response, this feeling of conviction, responds to a mass of information ('the interplay of chain stimulations in their various strengths') that is not available to our reasoning and reflection (the '*blind* resultant').

So we need to make sense of a kind of immediately justified belief that is required to make sense of how our beliefs and knowledge are employed in reasoning, but which is compatible with the fact that immediately justified beliefs may yet depend, for their justification, upon their relations to a mass of background information and unconscious processing.

This example complements our earlier remarks about the connections between this discussion of immediacy and some of the debates about internalism and externalism in epistemology. One kind of internalist argues that an item of information cannot

be relevant to whether beliefs are justified unless it (and its justificatory role) are accessible to the believer. On such a view, a belief can be immediately justified only if its justification does not depend upon anything that can be offered as a reason for accepting it. This sort of view emphasizes the fact that much epistemic evaluation is concerned with putting us into a position to reason well and to offer reasons to support our opinions: a theory of epistemic evaluation must show how we can exercise rational self-control over our beliefs and inquiries. And facts about the reliability of our belief-forming processes, or about unconscious cognitive processing, are not facts that we can *use* in exercising such self-control. And it follows from this that such facts have no role in justifying our beliefs, for example.

A plausible kind of externalism must still allow us to understand how (and how far) we can exercise rational self-control in these areas. Information we can *use* in reasoning must, of course, be accessible to us, and we must be confident in our right to make use of it. Our account of epistemic immediacy must thus be sensitive to the role of immediate knowledge in the exercise of rational self-control: we must give an account of such immediacy that shows how self-control is possible. This may require a variety of forms of immediate knowledge.

3. Epistemic Evaluations

Quine's discussion of induction is very valuable in drawing our attention to one kind of epistemic immediacy. But we must not lose sight of the fact that we make a wide range of immediate evaluations when we try to exercise rational self-control over our beliefs. Here are some examples:

1. Perceptual knowledge about our immediate surroundings and, perhaps, introspective knowledge about how our reasoning is progressing.
2. Knowledge of epistemic *relevance*: of whether something is a reason for some cognitive attitude.
3. Knowledge of the *strength* of reasons for belief and of evidence, for example, whether our reasons are sufficient to support our opinions.
4. Knowledge of the strength of reasons for doubt – for asking questions and trying to answer them.

In many of these cases, the information that is pertinent to relevance, strength etc. will not all be available to us, and nor will the standards to be employed in making judgements of strength and relevance. Yet it is important that we make the judgements and possess confidence in them when we do so. Unless we can possess such confidence, we cannot be sure of the reliability of our epistemic evaluations. I am suggesting that such judgements must possess a kind of immediacy, and the next task is to explore more fully what that involves.

When we engage in inquiry, reasoning and deliberation, one thing that occurs is that we *attend* to particular facts, propositions, issues and so on. Indeed, the success of our inquiries will depend upon our attending to the right things: we can go wrong if we fail to attend to relevant matters; and we can also go wrong if we attend to

matters that are irrelevant. For the moment, let us consider beliefs or items of putative knowledge: these invite attention of different kinds. First, we can attend to them when we use them as premises: if an item of information provides a reason for accepting some controversial proposition, then I should attend to it and use it as such a reason when I try to fix my beliefs concerning that contentious matter. If I want to know whether one of my colleagues is in her office, then I can use my information that someone told me that they had seen her leaving the building as a reason for thinking that she is not there. If I fail to attend to this information when it is relevant to my cognitive concerns, then my inquiries will go badly. But, second, I can also attend to them when I raise questions about their epistemic standing: I can question my memory of the conversation about my colleague's movements; or I can question its relevance – even if she was seen leaving the building, she would, by now, have had time to return. I attend to this information by using it as a premise or by asking questions about it. There are complexities here: I attend to the information both by doubting it and by wondering whether I should perhaps doubt it. With respect to these kinds of attention (and others) we can ask both: was I right to attend to the position when I did, and in the way that I did? And what reasons can I offer to defend the rightness of the ways I use my attention?

It is easy to see that I cannot *always* attend to, and question, all of the ways in which my attention is exercised and employed in the course of my deliberations. There are practical constraints: time and energy are limited. But the limits here are not just 'practical'. There are two related ways to support this. The first has already been alluded to: much of the cognitive processing that underlies reasoning and the direction of attention is unconscious, its details not being available to reflection. As Quine would put it, reflection is shallow: we simply *cannot* reflect on many matters that are relevant to the success of our deliberations.

The second appeals to epistemic holism. Suppose, once again, that I am eager to discover my colleague's whereabouts and somebody offers me some information as a reason for believing some proposition about where the colleague is. That this information does indeed serve as a reason depends upon a huge amount of complex background knowledge. First, I must judge that the testimony about my colleague's movements is *relevant* to the question whether she is in her office. Whether it is relevant depends upon a great mass of background knowledge. Second, I must make a judgement about whether, among the many facts that are relevant to her current whereabouts, this item is sufficiently prominent, sufficiently salient, to be offered as a reason for fixing my belief in one way rather than another. The scale of this task means that I must trust my judgement on this matter – I cannot list all possible candidates for relevance, evaluate each of them, then rank the facts that I do take to be relevant so that I can apply some rule to determine which is sufficiently prominent to be presented as a reason. The appeal to judgement is a recognition of a kind of immediacy here, even if I can point to some of the considerations that have shaped my judgement.

Some simple terminology can help us to sharpen the issue here. When we are evaluating our cognitive position, or carrying out some deliberation or inquiry, there are normative issues about *relevance*: Is this fact relevant to the acceptability of that proposition? Will it contribute to the success of our inquiry if we raise *these* questions

about the epistemic credentials of these opinions? Second, some judgements of the relevance of reasons and questions about them will be *salient* for us: we will immediately recognize them as matters we should take account of. We find ourselves disposed to attend to particular items of information in particular ways in particular contexts. And, as Quine might have put it, this sort of salience can be a strangely passive matter. Epistemic success requires that *salience tracks relevance*: what we find ourselves disposed to attend to should be what it is relevant that we attend to. The next task is to explain how this is possible.

4. Explaining Immediacy

We must now ask why affective states (including emotions) should be thought to be suited for a role in the regulation of inquiry. There is ample evidence that many of our deepest *evaluations* are manifested in emotional responses, and that emotional responses typically express evaluations. First, when I am angry with someone, I make a negative evaluation of their character or behaviour; and similar remarks could be made about responses of sympathy, pride, irritation and so on. There can be no doubt that emotional responses express evaluations. But, secondly, the grounds of evaluations that are expressed in emotional responses are often tacit and may not even be acknowledged by the agent. We may not be able to identify the considerations that make anger or sympathy appropriate, and we will very often not be able to express the standards or rules that make it the case that these considerations make the evaluation in question appropriate. Even if we are vaguely aware of the standards at issue, and even if we can reflectively find reasons for believing that the response is appropriate, there is no reason to suppose that we made use of an articulated awareness of these considerations in forming the response that we made. We can formulate our evaluative standards only as a result of a search for an explanation of our habits of evaluation. So, affective and emotional responses are appropriate vehicles for important evaluations for at least two reasons: emotional responses typically express evaluations; and they may be the only way of expressing evaluations when the standards that support them are not articulated explicitly in our reasoning.

Our epistemic evaluations have other roles that are more easily realized if some of these evaluations take the form of affective or emotional responses. An account of our practices of inquiry must explain how our evaluations and deliberations can *motivate* us to act. The standard Humean account of reasoning urges that beliefs cannot motivate us to act without the aid of appropriate desires: my belief that there is food in the kitchen will motivate me to enter that room only if I desire to eat. And even if I recognize that I have reason to eat, this may not motivate me to act if I lack an appropriate desire. It is an attractive feature of emotional or affective evaluations that they possess both belief-like and desire-like characteristics. My anger is a manifestation of my assurance that my friend has behaved badly. But it is also something that motivates me to act in ways that appear to be appropriate in the circumstances. These affective responses are both evaluations and sources of motivation. This feature will be particularly important when we examine states of *doubt* below.

This aspect of emotional evaluations also explains how our evaluations *spread*. My anger at somebody can lead to indignation at their actions and admiration of those who harm the object of my anger. Emotional evaluations can influence other evaluations in appropriate ways without the intervention of deliberation and reasoning. The system of emotional evaluations brings about a harmony in our affective evaluations that could not be produced (or not produced so easily) through reflection and reasoning.

To summarize: evaluations which are manifested in our emotional and affective responses can reflect unarticulated standards of evaluation, can explain how our evaluations lead us to act in accordance with them, and can explain how the influence of our evaluations can spread through our systems of beliefs and attitudes in appropriate ways, in spite of the fact that their influence is not subject to reflective rational control.

Developing a conjecture from section 2, we can now recognize that many epistemic evaluations possess just the same set of features that we have recognized as characteristic of emotional evaluations. First, of course, we make *evaluations* when we judge that beliefs or inferences are justified. In many areas (e.g. induction or inference to the best explanation), the standards that govern these evaluations standards are not *articulated*: as Quine has insisted, and as we noted in our discussion of self-control in section 2, epistemic reflection is generally very shallow and non-reflective. As Quine wrote, the goodness of an inductive argument (or an explanation) is more easily *felt* than articulated, we generally cannot *say* what makes such an argument a good one (Quine 1960, 23). And, as has already been noted, epistemic evaluations must have '*motivational*' force too: if I evaluate a proposition as poorly supported by evidence, I must be motivated (when appropriate) to reconsider other beliefs that depend upon it, or to search for more evidence when it becomes important for my inquiries. And epistemic evaluations have to *spread* too: we become suspicious of the further testimony of someone who has misled us once; and even if we frequently find it difficult to do so, we ought to question beliefs that have been inferred from others that have subsequently been doubted.

In the light of these parallels between the role of affective states such as emotions, on the one hand, and epistemic evaluations, on the other, why should anyone doubt the relevance of affective states such as emotions to epistemic evaluation? Some philosophers may argue that there is a significant difference, one which reflects some fundamental differences between theoretical reasoning and practical reasoning, theoretical inquiry and practical inquiry. First, the similarity: theoretical inquiry and practical deliberation are both problem solving activities, and the norms we follow in carrying them out are internal to these activities of solving problems – in the one case, problems of what is the case, in the other case, problems about what to do. One reason for taking emotions seriously in the practical case is that it could provide a (Humean) account of how we recognize our ethical and practical *ends*. Our fundamental ends, interests and values are reflected in our emotions and sentiments and so it is unsurprising that emotions have a role in determining our practical ends. In the epistemic case, the position seems to be different. Although we may seek a (practical) explanation of which questions interest us on any particular occasion, it can *seem* unnecessary to seek an emotive explanation of the standards we want

our conclusions to meet, of the fundamental norms which support our epistemic evaluations of beliefs and inquiries. We want our conclusions to be true; we want them to be justified in the light of the evidence: neither of these properties seems to be grounded in desires or emotions. This means that even if emotional responses serve as the *vehicles* of our evaluations, as the form taken by our evaluation of some belief as justified or as unjustified, they are irrelevant to the *content* and to the *authority* of these evaluations. As someone sympathetic to pragmatism, this does not convince me – nor am I convinced that the point would be decisive even if correct. However, I shall not pursue that here. Rather, in the next section, we shall consider a particular example of how emotional responses can have a role in regulating what it is rational to *doubt.* This discussion will be entirely consistent with the conclusions of the argument just described. The importance of affective evaluations is not primarily concerned with the evaluation of ends, even in the case of practical reasoning. The argument for a role for the affective in epistemic evaluation is consistent with recognizing that the aim of theoretical inquiry is to find the truth.

The role for emotions that I have identified here reflects Ronald de Sousa's claims about the functions that emotions can serve in *The Rationality of Emotion* (de Sousa 1987). The holistic character of decision and of belief formation point to what he calls 'the strategic insufficiency of reason': if our activities are directed wholly by reflective deliberation, there will always be questions that arise but which we lack the time or ability to answer. Searches for relevant information can always continue. We can always ask more questions about how well we have conducted our search. De Sousa claims that the 'function of emotions is to fill gaps left by (mere wanting plus) "pure reason" in the determination of action and belief, by mimicking the encapsulation of perception' (de Sousa 1987, 195):

> For a variable but always limited time, an emotion limits the range of information that the organism will take into account, the inferences actually drawn from a potential infinity, and the set of live options from among which it will choose. (de Sousa 1987, 195)

They can do this because emotions provide 'determinate patterns of salience among objects of attention, lines of inquiry, and inferential strategies. Emotions can serve rationality by: dealing with the insufficiencies of (conscious deliberative) reason by controlling salience' (de Sousa 1987, 201). So long as our emotional responses are appropriate, they control salience in ways that serve our purposes. They make a positive contribution to rationality.

5. Doubt and Anxiety

In this section, we should consider one particular epistemic evaluation that lends itself to the sort of explanation that I have just considered. I shall begin with two historical remarks. When Peirce discusses the Cartesian tradition in philosophy, he is very critical of the method of doubt. In 1868, he defends the claim that we cannot begin with complete doubt by saying:

- The assumptions with which we begin philosophy 'are not to be dispelled by a maxim, for they are things which it does not occur to us *can* be questioned'.
- This 'initial scepticism will be a mere self-deception and not a real doubt'.
- 'No one who follows the Cartesian method will ever be satisfied until he has formally recovered all those beliefs which in form he has given up.'
- 'Let us not pretend to doubt in philosophy what we do not doubt in our hearts.' (Peirce 1992, 28–9)

A 'real' doubt is one that can genuinely motivate us to inquiry – a pretend doubt may appear to motivate us to inquire but, as the third of these remarks indicates, does not really motivate us to serious inquiry. What counts is what we doubt *in our hearts*. Ten years later, in the 'Fixation of belief', he returns to this point about motivation: '[The] mere putting of a proposition into the interrogative does not stimulate the mind to any struggle after belief. There must be a real and living doubt, and without this all discussion is idle' (Peirce 1992, 115).

Doubt must be felt ('in the heart'), it must be real, it must motivate us to inquiry, and, as he says elsewhere, it must be rational: doubt requires a reason. We may think we have a reason for doubting some proposition, but the fact that we are not motivated to inquiry – to a 'struggle after belief' – indicates that we deceive ourselves when we say that we are aware of a reason for doubting the proposition. (Presumably there is no guarantee that there could not be people who experienced a genuine doubt on Cartesian grounds but, prima facie, it is reasonable to assume that the unarticulated standards of evaluation that inform our sense of doubt carry authority.)

It is fundamental to Peirce's position that a proper understanding of doubt and its epistemic importance must take seriously the idea that *real doubt*, the only kind of doubt that we need to take seriously, presents a question or proposition as requiring our attention or consideration. If doubts lack this kind of salience, then it is reasonable for us to work on the assumption that they do not demand further inquiry to settle them. This sort of motivational force that is part of real doubt must be explained if we are to understand how we can inquire effectively without being constantly side-tracked by questions that do not require our attention.

The seventeenth century common-sense philosopher John Wilkins employed a similar understanding of doubt in order to criticize scepticism thus: '*Doubt* is a kind of *fear* … and 'tis the same kind of Madness for Man to *doubt* of any thing, as to *hope for* or *fear* it, upon a mere possibility' (Wilkins [1675], 25–6). In coming to doubt some proposition, I evaluate that belief as 'epistemically unsafe', as something it would be rash to believe on the basis of the reasons I currently possess. Wilkins recognizes that this evaluation takes the form of an emotional reaction: when I become anxious about the uses I make of some item of information, then I am motivated to reassess its credentials and look for further support before relying upon it in future. I become anxious about my belief. He also suggests that the appropriateness of emotional responses (of fear or hope) provides a suitable model for thinking about the rationality or appropriateness of doubt. Doubt is a form of epistemic anxiety, and rational doubt is reasonable anxiety. Those who take sceptical doubts too seriously are accused of timidity. Both Peirce and Wilkins thus attach epistemic importance to 'real' or 'felt' doubt, over the more speculative claims of philosophers about what

ought to be doubted. So long as we are rational and, presumably, well educated, our dispositions to doubt are likely to be reliable.

These philosophers think that when we feel doubt of some proposition, hypothesis, method, strategy, argument or whatever, we are anxious about relying on it. The anxiety (the real felt doubt) reflects the possession of standards of evaluation that we may not be able to articulate. And the anxiety can motivate us to inquire further, seeking the source of the anxiety, evaluating its appropriateness, or acting to revise our opinions so it does not arise any more.

The case of doubt shares the features of emotional evaluations we have noticed above (see Hookway 1998). First, as we have seen, doubt of a proposition is an evaluation of it. Second, doubt often has a holistic character: although I may be able to point to the new information that led me to doubt a favoured belief, the doubt is usually only assessable as appropriate or otherwise by reference to the totality of information that I possess. It is unusual for me to be able to point to the principles and background assumptions that led me to be pushed over into doubt at just this point. Confidence in my doubt rests upon trust that any of my information that did not come to consciousness during the formation of my doubt was not relevant to the evaluation in question. As our quotation from Quine suggests, the doubt reflects weightings attached to a large number of items of information employing standards that (often) I cannot articulate. And thirdly, the doubt has a 'motivational' force: real doubts motivate me to collect more evidence, to reflect upon the credentials of the erstwhile belief, to worry about other beliefs which depend upon this one and so on. Philosophical doubts generally lack this motivational force. We might also think that doubt can have *felt quality* – a phenomenology – analogous to some other strong emotions. There is a 'gut reaction' – a 'felt doubt'.

In a more recent treatment of these issues, Paul Thagard (2004) has exploited the distinction between hot and cold cognition in thinking about doubt. We might define cold doubt by saying that someone doubts a proposition if they disbelieve it or suspend judgement in it (Salmon 1995). And we might then define reasonable doubt as doubt for which we have sufficient reason. Hot doubt involves a relation to a proposition that is 'emotional as well as cognitive', and Thagard explains it as a kind of 'emotional incoherence'. This is in line with the views of both Peirce (who he appeals to as a forerunner of his position) and Wilkins, and a central contention of the present chapter can be expressed by saying that a satisfactory account of epistemic rationality must give a central role to 'hot cognition', which 'involves emotions tied in with personal goals and motivations' (Thagard 2004).

This discussion does, of course, draw attention to an issue that was mentioned earlier. That there be this kind of role for our habits of feeling in determining what doubts we can take seriously does not provide an immediate response to scepticism. A philosopher who is impressed by the power of sceptical arguments may respond that the fact that such arguments produce no real epistemic anxiety should be explained by saying that our emotional and motivational capacities here function, not as aspects of epistemic rationality and reasonableness but rather as devices that help us to live with our dire epistemic position. The fact that we *feel* no doubt does not show that we should not do so. Experimental evidence shows how our habits of reasoning and inquiry are often irrational, often because of the impact of emotional

responses upon our evaluations of our beliefs. Perhaps philosophical reflections about sceptical arguments provide a more clear-eyed perspective upon our epistemic position than our hot cognitions will enable us to take seriously.

The fact that sometimes, in special circumstances, our emotional evaluations can lead us in the wrong direction is an indication of our fallibility but in itself provides no reason for taking sceptical arguments seriously. We can learn to take account of our frailties, and, even if error and irrationality cannot be avoided in all circumstances, we can still be confident of many of our cognitive successes. Both Peirce and Wilkins (and even Descartes) would agree that empirical evidence of error in special kinds of circumstances provides no reason for treating all of our beliefs as requiring further scrutiny. The fact that some people – obsessives for example – are unable to function cognitively because of an excessive capacity for epistemic anxiety does nothing to show that the rest of us are not in a much better and largely satisfactory epistemic position.

6. Conclusion

For several decades, debates about the merits of internalist and externalist approaches to epistemology have dominated epistemic debate. Internalists hold that all of the information relevant to whether our beliefs and doubts are reasonable should be available to the believer; our epistemic position should be evident to careful first-person reflection. Externalists deny this, holding that facts to which we have no introspective access may be relevant to the evaluation of our epistemic position. In concluding the chapter, I want to place several of the issues discussed within this perspective.

As we have seen, philosophers influenced by the pragmatist tradition often emphasize the difference between 'real doubts' and 'paper doubts'. The former, unlike the latter, are rational or reasonable; and the considerations that lead us to entertain these different dates are respectively, good *reasons for doubt* and apparent reasons that are not reasons for doubt at all. If we spend too much time thinking about 'paper doubts', we waste time on inquiries that serve no useful purpose. It is then hoped that doubts grounded in sceptical possibilities about whether we are deceived by an evil demon or brains in vats being deceived by wicked scientists, can be ignored because they are paper doubts for which we have no good reasons. On the other hand, when we entertain doubts for which we have reasons (*real doubts*) doing so is likely to contribute to the success of our inquiries and other projects. If we are to inquire well, we need to be sensitive to which doubts are real, and we need to be able to evaluate the strength of reasons for doubt. Some pragmatists even treat reasons for doubt as more important than reasons for belief. In many cases, we cannot give specific reasons for our most confident beliefs, and we see no need to do so. Reasons for belief matter when we form new beliefs, and when one of our beliefs faces a reasonable doubt which must be defeated (e.g. Levi 1998).

Real doubts are ones we should attend to, ones we should think about when deliberating and inquiring. Success in inquiry depends upon attending to the right worries: raising the right questions, responding to the right doubts. If we do not

reflect at all about the legitimacy of our beliefs and methods of inquiry, then we cannot be held responsible for the strength of our epistemic position. This, we might think, is the truth in epistemic internalism, in the idea that we must be aware of considerations which support or challenge our beliefs. On the other hand, if our reflections are not limited, if, obsessively, we consider every possible doubt and demand reasons for every belief we employ in our reasoning, then we will be unable to exploit all the knowledge that is embodied in habits and emotional responses. Here we encounter the strength in externalism, the idea that the fact that our methods are reliable, even if we do not understand how or why they are reliable, can be enough. An adequate account of epistemic evaluation must incorporate both strengths and provide an account of how it can be that we know, often unreflectively, just how reflective we need to be. We need to be correctly confident that we are sensitive to the limits of rational reflection.

The fundamental thesis of this paper is that affective and emotional evaluations can help us to understand how there can be limits to how reflective we need to be. Our epistemic anxieties, for example, suggest to us when doubts need to be taken seriously. They help to draw the boundaries within which the burdens of reflection emphasized by internalist theories are reasonable. Of course, there is no guarantee that the habits of evaluation held by an individual person at a particular time will draw these boundaries in the right place. Some people are careless in their reasoning and others are obsessive. And we need to be alert for clues for when our own dispositions may be flawed. But, so long as we are well educated, alert for signs of irrationality, and successful in our inquiries, it is rational to trust our dispositions, hoping that they will not lead us wrong. Here lies the truth in externalism.

References

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.

Griffiths, Paul E. 1997. *What Emotions Really Are. The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Hookway, Christopher. 1998. 'Doubt. Affective States and the Regulation of Inquiry'. In: Misak, Cheryl (ed.). *Pragmatism*. *Canadian Journal of Philosophy,* suppl. vol. 24: 203–25. [Reprinted in: Hookway, Christopher. 2000. *Truth, Rationality, and Pragmatism*. *Themes from Peirce.* Oxford: Oxford University Press. 246–64.]

Levi, Isaac. 1998. 'Pragmatism and Change of View'. In: Misak, Cheryl (ed.). *Pragmatism*. *Canadian Journal of Philosophy,* suppl. vol. 24: 177–201.

Peirce, Charles S. 1992. *The Essential Peirce.* Vol. 1. Ed. Nathan Houser and Christian Kloesel. Indianapolis: Indiana University Press.

Quine, Willard V. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.

Salmon, Nathan. 1995. 'Being of Two Minds. Belief with Doubt'. *Noûs* 29: 1–20.

Sellars, Wilfrid. 1963. *Science, Perception and Reality*. London: Routledge and Kegan Paul.

Thagard, Paul. 2004. 'What Is Doubt and When Is It Reasonable?' In: Ezcurdia, Maite; Stainton, Robert; Viger, Christopher (eds). *New Essays in the Philosophy*

of Language and Mind. *Canadian Journal of Philosophy,* suppl. vol. 30: 391–406.

Wilkins, John. [1675] 1715. *Of the Principles and Duties of Natural Religion*. 7th edition. London: R. Bonwicke.

This page intentionally left blank

Chapter Three

Virtues, Emotions and Fallibilism

Alessandra Tanesini*

In this chapter I develop an account of fallibilism in terms of intellectual virtues and related emotional states. After a very brief review of the failure of various attempts to define fallibilism in terms of the ever-present possibility of error, I present my alternative account. This account constitutes a development of the view of emotions that I developed in my paper 'Emotions and Rationality' co-authored with Mark Lance (published in a special issue of *Canadian Journal of Philosophy*). In that paper, we argued that emotions are salience generators which play a role in the selection of which lines of inference are to be pursued in a given context. In the present paper I develop the idea, briefly discussed at the end of that paper, that fallibilism is to be understood in terms of various intellectual virtues. These involve the presence of emotional states which are salience generators directed towards the people with whom we engage in dialogue and their opinions.

> For years … I used for myself to collect my ideas under the designation fallibilism; and indeed the first step toward finding out is to acknowledge you do not satisfactorily know already; so that no blight can so surely arrest all intellectual growth as the blight of cocksureness; and ninety-nine out of every hundred good heads are reduced to impotence by that malady – of whose inroads they are most strangely unaware!
>
> Indeed, out of a contrite fallibilism, combined with a high faith in the reality of knowledge, and an intense desire to find things out, all my philosophy has always seemed to me to grow … (Peirce 1960, 13–14)[1]

The main purpose of this chapter is to develop an account of fallibilism as the intellectual virtue of epistemic humility. Further, I argue that this virtue, like other intellectual virtues, plays its distinctive role in the regulation of cognitive conduct at least partly in virtue of possessing an emotional component whose distinctive epistemic role is to generate saliences.

The chapter consists of three sections. In the first I discuss some desiderata that must be met by any satisfactory account of fallibilism and I review the failures of various attempts to provide accounts of this notion in terms of the ever-present possibility of error. In the second section I defend the view that emotions play

* An earlier version of this paper was presented to the Philosophy Department at the University of Hull. I would like to thank Richard Gray, Michael Durrant, the editors of this volume and the audience in Hull for several very useful comments and queries.

1 I would like to thank the editors of this volume for reminding me that Peirce himself thought of fallibilism as an intellectual virtue.

the epistemic role of salience generators in rational cognitive conduct and sketch an account of intellectual virtues as involving emotional components. In the final section, I discuss the specific virtue of epistemic humility and argue that an account that identifies fallibilism with this virtue meets all the desiderata specified in the first section.

1. Fallibilism

These days we are all, or nearly all, fallibilists. Michael Williams even makes fallibilism the trademark of the modern conception of knowledge (Williams 2001, 41). Yet, fallibilism is remarkably hard to define. We owe the first formulations of this position to C.S. Peirce who thought of it as a feature of scientific method. It is now generally understood as an expression of the underlying idea that human beings are prone to errors when engaging in the pursuit of knowledge. Fallibilism is also often contrasted with dogmatism and presented as a certain openness to the possibility that the issues under consideration have not been settled once and for all, so that even though our views are very firmly held we concede the possibility that we might in future see that we need rationally to revise them. Fallibilism thus seems to imply an alertness to the possibility that our views might be mistaken.

It is therefore not surprising that early accounts of fallibilism focused on this last aspect of our intuitive conception and defined fallible knowledge and belief in terms of the possibility of error. Thus, one might hold that:

(1) For any proposition p, S's knowledge that p is fallible if and only if S's belief that p could have been false.

It is now widely agreed that this account is inadequate, since it has the immediate consequence that no knowledge of necessary truths is fallible.[2] After all, if *p* is necessarily true, belief in it can never be false. This consequence is seen as undermining the account because our knowledge of arithmetical truths is fallible.

Another common account of fallibilism also falters at the same hurdle. It is often heard that fallible knowledge is knowledge without certainty, or more precisely, it is knowledge which is based on evidence that does not entail or guarantee the truth of the belief. This conception can be formulated thus:

(2) For any propositions p and q, S's knowledge that p on the basis of evidence q is fallible if and only if q does not entail that S's belief that p is true.

This definition of fallibilism is initially attractive because it establishes a connection between fallibilism and ampliative inference (i.e., an inference to a conclusion that contains information which is not already present in the premises). There is

2 Traditional definitions of infallibility suffer from the same problem. Thus, *S*'s knowledge that *p* is said to be infallible if and only if *S*'s belief that *p* could not have been false. This definition has the implausible consequence that all knowledge of necessary truths is infallible.

some plausibility to the idea that fallible knowledge is knowledge acquired by fallible methods of reasoning. The idea is naturally supplemented by the thought that a method of reasoning is fallible if it can lead from truth to falsity. Ampliative inferences are paradigmatic in this regard since neither in induction nor in inference to the best explanation is the truth of the conclusion guaranteed by the truth of the premises.

Appealing as it might be, this definition also has the infallibility of knowledge of necessary truths as an immediate consequence since if *p* is a necessary truth, belief in it is also necessarily true, and, as is well-known, a necessary truth is entailed by anything whatsoever. Thus, we never have knowledge of necessary truths based on evidence that fails to guarantee the truth of the belief in question.[3]

These difficulties are widely acknowledged. Various attempts have been made to rescue the situation. Thus, for instance, it has been proposed that the modalities involved in the formulation of fallibilism should be read epistemically. *S*'s knowledge that *p* would be fallible if and only if, for all we know (i.e., possibly), not-*p* is true. But this idea falters when considering the case of known truths. It is not true that as far as *S* knows, or as far as those who attribute knowledge to *S* know, not-*p* is true, because both *S* and the attributors know that *p* is true and thus know that not-*p* is false.

At first sight, the attempt to make fallibilism a second-order issue appears more promising. Thus, one might define fallible knowledge as follows:

(3) For any proposition p, S's knowledge that p is fallible if and only if S knows that p and S does not know that S's belief that p is not false.[4]

Despite its initial appeal and its ability to address the problem of fallible knowledge of necessary truths, this account has some quite obvious problems. First, it entails the implausible thesis that only creatures capable of having second-order beliefs can have fallible knowledge. Second, it would seem perfectly possible for *S* fallibly to know that *p* and also to know, equally fallibly, that *S*'s belief that *p* is not mistaken. And yet this possibility is unwarrantedly ruled out by this formulation.[5]

More recently Baron Reed has proposed an account of fallible knowledge as knowledge that could fail to be knowledge. Thus,

(4) For any proposition p, S's knowledge that p is fallible if and only if S knows that p on the basis of justification j and yet S's belief that p on the basis of j could have been either false or accidentally true (Reed 2002, 150).

Reed's idea is that knowledge of a proposition is fallible if and only if that proposition is believed and the belief is justified, but the belief could have failed to be knowledge

3 This problem could be overcome by abandoning classical logic and adopting a logic, such as relevance logic, in which necessary truths are not entailed by anything whatsoever. I shall not consider this option in this paper.

4 This definition has been proposed by L.S. Carrier (1993).

5 For these criticisms of the view see Baron Reed (2002, 148).

either because the proposition is not true or because although the proposition is true, its truth is accidental since its justification could have been unrelated to its truth.

Reed's proposal does not fare much better than the earlier proposals in providing an account of how fallible knowledge of necessary truths is possible. Suppose Mary knows some mathematical theorem on the basis of a proof. Intuitively, we wish to include such cases within the domain of fallibility. And yet, they are not captured by Reed's definition. Patently, Mary's belief about the mathematical truth could not be false. Hence, it could count as an instance of fallible knowledge only if its justification could have been unrelated to its truth. In the kind of case under consideration, this Gettier-like failure is impossible. Since Mary's justification is a proof for the theorem, that justification could not fail to guarantee the truth of the theorem. So given Reed's definition Mary's knowledge is infallible. But this conclusion contradicts the intuition that even knowledge of necessary truths on the basis of proofs that guarantee their truth is fallible since, although no errors were actually made in the proof, one could have failed to provide a genuine proof.[6]

I have presented a review of this history of failures to provide an adequate account of the notion of fallible knowledge in order to motivate a change in orientation. The definitions discussed so far take fallibility to be a feature of epistemic states such as knowledge. Instead, I propose that fallibilism is primarily a feature of epistemic agents, and only derivatively of their states and of their methods of inquiry. This proposal gains some initial plausibility from the fact that fallibilism is the opposite of dogmatism, and the latter is typically thought primarily as a characteristic of agents. I shall defend this view in the final section of the chapter, but for now I want to provide some criteria that I take to be fairly uncontroversial, and which must be met by any satisfactory account of fallibilism.

First, fallibilism is related to the idea of the possibility of mistakes. A fallible knower is an agent that is liable to making errors. This much is uncontroversial. What might be taken to be more controversial is the second criterion which I endorse, but do not defend here, that the whole of human knowledge is fallible. In particular, our knowledge of necessary truths is fallible even when that knowledge is based on proofs. I take global fallibilism to be an immediate consequence of the view that human beings can be mistaken about absolutely anything that can be put forward in the form of an assertion.

The third criterion is that fallibilism must not preclude the possibility of knowledge. Thus, it must not entail the truth of scepticism. Obviously, a discussion of scepticism is beyond the scope of this chapter. The point of this criterion is simply that the sceptic should not be able to claim an automatic success as a consequence of our endorsement of fallibilism.

The fourth criterion opposes fallibilism to dogmatism understood, among other things, as an unwillingness to entertain the possibility that one might have made a

6 My intuition that our knowledge of necessary truths on the basis of a proof is also fallible might be a result of my commitment to the idea that fallibilism is global. This is an idea that I do not defend in this chapter. Nevertheless, if Reed intends to claim infallibility for this kind of knowledge, the onus is on him to show why knowledge of necessary truths on the basis of proofs deserves a special treatment.

mistake or that one's views might be false. Thus, the dogmatist is simply not prepared to reconsider his or her position in light of criticism. The fallibilist, instead, and this is the final criterion, is the agent that is open to the possibility that his or her position might need to be revised in light of relevant criticism.

To summarize, I propose that the following criteria must be met by any satisfactory account of what fallibilism might be:

1. It must relate fallibilism to the possibility of mistakes.
2. It must apply globally to human knowledge.
3. It must not rule out the possibility of human knowledge.
4. It must oppose fallibilism to dogmatism.
5. It must relate fallibility to the possibility of rational revision.

I have also hinted that it might be useful to think of fallibilism primarily as a character trait of agents, and only secondarily as a feature of the methods of inquiry adopted by agents and of the epistemic states of agents that engage in such inquiries. Part of the motivation for this suggestion lies in considerations pertaining to the role played by emotional states in the cognitive economy of rational agents.

2. Rationality, Emotions and Virtues

I have argued elsewhere that emotion-like states play an essential and distinctive role in rational conduct (Lance and Tanesini 2004). The argument is based on some observations about the character of rationality and the nature of emotions. Firstly, it is important to note that rationality does not merely concern the ability to draw inferences. The rational agent must also be able to select which, among infinitely many, lines of reasoning are worth pursuing. Any proposition is entailed by and entails infinitely many others. This fact does not have a paralysing effect because we do not even consider most of these possible inferences and focus only on a few.

The rational agent must be able to discriminate some, among all the possible inferences, as salient in his or her current situation. An agent that has excellent inferential abilities but cannot select among valid inferences those which are promising is a poor thinker. She would not be able to see that some lines of inquiry are more likely than others to produce a fruitful outcome; she would not be able to focus on any task since she would be busy considering all sorts of possibilities and their consequences, which to the trained mind clearly appear as irrelevant.

The point is obvious, although, perhaps, insufficiently noticed. It can be brought home by means of a couple of examples. Anybody who has ever tried to solve a problem of any kind knows the importance of not wasting time and effort considering some pretty unlikely possibilities. For instance, we would think that something is amiss if a philosopher, familiar with a traditional notion of species that defines a species as a group of organisms that can successfully inter-breed and knowing that horses and asses are normally taken to belong to different biological species although they are capable of inter-breeding, tried to solve the tension between these claims by considering first of all the possibility that classical logic might be wrong

and some contradictions true.[7] Whatever we might think of paraconsistent logic, we do not think of examples such as this one as offering support in favour of this position. It is just clear to us that this line of thought is simply not worth pursuing in the case at hand.

The importance of knowing what to investigate and what to ignore is apparent even in that most calculative pursuit which is the game of chess. Good chess players know where to look on the board, they are capable of picking out salient features of the game they are playing. Competent players do not begin by calculating the outcomes of every possible move and select that which is best. On the contrary, they ignore many possibilities and they focus on perceived threats, dangers and opportunities. It is at that stage, and only at that stage, that competent players engage in calculation.[8]

The second observation is perhaps less obvious: the states involved in our ability to select promising inferences cannot all be propositional attitudes such as beliefs. The role of these states is to present some avenues of thought as important; thus, these states are best characterized as salience generators. Their role is to pick out some features of the situation or problem as salient and thus to select some inferences as worth making. States that have propositional contents are themselves fodder for more inferences, rather than functioning as a mechanism for selecting some inferences over others.

It might be objected that the role of some propositional attitude, typically a belief, may be to help us select some among a group of possible inferences. In other words it is possible to form a belief about what is salient in a given situation and then reason inferentially about which inferences to focus one's attention on. Thus, for instance, the competent chess player could form a belief whose content is that a given number of features of the current position are relevant and use that belief to select various lines of calculation as worthy of pursuit. I do not wish to deny that something of this sort can at times be done. What I wish to claim instead is twofold. My first contention is that not all salience generators can be of this propositional kind. My second contention is that phenomenologically speaking human beings rarely, if ever, proceed in this manner.

A reason why not all the states that function as salience generators could have propositional contents is that the procedure sketched above requires the formation of beliefs whose contents specify the class of propositions that happen to be salient in the given situation. Yet, given the often gerrymandered nature of salient features in a given situation, it is hard to see how we could come to form the belief with the relevant content unless something of a different nature has already focused our attention on that class as relevant in the current circumstances.

For instance, the competent chess player is able to focus on a few possible moves and ignore at any point in the game most of the legitimate moves available to her.

---

7 The actual biological definition requires that the organisms can inter-breed and that the offspring is fertile.

8 Not even computers can play chess exclusively by means of sheer calculative brute force. Basic physical constraints combined with magnitude of the number of possible distinct continuations of the game make such successful calculations impossible.

The moves she focuses on, if she is a good player, are those which offer promising defensive and attacking possibilities. The objection to my view I am considering holds that the player selects these possible moves for consideration in virtue of holding the belief that these are the only promising moves given the current state of the game. But whilst we can make sense of the idea of forming a belief such as this one by considering all the possible moves, I have already shown that sheer calculative force cannot be used to play chess. This putative belief then would be formed in a different way and not as a result of exhaustive inferential reasoning. The most plausible candidate is something, other than a belief, that focuses the attention of the agent on a specific set of moves as having something in common: namely, they are promising in the context of the current game. In this particular example, it is the ability to see some configurations on board as threatening and others as exciting that guides the player's calculations.

This consideration generalizes. When reasoning about any topic whatsoever, inferential thinking and the formation of various beliefs are typically preceded by a process of selection of some alternatives as relevant to the issue at stake. More specifically, when faced with an infinite number of inferential alternatives among which they need to select a few, thinkers rely on pre-doxastic states or mechanisms that attract them to some connections and lead them away from others.[9]

The idea that salience generators are pre-doxastic is also supported by phenomenological observations. In ordinary life and when engaging in a variety of intellectual pursuits we do not proceed by sheer inferential force. Quite the contrary, often what distinguishes a good from a mediocre practitioner is precisely the ability to select the options that are worth considering, focus on the right opportunities, be attuned to genuine risks and ignore what is irrelevant. Competent practitioners have a sense, a hunch, about what is worth exploring and what is not which they exercise before they make inferences and develop a view about the issue at hand.

The observation that salience generators function pre-doxastically should not be taken to imply that creatures that are not capable of making inferences and forming beliefs are nevertheless capable of detecting saliences in the relevant way. In typical cases, it is only the individual who has acquired a lot of propositional knowledge about the area of investigation under consideration who will be capable of having good hunches. It is only by playing many games and studying others that a chess player acquires the ability to perceive threats and see opportunities on the board. However, once the player has acquired the relevant background knowledge her perception of a threat is not based on inferential reasoning but precedes it and is clearly non-inferential in nature.

It might be objected that their pre-doxastic nature makes salience generators irrelevant to epistemology. The traditional distinction between context of justification and context of discovery could also be invoked to support the objection. The idea is that salience generators play a role in the context of discovery since they figure among the factors that lead to the formulation of hypotheses, but have no role in the

9 I call these states pre-doxastic because they often operate before beliefs are formed. They are also non-doxastic because they cannot always be identified with, or reduced to, beliefs or combinations of beliefs.

justification of the hypotheses once they have been formulated. The objection is not, in my opinion, particularly worrying. In the first instance, both the exclusive focus on the context of justification and belief in a clear-cut distinction between the two contexts has been largely abandoned by philosophers of science. But, secondly, and more importantly, any epistemologist who is willing to grant that we have knowledge and justification that falls short of certainty must grant that our justification in holding a belief *p* is dependent upon our entitlement to ignore (as irrelevant) some *q* which is incompatible with *p*. Thus, for example, I am justified in believing that I saw a zebra in the zoo, even though I did not even consider the possibility that a practical joker may have put a painted mule in the zebra enclosure. Justification, therefore, is dependent upon a selection of relevant alternatives. This is clearly one epistemic role of salience generators.[10]

So far I have argued that rationality requires the existence of salience generators and that the states that perform this role cannot always be propositional attitudes. I have not, however, said anything about what these states may be. In order to answer this question I would like to re-consider some examples. Competent chess players, I have claimed, are able to perceive threats and spot opportunities. What enables them to do this is not calculation. Instead, good players do it by becoming excited about some moves, scared or worried about their king's safety. In other words, emotional states, by focusing the players' attention on some aspects of the game and away from others, are what generates saliences in this case. This phenomenon is not unique to chess.

Fear, interest, apprehensiveness, excitement all guide our perceptions of given situations. They make us focus our attention on some hypotheses and on features which would have otherwise gone unnoticed. Parental love offers a good example of this phenomenon. Parents, of course, possess a lot of propositional knowledge about their children. Typically, however, they do not reason inferentially on the basis of this knowledge when assessing, for example, whether or not the child's cry is a symptom of true distress. Parents can pick up features of their child's behaviour that are often opaque to others. Their love for their child makes the child and her needs salient to the parent in most circumstances. This emotion involves a pattern of attention, lasting many years, focused on the child's needs, and thus contributes to the acquisition of further knowledge about the child. This in turn informs parents' loving attention directed towards the child. Parental love, as a salience generator, plays a crucial role in the parents' skilful dealings with the child; and through sustained emotional engagement with the child, the parents become more reliable at detecting saliences; their skills become more fine-tuned. In other contexts, a healthy sense of fear helps competent individuals to negotiate dangerous situations. An important part of learning to be a good yachtsperson is to learn of what to be truly afraid.[11]

It is not only in practical contexts that emotions play the role of salience generators. Scientists' sustained enthusiasm and interest in various problems plays an essential

10 But, as my previous examples illustrate, it is not their only epistemic role.

11 The view that emotions play the distinctively cognitive role of generating saliences has also been defended by Catherine Z. Elgin (1996). A similar position has also been defended by Ronald de Sousa (1987).

role in theoretical inquiry. Without such interest scientists would soon lose focus and abandon their inquiry. Also, scientists pursue in their investigations the hypotheses they are enthusiastic about and carefully consider those objections of which they are afraid. Emotions thus play a crucial epistemic role in all rational conduct.

There is also some evidence from neuroscience that appears to confirm the view that emotions are crucial to rationality because they are salience generators. A case in point is Antonio Damasio's patient 'Elliot' who, having suffered frontal lobe brain damage to areas known to govern emotional response, lost the ability to accomplish many tasks. Damasio notices that Elliot knew a lot of stuff, but was not able to focus on anything. If asked to read and classify some documents, Elliot might get distracted and spend the whole afternoon deliberating about principles of classification. In Damasio's view:

> One might say that Elliot had become irrational concerning the larger frame of behavior, which pertained to his main priority … . His knowledge base seemed to survive, and he could perform many separate actions as well as before. But he could not be counted on to perform the appropriate action when it was expected. (Damasio 1995, 36–7)

What Damasio's work appears to have uncovered is that the powers of reasons, specifically the ability to pick out relevant factors and ignore others, and the experience of emotions decline together. This work offers further evidence that salience generators are essential to rationality and that in the case of human beings emotions perform this role.

It should not be surprising that emotions can play a role in epistemology since they possess the features that make them suitable for the task.[12] Emotions are subject to normative evaluations and can figure in rational explanations. When judging emotional responses to various situations it makes perfect sense to consider whether the response was warranted or justified. The notions of warrant and justification used in these evaluations are often epistemic rather than prudential. So when we are trying to convince a friend that their jealousy is not warranted, we do so by discussing the facts and showing to the friend in question that their jealousy is ungrounded. Similarly, an instructor might in the context of a class try to teach students which kind of objection to a theory they ought to worry about and which should not unduly concern them.

Further, when trying to explain the reasons for a person's action, one often invokes that person's emotions. People often do things for their friends because they love them; they also set time aside to spend with friends because they enjoy their company. Analogously, a scientist might pursue a line of inquiry because he is enthusiastic about its prospects; he might dismiss another possibility because he finds it boring and unattractive. In conclusion, emotions, to use a Sellarsian phrase, inhabit the space of reasons.

Since emotions can figure as justifications and are themselves sometimes in need of justification, emotional states must presumably be intentional and therefore

12 I do not take these observations to deny that emotions might have other less clearly cognitive characteristics. If so, work would have to be done to explain why these features would all come together in emotions.

contentful. If we understand intentionality as roughly speaking 'aboutness', emotions appear to be intentional. Thus, for instance, fear of tigers is about tigers, worry for the king's safety is about the king's safety and excitement about an hypothesis is about that hypothesis. These examples highlight a notable feature of emotional concepts: their nominal objects. The typical intentional focus of an emotion is an object or a property-instance. This characteristic of emotions seems to distinguish them from propositional attitudes such as beliefs whose intentional focus is a proposition.

It is true that on occasion emotion concepts can take propositional objects. For instance, one might be afraid that the tiger will attack. Such uses demonstrate the conceptual richness of emotional contents. They do not undermine the point about the distinctive objectual intentional focus of emotions. The point stands because not all uses of emotion concepts can be recast by furnishing them with a propositional content. It makes perfect sense to say that a sailor is worried about the sea-state even though they are not in a position to state what precisely about the sea-state is of concern to them.

There is, of course, a vast philosophical literature arguing precisely that emotions are or entail belief-like states. I do not take this brief observation to refute such a view. My aim is more modest. I simply intend to claim that it is not obvious that emotions are beliefs and that there is some evidence that they are not. I have already mentioned the nominal object of emotion concepts as providing some evidence against taking emotions to be doxastic states. The observation that emotions cannot occur without the belief which they should entail or be equivalent to provides further evidence. There are plenty of individuals who are afraid of tiny garden spiders without for a moment believing that they are dangerous. But what I take to be the strongest evidence for the claim that emotions are not beliefs, is precisely the case I have been developing in this section. Emotions, I have argued, are salience generators, and in most cases beliefs cannot perform this role.

Before addressing the connection between emotions as discussed here and intellectual virtue, a short clarification might be in order. What I have argued so far is that salience generators are essential to rationality and that in human beings emotions function as salience generators. I have not established that emotions are essential to rationality. What is essential to it is that states which are like human emotions insofar as they generate saliences are deployed. Obviously, the argument developed here would not show that emotions are essential to rationality, if being an affective state rather than a generator of saliences is what makes a state an instance of an emotional state.[13]

Emotions, whatever their essential nature might be, are important components of the virtues both moral and intellectual. There are at least three aspects of emotional states that contribute to their role as components of virtues (Sherman and White 2003). Firstly, virtues have motives which are typically associated with them. Thus, for instance, the courageous person is usually motivated by courage when she acts courageously. Often, these motives are emotions. Thus, the benevolent person, for

13 There might be good reasons why affective states are salience generators. These reasons are connected to the role of affectivity for actions, and the crucial involvement of salience generators in decision making.

instance, is motivated by love and the open-minded person is motivated by curiosity, enthusiasm, care and respect for other people's views. Emotions can be motivational components of virtue because, as I have argued above, they can be reasons for actions.

The second contribution that emotions make to virtue has been the main topic of discussion of this section; namely, their role in tracking saliences. Virtues are excellences and intellectual virtues are intellectual excellences. These excellences either are or involve skills, and they are in some way connected to success.[14] These are vast topics which I cannot address here. For my purposes the following considerations should suffice. Persons who listen to every opinion and end up believing everything they are told, are not open-minded; they are gullible. Individuals who frequently change their minds on a given topic are also not open-minded. They simply lack any intellectual constancy. The open-minded person is the one that gives their due to all views. She does not waste her time trying to find out more about views that are clearly foolish. But she is prepared to take seriously those views that are a genuine threat to her opinions. This ability to identify among conflicting points of view those that require further examination is not merely a matter of inferential ability. We are all familiar with the existence of very smart individuals with whom it is often pointless to talk because they are never really prepared to take seriously anybody else's views. Their close-mindedness is not evidence of a deficiency in their inferential reasoning; it is a matter of their emotional sensibility. They simply are not capable of perceiving views which are in conflict with their own as deserving of consideration.

Thirdly, several intellectual virtues concern our relations to other rational thinkers and are often exercised in the context of our interactions with others. These interactions will include dialogue and practical team work. In these contexts emotions as modes of expressing our attitudes play an important role. The person who lacks intellectual humility often displays arrogance in their interaction with others. The intellectually humble person, on the other hand, will display a range of emotional attitudes which will foster co-operation and team work.

These brief considerations hopefully go some way to suggest that the attentional and motivational patterns that are essential to intellectual virtues are exemplified by emotions. Or to put it in a different way, the virtuous person is the one who has acquired a certain kind of sensibility which is, at least in part, characterized by a combination of apt emotional dispositions. I thus disagree with Paul Bloomfield who states that 'we should not take emotions to be a necessary feature of the intellectual virtues' (Bloomfield 2000, 34). Bloomfield motivates his conclusion by virtue of a single example: a judge's exercise of justice. He acknowledges that pity or anger can provide information to the judge that might be valuable when deciding on a verdict. But for Bloomfield it would be wrong for the judge to be 'in the grip of such emotions' because he or she must exercise impartiality and emotions often interfere with impartiality.

---

14 I do not rehearse the debate on whether virtues are skills here. Aristotle, alone among the Greeks, held that virtues are not skills. The same view is also defended by Linda Zagzebski (1996) and opposed by Julia Annas (2003) and by Paul Bloomfield (2000).

Bloomfield's argument is unconvincing. Even if we ignore the exceptional features of this single example which is centred on the virtues of impartiality, the inessentiality of emotions to the exercise of intellectual virtues does not follow. Quite the contrary, in order to reach a good verdict the judge must deploy his or her emotional sensibility in order to pick out as relevant some features of the situation. The judge must also be able to sustain her attention on the details of the case without being distracted by other matters, and in order to do so the judge must be interested in the case. Further, the judge might be suspicious of some of the strategies of the defence lawyer and must be able to rule out some objections as irrelevant. He or she must be able to sympathize with the position of some witnesses and to evaluate the motivations of the defendant. These are all abilities that require the deployment of an emotional sensibility. A judge who is lacking all emotion would not be impartial; in all likelihood he or she would be perceived as cold, perhaps even callous.

The overall picture which has emerged from this discussion is, I hope, clear. Human rationality requires more than the ability to draw valid inferences. It also requires the capacity to select some inferences as relevant in a given context. This capacity is a skill which, in human beings, has an emotional component, for in human beings it is emotions that play the required role of being pre-doxastic salience generators. I say that the capacity to select some inferences as salient is a skill because it can be learnt and fine-tuned.[15] In many activities, the beginner is often scared of things and situations which are not dangerous, but is not afraid of the real dangers because he is unable to spot them. One of the important features of experience is precisely the development of apt emotional responses to one's circumstances. The idea that rationality requires the development of skills, including the ability to be attuned to salient features, finds in the notion of intellectual virtue its natural home. Whatever else intellectual virtues might be, they are human excellences. They involve skills and are connected with character traits. They also have emotions as components. More specifically, it is by developing a properly attuned emotional sensibility that the virtuous knower is able to pick out saliences, to act in accordance with the demands of the situation and to interact fruitfully with others.

3. Fallibilism and Epistemic Humility

In the first section of this chapter I have detailed the failures of some accounts of fallibility as an epistemic property of knowledge states and methods of inquiry. In this section I propose an account of fallibility in terms of virtue which I believe can meet the criteria also detailed in the first section. More specifically I propose that we take fallibilism to be the virtue of epistemic humility and that we understand that virtue as openness to views that are incompatible with one's own. Appreciation of one's own fallibility is a virtue; it does not require possessing the belief that it is possible that one's views are false. Instead, it is manifested in the ability to listen, assess fairly and explore the opinions, conflicting with our own, that are put forward

15 It is also a skill in the sense that to possess it one does not need to be able to explicitly state one's guiding principles.

by others. The intellectually humble person does not necessarily lack conviction nor is she necessarily gullible. Rather the intellectually humble person is neither vain nor arrogant.[16]

It is, therefore, helpful in order to begin to get a grip on intellectual humility to contrast this virtue with its corresponding vices: intellectual arrogance and vanity. The intellectually arrogant are supremely confident of their intellectual superiority and derive from it an inflated sense of entitlement. For this reason intellectually arrogant people are inclined to think that they never make mistakes and that the arguments and views put forward by others are less interesting than their own. More specifically, the intellectually arrogant person is prone to inflate her own intellectual credibility and to underestimate the credibility of the views of others. Thus, intellectual arrogance fosters dogmatism. It generates a certain inability to adopt a proper critical attitude towards one's own views and a receptive attitude towards the positions defended by others.

Intellectual vanity is also an obstacle to open-mindedness. The intellectually vain person is mostly concerned with making a good impression and with being highly regarded by powerful individuals. Such an individual is likely not to admit that she does not know something when asked by students, since she thinks that any admission of ignorance would diminish her status in their eyes. Thus, the intellectually vain person might make an educated guess and present the answer as knowledge. For similar reasons, the intellectually vain person might prefer to fudge an argument, rather than to acknowledge publicly the force of an objection. Also, since she excessively cares about being well-regarded by powerful individuals, the intellectually vain person cares about being seen as belonging to the right circles and thus ignores or avoids discussing matters with people she does not regard as sufficiently important. In a word, in the intellectually vain care for social status trumps any concerns she might have for intellectual honesty.

We have all come across both intellectual arrogance and intellectual vanity. We also know that they have a tendency to make conversation pointless. The individual whose behaviour exhibits these traits to a significant degree is not able to engage in genuine dialogue with others. She might be smart and have exceptional inferential abilities; she might also be very knowledgeable in her field. And yet, we want to say that her vision is blinkered. The root of the problem is that this person has not learnt to behave in intellectually humble ways.

The intellectually humble person shows instead a healthy disregard for concerns with social status. She cares more about figuring out the answers to issues than about making a good impression on the right persons. Further, the intellectually humble person is prepared to listen to what others might have to say, because she does not arrogate for herself the entitlement to have the final word. Intellectual humility, however, is not the mere absence of arrogance and vanity. It involves the presence of specific motivational patterns and character traits.

The intellectually humble person is genuinely motivated by a desire to learn from other people and is thus more likely to approach them with an open mind. She is also more likely to be open to the views of a broad spectrum of individuals because her

16 On humility as the opposite of vanity and arrogance see (Roberts and Wood 2003).

lack of arrogance and vanity will prevent her from caring only about the opinions of individuals who occupy social positions of influence. Of course, the intellectually humble person can have a healthy sense of the value of her own opinions. Also, her humility does not lead her to consider carefully every possible opinion. It is not a sign of arrogance not to engage with the views of quacks and frauds. What is characteristic of the intellectually humble person is that she treats every opinion with the consideration, be it high or low, it deserves.

In the first section I have individuated five criteria that any satisfactory account of the notion of fallibilism must satisfy. In what follows I intend briefly to show that an account in terms of the virtue of intellectual humility succeeds in this regard.

1. The account connects fallibilism with the possibility of mistakes.

In my view the fallible knower implicitly acknowledges the possibility that her views might be wrong by maintaining an attitude of open-mindedness towards the opposing views held by others. Fallibilism, however, does not require that every belief could be false, since this is patently impossible in the case of necessary truths. Nor does it require that all knowledge based on evidence could fail to be knowledge, since knowledge based on proofs is not open to this sort of failure. Fallibilism also does not require that knowers have any explicit belief about the possible falsity of all of their beliefs, since such belief would probably lead to pragmatic contradictions. The trademark of fallibilism is instead found in the attitudes of knowers both towards their own views, which they regard as subject to rational revision, and towards those of others, which they consider with an open mind.

2. The account applies globally to human knowledge.

Intellectual virtues have global application to all areas of inquiry. The intellectually humble thinker brings her open-mindedness to bear in all investigations, including her study of necessary truths in arithmetic or logic. Fallibilism as the attitude of humility has global application.

3. The account does not rule out the possibility of human knowledge.

Fallibilism conceived as the virtue of intellectual humility has no tendency to give rise to scepticism. The intellectually humble person can be deeply committed to her views and she might have genuinely gained entitlement to them. Thus, fallibilism is not incompatible with the possibility of knowledge. Further, when fallibilism is properly understood it becomes apparent that the character traits it encourages are not those characteristic of the Cartesian sceptic.[17]

17 There might, instead, be a connection with scepticism in its ancient version. Pyrrhonian sceptics thought of themselves as investigators who were open-minded about every claim to knowledge. However, they took this attitude to be incompatible with the fixation of belief. The practice of intellectual humility has no such consequence. One needs not suppress a tendency to have opinions in order to be humble.

4. The account contrasts fallibilism to dogmatism.

In my account, being open-minded is one of the defining features of intellectual humility. The intellectually humble is opposed to the dogmatist who is close-minded and not prepared to consider revising her views in light of opposing evidence.

5. The account relates fallibility to the notion of rational revision.

Finally, the account explains how fallibilism relates to the rational revision of belief. Fallibilism neither requires that every view must be subject to the possibility of rational revision, nor that thinkers must believe they are. Instead, it is because the virtuous knower always keeps an open mind that she implicitly manifests in her behaviour a preparedness to rational revision as and when her circumstances demand it.

In conclusion, I propose that we conceive of fallibilism as a virtue that is possessed by some epistemic agents rather than as a feature of doxastic states or methods of inference. This approach, I have shown, avoids some of the problems encountered by earlier accounts and offers a plausible account of some aspects of epistemic conduct.

References

Annas, Julia. 2003. 'The Structure of Virtue'. In: De Paul, Michael R.; Zagzebski, Linda (eds). *Intellectual Virtue. Perspectives from Ethics and Epistemology*. Oxford: Clarendon Press. 15–33.

Bloomfield, Paul. 2000. 'Virtue Epistemology and the Epistemology of Virtue'. *Philosophy and Phenomenological Research* 60: 23–43.

Carrier, Leonard S. 1993. 'How to Define a Nonskeptical Fallibilism'. *Philosophia* 22: 361–72.

Damasio, Antonio R. 1995. *Descartes' Error. Emotion, Reason and the Human Brain*. London: Picador.

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.

Elgin, Catherine Z. 1996. *Considered Judgment*. Princeton: Princeton University Press.

Lance, Mark; Tanesini, Alessandra. 2004. 'Emotion and Rationality'. In: Ezcurdia, Maite; Stainton, Robert; Viger, Christopher (eds). *New Essays in the Philosophy of Language and Mind. Canadian Journal of Philosophy*, suppl. vol. 30: 275–95.

Peirce, Charles S. 1960. *Collected Papers*. Vols 1 and 2. Ed. Charles Hartshorne and Paul Weiss. Cambridge, Mass.: Belknap Press of Harvard University Press.

Reed, Baron. 2002. 'How to Think about Fallibilism'. *Philosophical Studies* 107: 143–57.

Chapter Four

Conflict Without Contradiction

Sabine A. Döring

How to account for the fact that content-involving rational conflicts between emotions and judgements do not imply contradictions? Why is it that there is no contradiction involved, e.g., in feeling fear of falling although one judges, and even knows, that one is safe? I discuss two answers that have been given to the question of how 'conflict without contradiction' could be logically possible. In the first case, conflict without contradiction is explained by difference in content. In the second case, it is explained by difference in attitude towards content. I argue that conflict without contradiction is due to both difference in content and difference in attitude towards content. It is shown that emotions are analogous to sense perceptions in this respect. Furthermore I claim that emotions resemble sense perceptions in being a non-inferential source of reasoning and knowledge.

1. The Problem

It is common ground to contemporary theorizing about the emotions that they can rationalize, rather than merely explain, other states and actions. The common ground is the rationalizing role itself – not its conceptualization and logic. Many have insisted that this role cannot be captured by a judgementalist model of emotion, pointing to the fact that, by contrast with judgements or beliefs, emotions need not be revised in light of the subject's better judgement and knowledge (see, e.g., de Sousa 1987; Goldie 2000; Goldie 2006; Greenspan 1981; Greenspan 1988; Helm 2001; Helm 2006; Prinz 2004; Roberts 1988; Roberts 2003; Tappolet 2000; Tappolet 2003; Döring 2003; Döring 2004; Döring 2007).[1] Hume provides a prime example: your fear of falling may persist and represent it as dangerous to be so high up above the ground although you judge and even know that you are safe.[2] This kind of conflict between emotion and judgement is readily intelligible and happens all too often. It need not be a pathological case of vertigo, say, but occurs as an ordinary experience of ordinary people.

1 I shall also count as judgementalist theories of emotion those theories which employ the notion of belief. For the present, I shall not distinguish between judgement and belief.

2 In his *Treatise*, Hume invites us to consider 'the case of a man, who being hung out from a high tower in a cage of iron cannot forbear trembling, when he surveys the precipice below him, tho' he knows himself to be perfectly secure from falling, by his experience of the solidity of the iron, which supports him' (Hume [1739/40], 148).

Yet it is not just the psychological phenomenon which is at stake here. The point at issue is the logical possibility of 'conflict without contradiction', as I shall call it.[3] Conflicts between emotions and judgements are rational conflicts, i.e. conflicts in intentional content about how the world actually is, and yet they differ from rational conflicts between judgements in that they do not involve contradictions. Of course, this is not to say that rational conflicts between judgements are always contradictory. But if, in the case of two conflicting judgements, one judgement predicates that a certain thing has a certain property, whilst the other judgement denies that the thing has that property, we are faced with a contradiction. Even if it should be psychologically possible, it would clearly be contradictory to judge that you are safe whilst at the same time judging that you are in danger (not safe). By contrast, it is not contradictory to judge that you are safe whilst at the same time feeling fear, and that is: experiencing the situation as dangerous. Although judgement and emotion are about the same thing and seem to contradict each other in how they represent that thing, there is, in fact, no contradiction.

The possibility of conflict without contradiction is a touchstone for any cognitive theory of emotion, i.e. for any theory that attributes intentionality to the emotions.[4] My aim in this paper is to account for this possibility. My focus will be on the analogy that has repeatedly been drawn to perceptual illusions such as the famous Müller-Lyer illusion (see, e.g., the illustration in Crane 1992). Just as, despite his better knowledge, the protagonist of Hume's example cannot help but see himself in danger, the perceiver of the Müller-Lyer illusion cannot rid himself of seeing the two lines as being of a different length even when careful study has convinced him that they are the same length. I shall argue that conflicts between emotions and judgements are to be explained along the same lines as conflicts between perceptions and judgements. In both cases, so I shall claim, conflict without contradiction is due to the fact that the state in question differs from judgement in two respects: in content and in attitude towards content.

Let us be clear that, to be capable of playing a rationalizing role and of entering rational conflicts, both sense perceptions and emotions must have an intentional content of a certain kind. They must be about the world and represent it as being a certain way, so as to be assessable for correctness, i.e. for adequacy to the actual state of the world that they purport to represent. It is Ronald de Sousa (1987, 149ff.) who put forward a pioneering discussion of the analogy between sensory perception and the emotions in this cognitive respect. Like de Sousa, I shall not assimilate the emotions to sense perceptions. Emotions are not sense perceptions, but are best understood as what I have elsewhere characterized as 'affective perceptions' (see, e.g., Döring 2003; Döring 2007; Döring 2008). There are a number of obvious disanalogies between these two kinds of states. None of them undermines the analogy in question, for the latter concerns only the rationalizing role of both kinds of states

3 I owe this expression to Peter Goldie.

4 I shall not restrict emotional cognitivism to the judgementalist model, but use the term 'cognitive' to characterize any theory in which emotions are defined as intentional states. Conversely, any theory that denies the intentionality of emotion will be characterized as 'non-cognitive'.

that they play because of their specific kind of intentional content. The possibility of conflict without contradiction suggests that emotions are like sense perceptions, and unlike judgements, in that they are not inferentially related to other states. That is, in the rationalization of other states and actions the emotions play the non-inferential role of perception, and in that sense, so I claim, they *are* perceptions.

Let me conclude these introductory remarks by stressing that I do not mean to account for every state that is called an 'emotion' in ordinary language, nor do I claim to cover every state that was ever called so in the history of philosophy. It suffices if my argument works for paradigm cases of human emotions and can, on that basis, integrate or possibly exclude less paradigmatic cases.[5] As paradigm cases of human emotions I regard occurrent conscious states with an intentional content of a certain kind, such as the fear you may feel at the sight of a precipice below you, the anger and indignation you may experience when you feel that you have been offended, or the envy that may overcome you when you see your colleague driving in his new Maserati.

2. Conflict in Content or Conflict in Attitude?

The renaissance the emotions are celebrating in contemporary philosophy is mainly due to the insight that emotions are intentional mental states. At the beginning of the debate this insight led to a strong opposition between (supposedly) new cognitive theories of emotion and traditional non-cognitive feeling theories. In its early stage cognitivism amounts to the claim that an emotion is defined by a certain belief or judgement which it necessarily involves (see, e.g., Kenny 1963; Lyons 1980; Solomon 1993). To qualify as an instance of fear, for example, a token emotion must involve the judgement that the thing which it is about – its intentional object – is fearsome or dangerous. Since emotions are not neutral on things but evaluate them in light of the subject's concerns (as fearsome, annoying, enviable or the like), this means that an emotion necessarily involves a certain value judgement. Even when, as in the more sophisticated early theories, the cognitivist does not deny that emotions also involve other factors such as, in particular, certain feelings, these factors are taken to be irrelevant for distinguishing emotion types from each other. This is in contrast to the feeling theory which analyses an emotion in terms of its phenomenology alone, i.e., in terms of what it is like to experience the emotion.

By now, the extreme forms of cognitivism and non-cognitivism are mostly abandoned in favour of more balanced views. It is generally accepted that a theory of emotion, although it must not reduce emotion to feeling, has to account for the fact that emotions – carefully speaking, paradigm cases of human emotion – are by their nature felt states. The classical opposition between cognitivism and feeling theory is watered down. Peter Goldie, for one, emphasizes that the emotions' intentionality

5 This in contrast to Amélie Rorty's (1984) and Paul E. Griffiths's (1997) views. Both deny that there is such a thing as one single category of *the* emotions, although for somewhat different reasons. I shall not discuss this point further here. This has already been done by Robert C. Roberts (2003, ch. 1) who presents the most comprehensive defence to date of the coherence of the category 'emotion'.

cannot be understood independently of their phenomenology (see Goldie 2000, ch. 3). As Goldie points out, a theory of emotion that merely 'adds on' non-intentional feelings, such as bodily feelings, to otherwise feelingless intentional states, such as judgements, does not do justice to the phenomena.[6] Opposing 'add-on views' (and, implicitly, the functionalist view of the mind behind them), Goldie has it that emotional feelings are not non-intentional feelings, but are directed towards a certain thing, which is the emotion's intentional object. Thus the emotions' intentionality cannot be separated from their phenomenology but is built into it (so to speak): what an emotion is about is part of its conscious, subjective character, of what it is like to experience the emotion.

One of Goldie's arguments for affective intentionality, the claim that the emotions' intentionality must necessarily be felt, is the argument from conflict without contradiction (see also Goldie 2006). Neither the feeling theory nor the judgementalist theory of emotion, Goldie says, would allow us to account for the way in which our emotions sometimes conflict with our better judgement and knowledge. As Goldie puts it, non-intentional feelings give us less than we want – no conflict, or at least no rational conflict –, and feelingless judgements give us more than we want – contradiction. On the one hand, rational conflict holds between the intentional contents of two states, not between the states as mere mental events. Therefore rational conflict between emotion and judgement requires that our emotions resemble our judgements in having an intentional content. Like judgements, the emotions must be about something. When it comes to conflict between emotion and judgement, both emotion and judgement are about the same thing and the emotion represents that thing in a different way than the judgement does. On the other hand, to avoid contradiction, the emotions' intentional content must not assume the form of the intentional content of judgement. On such grounds Goldie suggests that the emotions' intentional content must be of a specific kind. On his view, the difference in intentional content between emotion and judgement is best explained by conceiving the emotions' intentionality as felt or affective intentionality.

This view is not without alternatives, however. Could not the possibility of conflict without contradiction equally be explained by a difference in attitude

6 Within 'add-on views', as Goldie labels the theories in question, emotional feelings are typically interpreted as a Jamesian bodily feeling, i.e., as awareness of the bodily changes involved in emotion, such as muscular reactions, including changes in facial expression, hormonal changes and changes of the autonomic nervous system. Add-on views are nourished by the hope that emotions can be fully characterized in functional terms amenable to the impersonal perspective of the sciences. While intentionality is taken to be explainable in functional terms and therefore seen as an 'easy problem', phenomenal consciousness is regarded as a not yet solved and probably insoluble 'hard problem' (to quote David Chalmers). This is why the add-on theorist isolates emotional feelings from intentional judgements or belief-desire pairs as far as possible. On his account, feelings only enter the scene after an emotion's intentionality has been fully characterized in functional terms, and they do so only as the non-intentional felt feedback of the bodily changes involved in emotion. Descriptions of what it is like to experience such changes from a personal perspective are dismissed as relatively unimportant, because they are compatible with entirely different descriptions of the nature of those changes.

towards content, rather than by a difference in content? This seems to be the view of Robert C. Roberts (1988; 2003). Roberts is unclear on whether there is always a something-it-is-like, some affect, attached to emotion (see also Lacewing 2004). He introduces the emotions as 'concern-based construals'. As construals, they are ways of presenting their objects, without this implying that the subject regards the presentation as true. According to Roberts (1988, 191; see also 2003, 92), the content of an emotion is 'verisimilar' by which he means to say that, for the subject, the content has 'the *appearance* of truth, whether or not she would *affirm*' its truth. This is to say, the emotions differ from judgements or beliefs, in that the subject does not regard the content of his emotion as true.[7] On Roberts's account, when it comes to conflict between emotion and judgement, no contradiction arises because the subject only affirms the truth of his judgement's content, whereas the content of his emotion merely appears to be true to him. Thus, in feeling fear of falling, it merely appears to Hume's protagonist that he is in danger. Because he does not affirm the truth of his fear's content, he may still consistently judge and regard as true that he is safe. In Roberts's view, the difference between emotion and judgement required to account for conflict without contradiction is best explained as a difference in attitude towards content.

The same difference Roberts finds between the attitude of perception and the attitude of judgement. Difference in attitude, Roberts claims, explains also why one can consistently see a stick half-submerged in water as bent even after one has judged that it is straight; or why there is no contradiction involved in seeing the two lines of the Müller-Lyer illusion as being of a different length even after one has found out that, actually, they are the same length. On Roberts's view, we can understand conflict without contradiction in the case of emotion as analogous to perceptual illusion: an appearance that may consistently persist even after we repudiate it in judgement.

So we have at least two competing explanations of why conflicts between emotions and judgements are not contradictions. Do emotions differ from judgements in content or in attitude? My own view is that emotions and judgements differ in both respects, where difference in content is prior to difference in attitude. So let me begin with difference in content.

3. Non-Inferentiality and Affective Intentionality

I have argued that, in order to allow for rational conflict with judgement, an emotion must have an intentional content. Yet intentionality is not enough. More precisely, an emotion must have a 'representational content', to use Christopher Peacocke's influential notion (see Peacocke 1992; see also Döring and Peacocke 2002). Although representational content is also intentional, it differs from non-representational intentional content in being subject to a correctness condition. It is content that represents the world as being a certain way, and can thus be correct

7 Lurking in the background of this distinction is the familiar metaphor of belief 'aiming at truth'.

or incorrect. In experiencing fear of falling, it seems to you that you are in fact in danger: your occurrent emotional state puts forward your fear's content as correct. This is in contrast to, say, your imagining yourself to win gold at dressage world championship. Under conditions specifiable as normal, you will be aware that this imagining is just a daydream. It will not seem to you, outright, that you could really ride like Isabell Werth.

I think that it is this difference between representational and non-representational intentional content that Roberts has in mind when he states that the content of emotion has the appearance of *truth*. Even though Roberts denies that the content of emotion is regarded as true, he does not mean to exclude that our emotions can rationalize our judgements and actions. To be capable of doing so, an emotion must neither be a non-intentional feeling nor a non-representational fantasy, but appear to represent something that is or was or could really be the case (see, in detail, Döring 2007). Fear could never rationalize your judging that you are in danger, nor your getting yourself to safety, had it either no intentional content at all, or contained no representation of the world. Without representational content, your fear would merely be a 'blind instigator', to borrow Robert Musil's apt metaphor, and it would be equally rational for you to take a drug that releases you from your negative emotion. Furthermore, there could be no rational conflict between emotion and judgement, for rational conflict is about how the world *is*, about what the facts are. Since an intentional but non-representational state like your fantasy of winning gold at dressage world championship does not normally purport to represent an actual state of the world, or fact, it could no more conflict with your insight that you will never be able to ride like Isabell Werth than a non-intentional feeling could.

The attitude corresponding to intentional but non-representational content is that of regarding as merely fictitious. Thus, the attitude towards emotional content must also be different from the attitude of regarding as merely fictitious. I shall turn to the question of what kind of attitude this could be in the next section. In this section, my concern is with the question of why emotional content, even though it is like the content of judgement in representing the world as being a certain way, yet differs from the content of judgement.

Goldie (2000, 58ff.) appeals to the phenomena (as others have done before him; see already Alston 1967). He points to the difference it makes whether one thinks of a certain thing as dangerous in a feelingless way or experiences that thing as dangerous. His example is about a gorilla in a zoo whom you see grimly loping from left to right in its cage. According to Goldie, the difference in question can be seen from what happens when you suddenly realize that the door to the cage has been left open. Until this moment, Goldie says, you see the gorilla as dangerous, but you do not feel fear because you believe the gorilla to be safely behind bars. (As always, your belief need not prevent you from feeling fear of the gorilla. Conflict without contradiction may well arise. But this is not the sort of case Goldie is looking at here.) When this belief turns out to be false, you become 'emotionally engaged with the world', as Goldie (2000, 61) puts it. Goldie emphasizes that the difference between the new, emotional evaluation and the earlier, calm evaluation is a difference in

content, despite the fact that both contents might be naturally expressed in the same words ('The gorilla is dangerous').[8]

The difference in content between emotion and judgement becomes particularly clear when we look at how emotions rationalize other states and actions, including other emotions. By contrast with judgements, they do so non-inferentially. To begin with, a judgement made rational by the representational content of the subject's emotion is not a matter of an inference from the occurrence of the emotion. When you are gripped with fear facing a furious gorilla that escaped from his cage in the zoo, fortunately, your judging, and regarding as true, that the gorilla is dangerous does not need to wait for inferential means. Instead, you immediately rely on your emotion. Your judgement that the gorilla is dangerous is non-inferentially justified by your taking your fear's content at face value.

Starting from this example, we may distinguish different senses of non-inferentiality. First, the rationalization of the judgement in question is non-inferential in the phenomenological sense that it does not involve conscious inferential reasoning by the subject. This does not preclude there being non-conscious inferential processing. Second, and more important here, the rationalization of a judgement can be non-inferential in the phenomenological sense, and yet be inferential in the epistemic sense, i.e. the judgement can yet be inferentially justifiable. If you are challenged about your judgement that the gorilla is dangerous, you might 'retreat' to an inferential justification. If such a justification were available, this would show that you could in principle have come to judge that the gorilla is dangerous by an inferential route, although your judgement is in fact not reached by inferential means.

My concern here is with 'epistemic' non-inferentiality, not just with 'phenomenological' non-inferentiality. While it may seem in Goldie's gorilla example that, were you asked why you think that the gorilla is dangerous, there would be an inferential justification available, things look different in other examples, depending on the emotional properties involved. We may again use one of Goldie's (2006) examples. Although it was designed for a different purpose, this example shows that inferential relations do also not hold *between* emotions. You feel that someone's behaviour is irritating, but then start to feel amused about how over the top his irritating behaviour is. That is, you are amused *because* the behaviour is so irritating. The 'because' is ambiguous, for it may indicate either a merely causal or a rational explanation (which is, of course, not to deny that rational explanations might also be causal). From your point of view, at least, your amusement is made rational by experiencing the behaviour as way over the top irritating. On the face of it, it seems plain that there is no such thing as inferring the experience of the behaviour as amusing from the experience of its being over the top irritating. It is not that you deduce the conclusion that his behaviour is amusing from premises stating that his behaviour is over the top irritating and that behaviour, if it is over the top irritating, is amusing. No *modus ponens* involved here. Instead, you immediately experience the behaviour as amusing.

---

8 What happens to feeling when we put words to it is explored by David Pugmire (2007) in his paper 'Go on*, Say* It!'.

Let us, for the sake of the argument, presume that your amusement about the person's behaviour is appropriate, and yet your friend, who is also present at the scene, fails to see it this way. All you can do to convince him of the appropriateness of your emotional reaction is to point to the salient features which, for your eyes, make the behaviour amusing. None of your hints will suffice to 'force' your friend to see the behaviour the same way, since this does not follow, and can thus not be inferred, from any of its saliences. To understand, he must immediately experience the behaviour as amusing.

Tim Crane (1992, 152) makes a similar point about the content of perception. The content of perception is defined as representational in the sense of Peacocke, i.e. as being subject to a correctness condition, but is distinguished from the content of judgement by its not being subject to inferential constraints. As Crane points out, there is no such thing as inferring the perception of a certain thing *a* as being *F* and *G* from the perception of *a* as *F* and the perception of *a* as *G*. In order to gain the perceptual content of *a* as *F* and *G*, both properties must be perceived at the same time.

To use an example that corresponds to the one of the amusement about the irritating behaviour: if you see a vertical line, and then you become aware of a dot above it so that you see the letter *i*, you do not, and cannot, infer the perception of an *i*. This perception is immediate, just as the amusement about the irritating behaviour is immediate. To make this clearer, compare the way we discern gestalt figures. Consider the well-known old woman / young woman gestalt figure (see, e.g., the illustration in Roberts 2003, 70). Let us suppose that you see only the young woman (as I first did). Clearly this is a perception, rather than a judgement or belief. You must see the figure as the picture of a young woman. In the same way, you must see it as the picture of an old woman to understand that it represents an old woman as well. Someone might describe the old woman to you by pointing to the saliences that constitute this picture. None of his descriptions will suffice to make you understand that the figure is the picture of an old woman since this does not follow from any description of the figure. To understand this, you must see the old woman yourself. Likewise, you must see the letter *i* with your own eyes in order to understand that it is constituted by the vertical line with the dot above it in front of you. A vertical line with a dot above it could equally be a semicolon, say.

As the properties of being amusing, or of being the picture of an old woman, or of being the letter *i* cannot, in the strict sense of the term, be inferred and inferentially justified in any case, a judgement based on the perception of these properties is not the product of an inference. Instead, the contents of perceptual judgements are arrived at non-inferentially, by taking the content of the perception, be it sensory or affective, at face value.

If it is agreed that emotions are like perceptions, and unlike judgements, in that they do not enter inferential relations, we may then ask how this could be so. A natural answer, also given by Crane, is that non-inferentiality is to be explained by a state's having a different content than the content of inferential judgement. It is natural to assume that inferentiality is determined by the specific kind of intentional content of judgement. The contents of our judgements must be so as to relate inferentially to each other. The fact that emotions and perceptions are not inferentially related to

each other, and to other states, suggests that their content is different from the content of judgement. Assuming that 'concepts are the *inferentially relevant constituents of intentional states*', Crane (1992, 147) defines the content of non-inferential perception as nonconceptual.[9] Although I agree with him (and others), I shall avoid mentioning nonconceptuality, as my focus in this chapter is on the phenomenal character, or affectivity, of perceptual content (cf., however, Döring 2008). Of course, explaining conflict without contradiction by a difference in content of whatever kind requires that different kinds of content can conflict. I see no reason to exclude this possibility as long as both contents are assessable for correctness.[10]

According to Crane (1992, 149ff.), the difference in content between judgement and perception is also responsible for the fact that perceptions may consistently persist in light of better knowledge. I second this view. Conflict without contradiction is due to difference in content in the first place. If the rational role characteristic of perception is best explained by difference in content in other cases, why should not the same difference account for conflict without contradiction? It would be implausible to treat this case as an exception from the rule.

We may then explain conflict without contradiction in the case of perceptual illusions, such as the Müller-Lyer illusion, as follows: Inferential relations holding between the contents of judgements include obvious logical relations. If you judge that *p*, you commit yourself to certain other judgements that are the obvious logical consequences of this judgement, among them the judgement that not (*p* & not-*p*). Hence you are not entitled to judge that *p* and that not-*p* simultaneously. Even if it should be psychologically possible to make simultaneous contradictory judgements, from a logical point of view, you would certainly make a mistake, as shown by Moore's so-called 'paradox': you would commit yourself to 'paradoxical' (in the sense of Moore) utterances of the form 'I know that the two lines are the same

9 As Crane points out, the idea to derive the notion of a concept from facts about the inferential relations among judgements or beliefs, is not new (see, e.g., Evans 1982; see also the introduction to Gunther 2003). Typically, the proponents of this idea appeal to Gottlob Frege's characterization of linguistic content as sense (*Sinn*).

10 Here we may consider the case of artistic content. Artistic content, the content of Musil's novel *Die Vollendung der Liebe*, for instance, is unlike the content of judgement or belief in that it cannot be fully expressed by a declarative sentence or a coherent system of such sentences. To capture the content of Musil's novel it will not do to say that Claudine, the main character, perfects the love to her husband by sleeping with another man. We cannot even understand this sentence, and our situation would not be improved by considering a whole system of declarative sentences meant to express the content of Musil's novel. For the content is by its very nature 'narrative', rather than 'propositional' (in the sense of Gottfried Gabriel 1991 and 1997). Nevertheless the content of Musil's novel can conflict with the judgement of a reader who is convinced that perfecting the love to one's partner has nothing to do with making love to perfect strangers. On Musil's view, provoking conflicts of such kind even makes the essence of art. As Musil well knows, this requires that the content of art presents a 'wirkliche Möglichkeit', a 'real possibility', rather than a mere fantasy, and is thus subject to a correctness condition.

length, but I do not believe it. I believe that one line is longer than the other.'[11] The Müller-Lyer illusion shows that there is no logical mistake involved in having a perception with a content that conflicts with the content of judgement and belief. It is not Moore-paradoxical to say 'I know that the two lines are the same length, but I do not see it. I see one line as longer than the other.' This is so because the conflict in question is no contradiction. It is no contradiction because contradictions arise only between the contents of inferential judgement and belief. Although perception is like belief (and unlike mere fantasy, for instance) in purporting to represent an actual state of the world, a real fact, Moore's paradox does not apply in cases like the Müller-Lyer illusion, the reason being that the content of perception is of such a kind that it is not subject to inferential constraints.

In the case of emotion, conflict without contradiction is to be explained in the same way. As its content is not subject to inferential constraints and does therefore not contradict the contents of other states, an emotion may persist and represent the world as being a certain way even when you judge, and know, that actually the world is not that way. This is not to deny that the content of an emotion or a perception sometimes *is revised* in light of better judgement and knowledge. How could this fact be accounted for? I shall address this question in the final section.[12]

Nothing that has been said so far about the non-inferentiality and the specific content of the emotions refers to their phenomenal character. Why should we assume that the representational content of an emotion is part of its conscious, subjective character, of what it is like to experience the emotion, as Goldie suggests? First, it seems that all states that have a representational content, but do not enter inferential relations, are phenomenal states. Both emotions and perceptions are occurrent states that can capture and occupy the subject's attention. As always, this does not imply that they are the objects of attention (see Sartre 1939, part III; Peacocke 1998; Goldie 2000, 63–70). We must distinguish between two ways in which a state can be conscious. In the first case, a state is 'non-reflectively conscious'. It is conscious in that it focuses the subject's attention on its intentional object without the subject's having to be aware of being in that state. In the second case, a state is 'reflectively conscious'. It is conscious in that the subject is conscious of having a conscious state as it is described as the first case: the subject is conscious of himself as being in a conscious state so that this state also becomes the object of attention. The first case does not entail the second. Thus you may fear a certain thing without being aware that you do. Nonetheless your fear involves a 'something-it-is-like' to experience that thing as dangerous. Both emotional and perceptual consciousness are consciousness of the world in the first place.[13]

---

11 Let me emphasize again that it is not the psychological possibility which is at stake here. Nor do I mean to say that people always draw the logical and material inferences which are available to them by what they believe.

12 I owe it to Neil Dillon.

13 My guess is that Roberts is unclear on whether emotions are always felt states because he does not make the distinction between reflective and non-reflective consciousness. Roberts (2003, 60) says that in 'feeling angry, I am aware of being angry'. But, as pointed out above, feeling angry need not involve reflective consciousness.

yet shrinks back in fear, he does not fully believe that the spider is harmless. There is conflict in such a case – the subject finds himself armed with reasons for and against the belief that the spider is harmless –, but the conflict does not amount to contradiction precisely because of the subject's uncertainty. For example, the subject might assign to his judgement that the spider is harmless a probability of 0.8, whilst the proposition that the spider is dangerous is held true with the counter probability of 0.2. Strictly speaking, the subject does thus not make either judgement, but holds true a certain probability distribution. The question, however, of how two judgements are logically related to each other is independent of the probabilities with which their contents are held true. Contradiction is not eliminated by the assignment of probabilities. The fact that the belief that *p* and the belief that not-*p* logically exclude each other is not affected when relative probabilities are assigned to them. Apart from this, probabilists have to deal with contradiction, if only in the case that the probabilities assigned to two conflicting judgements do not add up to 1. In this case, the probability distribution is inconsistent. Therefore, the judgementalist cannot resort to probability, or uncertainty, in order to account for the logical problem of conflict without contradiction. In any case, conflict without contradiction is not normally understood as a matter of uncertainty.

The Helm objection is helpful to clarify this point, but confusing in other respects. In particular, it confuses the attitude we take towards the content of an illusionary perception on the metalevel of judgement with the attitude that corresponds to that perception's content when it is experienced. To be sure, once we are certain that the two lines of the Müller-Lyer illusion are the same length, we judge that our visual experience deceives us. But this does not prevent the experience itself from having the appearance of truth. The occurrent perception of the two lines as being of a different length puts forward this content as correct. The content is representational, or verisimilar, as Roberts puts it in his early writings.

This is perfectly analogous to conflict between emotion and judgement. Once we are certain that an emotion is inappropriate we regard its content as a mirage in judgement. Yet in experiencing the emotion it still seems to us that the world really is as the emotion represents it to be.

In the case of emotion this is more obvious than in the case of perception because, in this case, conflict without contradiction entails practical as well as cognitive conflict. Typically, we are poised to act out of our emotions. In being afraid of a spider, typically, the subject is moved to act so as to avoid harm from the spider. If the spider is considered to be harmless, this action, as well as its underlying emotion, conflicts with the subject's better judgement. In the same way, the offended reactions of the person who is aware of her hypersensitivity occur in spite of what she thinks is appropriate. Practical conflict also explains why, by contrast with the perceiver of the Müller-Lyer illusion, the spider-phobic and the hypersensitive person are in trouble. Unlike the perception of the two lines as being of a different length, the occurrence of the fear and of the offence are not only beyond these persons' rational control; their emotions also have a distorting influence on their actions. That's why the spider-phobic and the hypersensitive person feel torn between judgement and emotion and go to the therapist.

However, the fact that there is also practical conflict in the case of emotion-judgement conflict, but not in the case of perception-judgement conflict, does not affect the analogy I have drawn between both cases with regard to cognitive conflict.[14] Cognitive conflict is prior to practical conflict, and the analogy means that both emotions and perceptions do not involve the attitude of regarding as merely fictitious.

5. True by Definition versus True by Default

Helm starts from the assumption that emotions are like judgements or beliefs in that they 'aim at truth'. But the way emotions aim at truth is said to be not the same as the way beliefs do. Therefore, Helm introduces the attitude of assent. His intuition is that this attitude must somehow be weaker than the attitude of regarding as true. The assent we give to the contents of our emotions must be so weak, on Helm's view, that these contents do not contradict the content of a conflicting belief. This is not to say that the content of emotion is regarded as less true than the content of belief (as in the case of uncertainty discussed above). Instead, I understand Helm as saying that assent to the content of emotion is a different, weaker attitude than the attitude of regarding as true. Contradiction is thus avoided with the help of incommensurable attitudes.

The problem with this account is not only that it is left unclear what kind of attitude assent should be. Furthermore, by offering an ad hoc solution to the problem of conflict without contradiction, Helm runs into deeper problems. Conflict without contradiction is at the focus of the debate not so much because we want to explain how an emotion that is considered inappropriate may yet persist in light of better knowledge. Understanding this case is crucial because it also allows to account for the more important case in which an appropriate emotion persists in light of a conflicting belief.

This is the point of the much-discussed example of Mark Twain's character Huckleberry Finn (see Bennett 1974; de Sousa 1987; McIntyre 1990; Tappolet 2000; Döring 2007). After having helped his friend Jim to run away from slavery, Huck decides to turn Jim in. But when he is given the opportunity to do so, he finds himself doing just the contrary. Instead of turning Jim over to the slave hunters, Huck lies in order to protect his friend. It is his sympathy and friendship for Jim which cause Huck to act this way, even though he does not endorse his emotional state but castigates himself for his weakness. Prior to Huck's acting against his better judgement is cognitive conflict between emotion and judgement: Huck's sympathy and friendship for Jim conflict with his normative judgement that he ought to turn Jim over to the slave hunters.

Now, Twain gets the reader to believe that, contrary to the case of illusion, it is in this case the emotion, rather than the judgement, which gets things right, and that his protagonist even deserves moral praise. This opens up the possibility that,

14 Practical conflict in the case of emotion-judgement conflict is due to the fact that the emotions are evaluations in light of the subject's concerns, rather than neutral representations of the world (see Döring 2007).

by persisting in light of conflicting judgements, our emotions need not deceive us, but may equally teach our judgements better. By introducing two incommensurable attitudes, the attitude of regarding as true and the attitude of assent, Helm excludes the possibility to reject a belief on the basis of an emotion. Firstly, if assent is weaker than regarding as true, how can assent given to the content of a state ever be strong enough to annul the content of a conflicting state which is regarded as true? Secondly, if assent and regarding as true are incommensurable attitudes, how could their contents logically relate to each other at all?

I have argued that Helm's objection against the analogy between emotion-judgement conflicts and perceptual illusions is mistaken. This analogy does not commit us to the claim that the attitude towards emotional content is that of regarding as merely fictitious. This is also Roberts's view. In his defence of the analogy in question Roberts (2003, 92) stresses that the notion of verisimilar content remains neutral on whether or not the subject regards the content as true. I think that he is right. The emotions do not 'aim at truth' in the sense that their content is regarded as true 'by definition'. To put it in de Sousa's (1987, 122) terms, truth is the formal object of belief, but it is not the formal object of emotion. Instead, each emotion type has its own formal object. The formal object is the property which the subject must necessarily ascribe to the object of his mental state in order to render the state intelligible (see also Kenny 1963). In the case of belief, the subject must regard the object of his belief as true. In the case of an emotion, fear, say, the subject must regard the object of his fear as dangerous or fearsome. Now, fearsomeness is typically ascribed to a real object (see de Sousa 1987, 116). This is in contrast to truth, which is the property of a proposition. As this contrast between formal objects already indicates, the emotions differ from belief in that content cannot be separated from attitude. Beyond regarding as fearsome, or annoying, or enviable, or admirable, or the like, there is no further attitude (conceptually) involved in experiencing an emotion. In the case of belief, which is a propositional attitude, all the representational work is done by the proposition involved. What makes a belief a belief, rather than a hypothesis, or a fantasy, or some other propositional attitude, is that the proposition involved is regarded as true. Belief is defined as the attitude of regarding a certain proposition as true. (In this sense the content of belief is regarded as true by definition.) The same proposition could be combined with other attitudes, which shows that attitude is independent from content. By contrast, representing a certain thing as fearsome *and* regarding it as fearsome *and* feeling fear of it is all the same. If a representational content is by its very nature a phenomenal content, then the representation of a real object is insolubly tied to the subject's attitude towards that object, and that is: to what it is like for the subject to experience the object the way he does. In other words, emotional intentionality is felt or affective intentionality.

This is, again, similar to the case of perception. To count as a red-perception, the subject must ascribe to the object of his mental state the property of being, precisely, red. Like fearsomeness, and unlike truth, redness is the property of a real object. Like fear, and unlike belief, a red-perception represents a real object which, by being represented as red, is experienced as red. Again, representational content and attitude are one and the same.

Having a representational content suffices for both the emotions and perception to qualify as an input to the rationalization of belief and action. To say that they have a representational content is to say that they are subject to a correctness condition. The correctness condition is provided by the formal object which corresponds to each emotion type and each type of perception.[15] In order to be correct, an emotion's or a perception's object must really have the property that is ascribed to it by the emotion or the perception. In order to be correct, fear must represent as fearsome what really is fearsome, and a red-perception must represent as red what really is red.

Having a representational content suffices for the emotions and perception to play their rationalizing role, especially so as we regard the content of emotion and of perception as true 'by default'. In contrast to the content of belief, the content of emotion and of perception is not necessarily (conceptually) regarded as true by the subject. But this need not prevent a thinker from recognizing his emotions and perceptions as reliable cognitive sub-systems, which function properly under standard conditions.

6. Epistemic Agency

I am here appealing to the picture of rational agency put forward by Karen Jones (2003), which I take to include 'epistemic agency' (on my view, judgements are actions, but I shall not argue for this view here). Crucial to this picture is the normative conception that we have of ourselves as agents. We do not regard ourselves as mere playthings of causes, but believe ourselves to be capable of guiding our judgements and actions via reasons. From our personal or first-person perspective it seems to us that we at least sometimes judge and act for reasons. In turn, this requires the capacity for reflection. The agent must see the reason as a reason for him. It does not matter in the first place whether what he sees as a reason actually is a reason. Of course, the rational agent is trying to act for what actually are reasons for him. But the only path of access to such 'good' or 'objective' reasons is through subjective assessment. Rationality understood as guidance via reasons is relative to what the agent sees as right.

Jones refers to this normative self-conception in order to make clear that, on the one hand, rational agency must not be reduced to what she calls 'reason-tracking'. As she puts it, our self-conception is not that of a 'reason-tracker', but that of a 'reason-responder'. A 'reason-tracker' is a creature that passively registers and reacts to environmental stimuli, much as a thermostat registers and reacts to changes in temperature. By contrast, a 'reason-responder' actively relates himself to and reflects on the information which he gains about his environment so as to treat informational inputs as reasons for judgement and action. A theory of rational agency, Jones says, must account for the fact that we see ourselves as 'reason-responders', rather than as mere 'reason-trackers'.

15 Since the content is specifically representational content, the correctness condition is also a success-condition. See also de Sousa (1987, 158ff.) for discussion of success-conditions.

On the other hand, Jones emphasizes that this does not commit us to the traditional picture of human agency either. Rejecting this picture as 'intellectualist', Jones maintains that it is also insufficient to accept exclusively the agent's judgements and beliefs as reasons in the normative sense. On her account, judging and acting for reasons involves the on-going cultivation and exercise of *all* abilities characteristic of a rational agent, and, according to her, this includes the ability to feel. The judgemental system is claimed to be just one reason-giving system among other such systems. This means that our judgements do not occupy a privileged role in the giving of reasons, but share that role with other cognitive sub-systems such as the emotions. As a consequence, an emotion need not first be sanctioned by a judgement in order to qualify as a reason for judgement or action. Against this requirement Jones argues that all that is needed for judging and acting for reasons is that the agent accepts some cognitive sub-system as reason-giving in the normal (or standard) case. 'Reason-giving' here means 'reason-tracking', and the different reason-tracking sub-systems are introduced as first-order mechanisms, which the agent can guide through his second-order conscious reflective and self-monitoring capacities. Rational guidance then amounts to 'regulative guidance': the rational agent directly controls his reason-tracking mechanisms, and indirectly his judgements and actions, by monitoring the mechanisms in order to step in when necessary, and where possible, to recalibrate them.

I fully agree with Jones that the emotions can provide us with reliable information about the world, that they are not inferior to judgements and beliefs in this respect, and that they can even teach our intellect better. In particular, I share her view that we treat our emotions as normally reliable cognitive sub-systems. We take for granted that the cognitive function is fulfilled under normal or standard conditions, and it does not normally occur to us to ask whether the conditions under which we experience the world are standard conditions. It is in this sense that we regard their representational content as true by default. In the default mode we operate so as to rely on the representational content of our emotions. For example, when we encounter a furious gorilla that escaped from his cage in the zoo, we do not normally wonder whether the conditions are really such as to justify fear.

This is again similar to the case of sensory perception. In the default mode it does not occur to us to ask whether the lighting conditions are really such as to justify the perception of a ripe tomato's redness. Like emotions, sense perceptions may be described as reason-giving or 'reason-tracking' sub-systems in the sense of Jones, which is to say that their representational content is regarded as true by default. Thanks to their salience in consciousness, emotions and sense perceptions are even more effective in the production of judgements and actions than judgements are. Other things being equal, seeing something to be the case with your own eyes is likely to produce a stronger belief than is yielded by inference or testimony to the same content. Similarly, an evaluation based on the representational content of an emotion will, other things being equal, produce a stronger belief than evaluations yielded by inference or testimony.

Nevertheless I think that, from the subject's first-person perspective, judgement and belief have a privileged position in the giving of reasons. The main problem

of Jones's analysis seems to me that it remains obscure how, in the case of conflict between two reason-giving sub-systems, the agent may decide which system is to be given preference. Clearly, it is not enough for him to accept that he may trust his emotions or sense perception in the normal case. Not every case is normal, and what the agent needs to know is which system he should rely on in this particular case. According to Jones, the decision is made by the agent's higher-order reflective and self-monitoring capacities. But how do they do so? Provided that these capacities do not consist of judgements again: which form do they assume?

Because the subject regards their content not merely as true by default but as true by definition, judgement and belief are privileged in the giving of reasons. This is not to deny that emotions can justify judgements and actions, and that they can do so without this requiring that the subject must authorize the emotion via a judgement in each particular case. Such a requirement would clearly be 'intellectualist' (or 'overintellectualizing'). But a rational agent must satisfy the condition that he *would* have authorized the emotion, had he had the opportunity to do so, or that he *would* affirm his emotion's content, were he asked to do so. In order to provide an agent with a reason here and now, the agent must at least 'hypothetically' authorize an emotion via a judgement. Thus, were you asked why you took flight from the gorilla, you would have to answer that the gorilla was dangerous, thereby endorsing the content of your fear by making a corresponding judgement. If an emotion's content is not even hypothetically endorsed, conflict with judgement arises, and a thinker who gives in to his feelings is irrational in the sense that, from his perspective, he has no reason to judge and do what he does.

Once we suspect that our emotions deceive us, we are ready to leave the default mode and to switch into a different mode. We withdraw our confidence in the experiential state, and the belief that it might immediately have generated in the default mode is revised. Therefore no contradiction arises. And yet there is conflict, if the experiential state, which is not subject to inferential constraints, persists. This *is* conflict because emotions have a representational content and are thus seen by the subject as being in the service of truth and knowledge, even though in a different way than judgements are.

Some of these conflicts allow resolution, others do not. If it is the judgement that goes wrong, resolution should be possible as soon as we realize its mistake. Here we may consider the example of Huckleberry Finn again. In this example the condition of at least hypothetic endorsement is not fulfilled at the moment of choice. In terms of his subjective perspective Huck is, at the moment of choice, no less irrational than Hume's protagonist or the spider-phobic, even though, from an objective point of view, it may be clear that it is Huck's emotional state, rather than his judgement, that represents the world correctly to him. As Huck does not even hypothetically endorse the content of his sympathy and friendship for Jim, from his perspective, he has no reason to judge that his friend ought to be protected from slavery and to act accordingly. This is why he castigates himself for his weakness. Nonetheless, Huck's emotional state is epistemically significant by virtue of its adequacy to the actual state of the world that they purport to represent. If emotions are not 'blind instigators', but can be adequate to actual states of the world, then rational agency

comprises the on-going cultivation and improvement of reasoning and knowledge through non-inferential emotion. Thanks to his persistent emotions Huck may come to know later that he made an error in judgement. In retrospect, he may recognize that it was right to protect Jim, even though he did not see this at the moment of choice. That is, he may recognize in retrospect that it was his emotional state, rather than his judgement, which represented the situation correctly to him. Since this does not follow from what he had hitherto accepted, the occurrence of his emotion may lead him to question his 'belief system' and to formulate new, better, more comprehensive beliefs.

If, in a conflict without contradiction, the emotion or the sense perception fails to represent things correctly, it may not be possible to resolve the conflict. For, as can be seen from the Müller-Lyer illusion or Hume's fear of falling, sometimes emotions and sense perceptions resist calibration. Still, in many cases they can be calibrated. For the case of sensory perception this is shown paradigmatically by the train illusion example: when sitting in a train waiting to depart from the train station and watching a train on a neighbouring track pulling out of the station, one gets the impression of moving oneself, even though it was in fact the train on the adjacent track that just started to move. By contrast with the Müller-Lyer illusion, it is easy to correct one's perception in the train illusion by calibrating it.[16]

I deliberately chose an example for the case of sensory perception, because it is sometimes believed that emotions are easier to calibrate than sense perceptions. As far as I can tell, this belief is wrong. The train illusion provides a counter example; and there are many more. The fact that both emotions and sense perceptions may or may not be open to calibration also explains why the content of an emotion or of a perception may or may not be revised in light of better judgement and knowledge. To complete my account of conflict without contradiction: if an emotion is revised in light of better judgement and knowledge, this is not a matter of contradiction but of calibration.

References

Alston, William P. 1967. 'Emotion and Feeling'. In: Edwards, Paul (ed.). *The Encyclopedia of Philosophy*. New York: Macmillan. Vol. 2. 479–86.

Bennett, Jonathan. 1974. 'The Conscience of Huckleberry Finn'. *Philosophy* 49: 123–34.

Crane, Tim. 1992. 'The Nonconceptual Content of Experience'. In: Crane, Tim (ed.). *The Contents of Experience. Essays on Perception*. Cambridge: Cambridge University Press. 136–57.

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.

Döring, Sabine A. 2003. 'Explaining Action by Emotion'. *Philosophical Quarterly* 211: 214–30.

16 Science may offer an explanation of why some perceptions and emotions resist calibration, whereas others are open to it. Thus the Müller-Lyer illusion is explained by the size constancy mechanism of our visual system.

Döring, Sabine A. 2008. *Gründe und Gefühle. Zur Lösung 'des' Problems der Moral*. Berlin: de Gruyter.

Döring, Sabine A. 2007. 'Seeing What to Do. Affective Perception and Rational Motivation'. *Dialectica* 61: 363–94.

Döring, Sabine A.; Peacocke, Christopher. 2002. 'Handlungen, Gründe und Emotionen'. In: Döring, Sabine A.; Mayer, Verena (eds). *Die Moralität der Gefühle*. Special Issue of *Deutsche Zeitschrift für Philosophie*: 81–103.

Evans, Gareth. 1982. *The Varieties of Reference*. Ed. John McDowell. Oxford: Oxford University Press.

Gabriel, Gottfried. 1991. *Zwischen Logik und Literatur. Erkenntnisformen von Dichtung, Philosophie und Wissenschaft*. Stuttgart: Metzler.

Gabriel, Gottfried. 1997. *Logik und Rhetorik der Erkenntnis. Zum Verhältnis von wissenschaftlicher und ästhetischer Weltauffassung*, Paderborn: Mentis.

Goldie, Peter. 2000. *The Emotions. A Philosophical Exploration*. Oxford: Oxford University Press.

Goldie, Peter. 2006. 'Getting Feelings into Emotion in the Right Way'. Unpublished manuscript.

Greenspan, Patricia S. 1981. 'Emotions as Evaluations'. *Pacific Philosophical Quarterly* 62: 158–69.

Greenspan, Patricia S. 1988. *Emotions and Reasons. An Inquiry Into Emotional Justification*. New York: Routledge.

Griffiths, Paul E. 1997. *What Emotions Really Are. The Problem of Psychological Categories*. Chicago: Chicago University Press.

Gunther, York H. (ed.). 2003. *Essays on Nonconceptual Content*. Cambridge, Mass.: MIT Press.

Helm, Bennett. 2001. *Emotional Reason. Deliberation, Motivation, and the Nature of Value*. Cambridge: Cambridge University Press.

Helm, Bennett. 2006. 'Affective Intentionality – Holistic and Multiply Layered'. Unpublished manuscript.

Hume, David. [1739/40] 1978. *A Treatise of Human Nature*. Ed. Lewis A. Selby-Bigge and Peter H. Nidditch. Oxford: Clarendon Press.

Jones, Karen. 2003. 'Emotion, Weakness of Will, and the Normative Conception of Agency'. In: Hatzimoysis, Anthony (ed.). *Philosophy and the Emotions*. (*Royal Institute of Philosophy Supplement* 52). Cambridge: Cambridge University Press. 181–200.

Kenny, Anthony. 1963. *Action, Emotion and Will*. London: Routledge and Kegan Paul.

Lacewing, Michael. 2004. 'Emotions. An Essay in Aid of Moral Psychology'. *Journal of Moral Philosophy* 1: 105–8.

Lyons, William. 1980. *Emotion*. Cambridge: Cambridge University Press.

McDowell, John. 1985. 'Values and Secondary Qualities'. In: Honderich, Ted (ed.). *Morality and Objectivity. A Tribute to J.L. Mackie*. London: Routledge and Kegan Paul. 110–29.

McIntyre, Alison. 1990. 'Is Akratic Action Always Irrational?' In: Flanagan, Owen; Rorty, Amélie O. (eds). *Identity, Character, and Morality. Essays in Moral Psychology*. Cambridge, Mass.: MIT Press. 379–400.

Mulligan, Kevin. 1998. 'From Appropriate Emotions to Values'. *Monist* 81: 161–88.

Musil, Robert. [1930–42] 1978. *Gesammelte Werke*. Reinbek: Rowohlt.

Peacocke, Christopher. 1992. *A Study of Concepts*. Cambridge, Mass.: MIT Press.

Peacocke, Christopher. 1998. 'Implicit Conceptions, the *A Priori*, and the Identity of Concepts'. In: Villanueva, Enrique (ed.). *Concepts*. (*Philosophical Issues* 9.) Atascadero, Ca.: Ridgeview. 121–48.

Prinz, Jesse J. 2004. *Gut Reactions. A Perceptual Theory of Emotion*. New York: Oxford University Press.

Pugmire, David. 2007. 'Go on, *Say* It!' Unpublished manuscript.

Reisenzein, Rainer; Horstmann, Gernot. 2006. 'Emotion'. In: Spada, Hans (ed.). *Lehrbuch Allgemeine Psychologie*. 3rd edition. Bern: Huber. 435–500.

Roberts, Robert C. 1988. 'What an Emotion Is. A Sketch'. *Philosophical Review* 97: 183–209.

Roberts, Robert C. 2003. *Emotions. An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.

Rorty, Amélie O. 1984. 'Aristotle on the Metaphysical Status of *Pathe*'. *Review of Metaphysics* 84: 521–46.

Sartre, Jean-Paul. 1939. *Esquisse d'une théorie des émotions.* Paris: Hermann.

Solomon, Robert C. 1993. *The Passions. Emotions and the Meaning of Life*. Indianapolis: Hackett.

Tappolet, Christine. 2000. *Émotions et valeurs*. Paris: Presses Universitaires de France.

Tappolet, Christine. 2003. 'Emotions and the Intelligibility of Akratic Action'. In: Stroud, Sarah; Tappolet, Christine (eds). *Weakness of Will and Practical Irrationality*. Oxford: Clarendon Press. 97–120.

Wiggins, David. [1987]. 'A Sensible Subjectivism'. Reprinted in: Darwell, Stephen; Gibbard, Allan; Railton, Peter (eds). 1997. *Moral Discourse and Practice*. Oxford: Oxford University Press. 227–44.

This page intentionally left blank

Chapter Five

Epistemic Immediacy and Reflection

Daniel Dohrn

Starting from the historic example of Descartes' theory, epistemologies stressing the relevance of reflective assessment are contrasted with approaches shifting the emphasis to the role of immediate affective valuations. It is discussed whether the insights of the latter demand a reshaping of normative terms used to appraise epistemic activity. Two claims have to be distinguished: Firstly, such immediate factors are indispensable ingredients of a blameless epistemic activity. Secondly, they have as such a justificatory role of their own. Whereas the first claim is endorsed, the second is rejected. It is proposed instead that justification comes with reflectivity.

1. The Cartesian Ideal of Full Reflectivity

I will start with a brief historical excursion into Descartes' epistemology of emotions. His account serves as a starting point to discuss the epistemic role of emotions. I will concentrate on their immediacy and an eventual demand to reflectively assess them. Descartes' epistemology of emotions can be derived from his theory of the passions. He describes the natural world view which does not involve epistemological reflection. Human beings are naturally disposed to immediately react to certain stimuli, especially sense perceptions. Incited by these stimuli, they form certain beliefs and perform certain actions. Upbringing and education might play some role in forming customs, habits and prejudices. So far nothing in this picture precludes our beliefs from being true. But nothing precludes them from being wrong either. On reflectively scrutinizing this natural world view, Descartes comes to an unfavourable verdict. It contains many inconsistent convictions and conflicting principles of action. Descartes proposes a well-known radical solution: Doubt once in a lifetime all beliefs against which even the slightest reason of doubt can be forwarded (Descartes [1644], 5)! He sometimes calls the metaphysical doubt a fiction and stresses that it must not be extended to suspending urgent actions. Yet he emphasizes that one is to break one's customs, habits and natural ways of acquiring beliefs (Descartes [1641], 22; Descartes [1649], § 211). This procedure is to ensure that we do not accept any proposition *merely* because of a natural or habitual disposition to believe it. Any belief which is to be accepted must be accounted for. Of course it would be too complicated to separately justify each and every single belief. Rather beliefs must be classified according to the natural ways in which they were acquired. Then the admissible ones must be singled out. The aim is to ensure the possibility of explicitly accounting for any single belief, of giving reasons why the ways in which it was acquired are acceptable.

Although he does not really focus on emotions in his genuinely epistemological discussions, Descartes develops a rather similar view of their role in our world view and behaviour. Emotions provide immediate impulses to think and to act in certain ways. Their function is to make one believe and do what is useful. But they fulfil their task not by themselves. As in the case of sense perception and other immediate sources of belief acquisition, one must develop a critical attitude which allows to assess and to control emotions. According to Peter Schouls, Descartes proposes an operation similar to the method of doubt. He aims at conscious self-control or self-mastery. Just as one suspends one's beliefs, one aims at interrupting the physiological mechanisms which cause emotions: 'To achieve self-mastery these "natural" links must first of all be broken' (Schouls 1989, 165).

I consider Schouls's interpretation too strong. In fact Descartes claims two things: Firstly, one must undergo a special training. This training ensures that no emotion can influence one's thoughts and actions without one's being able to resist it. Contrary to Schouls, one does not need to actually break all natural physiological links between sensory stimuli and emotions. One must merely be able to do so. Secondly, one must classify emotions in order to elucidate whether they are useful or harmful in certain situations. We can summarize Descartes' claim as follows: Although he admits that it is sometimes necessary to act blindly, he denies that it is appropriate to follow one's immediate emotional valuations. To be right to follow an emotional valuation in a particular situation, one must have done two things. Firstly one must have classified the valuation and the situation before. Secondly one must have formed an explicit judgement that following such an emotional valuation in such a situation is right. The criterion of this examination is set by a belief system which conforms to Descartes' epistemology.

Now it could be asked what use to make of emotions at all under these circumstances. Descartes must grant that emotions play some *immediate* role, a role not always mediated by reflection. Otherwise they could be replaced by a practical syllogism yielding what would be useful in a given situation. Thus his proposal must be understood as follows: Reflection on natural emotional responses and sufficient training must allow to establish a kind of 'second nature'. Immediate 'animal' emotions must be replaced by emotions conforming to an assessment. This assessment must show which emotional response would be appropriate in a given situation. Emotions form a 'second nature' because they are immediate. There is no need to assess them anew every time they occur. They immediately motivate to think and to act as appropriate in a given situation. They abbreviate chains of reasoning and measures of training. In contrast to Schouls's interpretation, many immediate emotional responses may rest untouched and be integrated into the second nature, provided that one could resist them and that a critical assessment shows their acceptability.

Why does Descartes demand that we regulate our natural perceptual and emotional responses? He denies that these responses are sufficiently useful or truth-conveying. This is shown by the many contradictory beliefs and incompatible actions they lead to. But imagine the following: They are by themselves sufficiently useful or truth-conveying. Or they mostly are erroneous but give rise to a sufficiently coherent web of beliefs and actions. In this case we have no motive within that web of beliefs or actions to be suspicious about them. Could there still be reasons for

such a scrutiny? This question is somewhat speculative. But an answer could lie in Descartes' idea that we are responsible for our activities. To him responsibility depends on accepting, i.e. consciously and voluntarily making a proposal one's own. But if it is required that such a decision is not a matter of luck, it should be informed. It should be guided by conscious reasons. Thus we may ask for a reason why to trust our immediate valuations.

However, we do not have to invoke such strong notions as responsibility. There could be a simpler reason for critical assessment. Imagine that someone in the natural situation Descartes describes accidentally asks herself: Is it an appropriate cognitive behaviour to form beliefs and act the way I do? When looking for an answer, one could start from reflecting on one's natural goals. It is an epistemic aim to correctly answer interesting questions. Now one could ask further, perhaps generalizing some special experience of failure: To what extent are the ways in which I am acquiring beliefs suitable to reach my epistemic aims? The natural reaction to such a question would be to examine these ways of acquiring beliefs. When our inquirer notices that her beliefs partly depend on her emotive responses, the natural reaction is to examine these emotive responses. This general question of suitability may arise from a specific case of epistemic activity. But if it is appropriate to raise this question in one case, it is appropriate in any other case unless there is a reason against doing so. Since it is too complicated to provide such an account case by case, the natural consequence is to classify kinds of emotional valuations and then to assess them generally. Such an inquiry should not be a purely theoretical or hypothetical matter. One should respond to results confirming or refuting the suitability of immediate behavioural patterns. The natural response is to adapt immediate behavioural patterns in a way that best suits one's aims. It is plain that there is something to be gained by raising such an issue. It opens the field of epistemological scrutiny which Descartes can claim to have pursued consequently.

2. The Ideal of Reflectivity in Current Epistemology

The ideal of reflectivity is present in current epistemological debate. Let us just briefly look at three paradigmatic positions to see how the demand for reflectiveness can be grounded. Jonathan Cohen opposes immediate belief and reflective acceptance. He claims that reflection is a prerequisite of methodically guided and unbiased belief formation. Linda Zagzebski opposes immediately formed beliefs and the ability of their voluntary reflective assessment. She argues that reflectivity is a requirement of responsible belief formation. Ernest Sosa opposes immediate and reflective knowledge. He argues that reflectivity as such provides a superior epistemic standing.

The first position to be summarized is Jonathan Cohen's. He distinguishes belief and acceptance. Belief involves immediately feeling something to be true (Cohen 1989, 368). In contrast, accepting something to be true is conscious and voluntary. Epistemic agents are responsible for what they accept and not for what they believe (Cohen 1989, 370). According to Cohen science requires acceptance. Beliefs which are not explicitly accepted should better not play a role in scientific practice. The

reason is that immediate affective responses like beliefs threaten the rationality of scientific reasoning. It cannot be ensured that they are formed according to the methodological requirements of science:

> Perhaps, there is not much harm in the scientist's in the end believing that *p* as well as accepting that *p*. But he would do better to school himself into practising a greater intellectual detachment. … Also, in the establishment of a belief that *p*, some factors might be influential in the black-box of the scientist's sub-conscious mind which he would reject as irrelevant or prejudicial if they came up for consideration before the tribunal of conscious acceptance. (Cohen 1989, 385)

Cohen pictures scientific knowledge as due to sources which must hold before a reflective tribunal. The problem of immediate affective valuations is that it is open whether they would survive such a tribunal. Thus they should be discarded when doing science. However, the possibility of such a fully enlightened practice is doubtful for the following reason: It is questionable whether sufficient standards of acceptance remain after discarding all immediate affective valuations. It is irresponsible to accept beliefs blindly. If on the one hand the basis of acceptance is not always a consciously accepted conviction, it must be something immediate like a mere belief, a perceptual experience, an affective valuation or so. But with such a basis the factors which the epistemic agent should ignore come into play again. If, on the other hand, the basis must be some consciously accepted belief, then a regress or a circle of such beliefs threatens. The regress is vicious when the beliefs providing the basis must be consciously accepted one by one, because we cannot actually accept an infinite series of beliefs in this way. But the circle also is vicious. It would be a strange move to consciously and explicitly base a belief on a series of accepted beliefs containing that very belief. Perhaps there are other possibilities. Be that as it may, Cohen owes an account of justification meeting his standards.

The second position to be discussed is Linda Zagzebski's. She interprets epistemic activity according to the pattern of responsible action. She claims that beliefs should be acquired in a responsible manner. Although she rejects strong Cartesian voluntarism, she emphasizes that the immediacy of belief acquisition must be reconciled with the ideal of epistemic responsibility. When beliefs are formed immediately, arise spontaneously, they cannot be attributed to one's agency. In Zagzebski's opinion being responsible for one's beliefs requires the possibility to reflect on them and in consequence to accept or to reject them consciously, to make them one's own or not:

> We have seen that the non-voluntary acts and beliefs of agents can differ in important ways from events that are produced by non-agents. If it is the act of belief of an agent, the agent's subsequent reflectiveness makes it voluntary on the second level. The agent either does or does not make the belief her own. Even non-voluntary acts/beliefs can therefore earn the agent credit (or blame), and in the case of beliefs, they may constitute knowledge. (Zagzebski 2001, 154)

Zagzebski does here not explain what relation to acts of reflection makes a non-voluntary belief knowledge. A very strong reading would be that any praiseworthy

belief or action demands actually exerting reflective control. I am not sure whether Zagzebski would really be willing to accept such a strong conclusion. It would probably be too strong a requirement if epistemic success were always to depend on actual reflective choice. Rather the agent must have control or mastery over her cognitive powers, a capacity to reflect and make a belief her own or not. Yet it would be strange to hold that epistemic success solely depends on the availability of a blind decision, a decision to make a belief one's own or not. If they are to win praise, reflective assessment and choice should be enlightened. They should involve comparing a belief to standards of belief acquisition. But then such standards must be available. If they are immediate, then they impose themselves on the epistemic agent, and if immediacy cannot be reconciled with responsibility unless reflective control is available, the problem Zagzebski addresses looms again. Perhaps some sort of 'access internalism' could be sufficient for responsibility (Boghossian and Williamson 2003, 228): It must be possible to develop standards allowing to reflectively assess anything which is immediate in belief formation. This option of a mere access internalism is not open to Cohen. To him conscious and explicit acceptance is a prerequisite of methodical and unbiased belief formation. Since such an acceptance should not be blind, the standards on which it rests must be conscious, too. In that respect Cohen's claims are stronger than Zagzebski's.

Although many normative notions like justification are used in epistemology, one might feel uneasy about transferring a notion like responsibility from moral philosophy to epistemology. The same holds for the strong assimilation of epistemic activity to a full-fledged conscious practice with all its implications of voluntariness and choice. But even if one refrained from such an assimilation, one could still ask whether natural or unconsciously acquired habits of belief formation are in tune with the demands of proper epistemic activity. This question may arise without any concrete motive for suspicion. A theory which does not depend on strong normative notions like responsibility is Ernest Sosa's. In his interpretation of Descartes, Sosa draws a distinction between animal or immediate knowledge and reflective knowledge. An activity which is right upon reflection is better than a right but unreflective activity: 'It is better to believe and act in ways that are reflectively right than in ways that happen to be right but unreflectively so' (Sosa 1997, 232).

Sosa's claim is somewhat more modest than Zagzebski's and Cohen's. It is better to be reflectively right, but reflective control however understood is not a *conditio sine qua non* of epistemic success. This modest position allows to avoid some of the difficulties mentioned so far. Sosa attributes to Descartes an ideal of reflective knowledge as a steadily growing coherent network which arises from our everyday epistemic practice (a historically questionable view):

> [Descartes] meditates along attaining the kind of epistemic justification and even 'certainty' that might be found in an atheist mathematician's reasonings, one deprived of a world view within which the universe may be seen as epistemically propitious. … Absent an appropriate world view, however, no such reasoning can rise above the level of *cognitio* [animal knowledge]. If we persist in such reasoning, nevertheless, enough pieces may eventually come together into a view of ourselves and our place in the universe that is sufficiently comprehensive and coherent to raise us above the level of mere *cognitio* into the realm of higher, reflective, enlightened knowledge, or *scientia*. (Sosa 1997, 240)

The idea of being reflectively right involves that reflection has an impact on one's actions and beliefs. It may be questioned that when something appears right upon reflection, it should always replace what appears right immediately. But in any case a cognitive fissure between both appearances should invoke further reflection. Since affective valuations shape immediate patterns of acting and acquiring beliefs, it would be better if they were subject to critical assessment, too. If they are judged to promote false ways of acting and acquiring beliefs, it is rational to reform them in order to avoid such negative consequences.

Now if it is an epistemic aim, or even – as in some strong accounts of epistemic responsibility – an epistemic requirement, to be reflective in acting and acquiring beliefs, then the Cartesian idea of assessing one's 'first nature' (the natural patterns of habits, prejudices, sense perceptions, emotions etc. forming our thinking and acting – in short, everything that is immediate) is an extreme but rational consequence. Unless there are considerations against such an idea, one should engage in the Cartesian enterprise. This does not necessarily mean to endorse the extreme consequences of hyperbolic doubt and doxastic voluntarism, although such consequences may prove to be necessary conditions for fully carrying out the Cartesian program. Nor does it necessarily mean that the critical examination must be carried out before one is able to act or believe rightly.

3. The Role of Epistemic Immediacy in Current Epistemology

Even if we are sympathetic to this ideal of unlimited reflectivity, we might still ask in how far it can and should be pursued. Philosophers like Ronald de Sousa and Christopher Hookway argue that the usefulness or even the possibility of such an enterprise may be very limited. They provide a thorough analysis of the role of emotions. In the following section I will summarize the reasons provided by de Sousa and Hookway why emotions are indispensable ingredients in any human activity. Then I will discuss the consequences they draw for the role of emotions. Two claims will be distinguished, both of which limit demands for epistemic reflectivity as hitherto considered: Firstly, de Sousa's claim that emotions are necessary prerequisites of cognitive activity without a fully reflective assessment of their function being possible; secondly, Hookway's claim that it is appropriate to base epistemic activity on immediate emotional valuations. Whereas the first claim will be accepted, the second one will be contested.

De Sousa envisages what he calls the philosopher's frame problem. In order to elucidate this problem, he refers to the example of a sophisticated robot which is able to perform a lot of human-like functions but has nothing like human emotions. The robot can draw inferences from the input it gets. But if the input is sufficiently complex, there will be a great number of inferences that could be drawn. If the robot is to draw all these inferences, it cannot operate efficiently. The robot will constantly be busy drawing irrelevant conclusions, unless there is some pattern of salience that allows for identifying the relevant ones (de Sousa 1987, 192–3). The problem with programming such a pattern of salience is the following: The robot has to draw a conclusion first before applying a standard of relevance to it. Sometimes a program

may exploit general features of irrelevant conclusions which can be apprehended without drawing these conclusions. A robot embodying this program does not actually need to draw these conclusions. But often there may be no such features. Then there must be some capacity to simply ignore certain reasonings, preferably the irrelevant ones. According to de Sousa, this capacity of singling out issues of potential attention defines the role of emotions:

> (BH1) *New Biological Hypothesis 1*. The function of emotions is to fill gaps left by (mere wanting plus) 'pure reason' in the determination of action and belief, by mimicking the encapsulation of perception: it is one of Nature's ways of dealing with the philosophers' frame problem.
>
> (BH2) *New Biological Hypothesis 2*. Emotions are species of determinate patterns of salience among objects of attention, lines of inquiry, and inferential strategies. (de Sousa 1987, 195–6)

If de Sousa is right, our capacities for explicitly assessing the role of an emotion are limited. An assessment would require to perform the task which the robot (guided not by emotions but merely by 'pure reason') fails to perform.

Does de Sousa hold that emotions are reliable in singling out the *right* conclusions, that the encapsulation by emotions leads to better results than a random process of encapsulation? The problem of such a claim can be illustrated by another problem of encapsulation which de Sousa's theory of emotions is to solve. In Bayesian decision theory it remains open what to do in situations of indifference among several mutually exclusive bets whose expected utilities are equal (de Sousa 1987, 193). This gap could simply be filled by some random process. I see no criterion to decide what could do better than such a process. If emotions are to fill this gap – as de Sousa holds – they may function like such a process. If it is claimed that emotions work better than a random process, we are in the situation the developer of the robot had to face. In order to explicitly show that emotions manage to single out the right issues of attention, one must already attend to them, for example draw and assess irrelevant conclusions. If 'pure reason' leaves open the issues settled by emotions, how is the achievement of the latter to be assessed? But if the encapsulation were blind, sceptical problems could follow. How can we confidently endorse any belief if it is open whether we notice factors relevant to its being true? Sometimes we may be able to conclusively rule out such factors –but nothing ensures that such favourable circumstances obtain sufficiently often. Such worries are not restricted to an internalist theory. Someone endorsing an externalist theory may also be uneasy if encapsulation conforms to standards of reliable belief formation.

Hookway sets out to defend a claim much stronger than de Sousa's. He replaces the Cartesian by a Humean picture. Responsibility involves tracing epistemic behaviour to traits of a suitable stable epistemic character but not necessarily reflection and voluntary action. Emotions play a much broader role than in de Sousa's 'new biological hypotheses'. Immediate Emotions are not only indispensable in human acting and thinking but play a crucial role in human rationality. It is *rational* to think and to act according to emotional patterns. From a more epistemological perspective

Hookway defines several framing problems. He refers to Quine who doubts that the path leading from individual sensory stimuli to inductive generalizations can be explained by explicit rules (Quine 1960, 19). The simplest generalization explaining sensory stimuli must be felt and cannot be further assessed. There is no inductive logic that would make such an assessment possible. Hookway also refers to Goodman's riddle of induction (Goodman 1954, 74): Having observed sufficiently many green emeralds, we tend to infer that all emeralds are green. However, we can define a concept 'grue'. Something is grue if, and only if, the following holds: It has either been observed until now and is green or has not yet been observed and is blue. Goodman asks why we may infer from our observations that all emeralds are green, but not that all emeralds are grue. Hookway argues that we cannot further explain why the first inference in contrast to the second one is sound: '… [we] find it compelling and we are right to do so. But how and why this is the case is not something that is transparent to us, it is not something that we can bring to full reflective consciousness' (Hookway 2003, 81).

When we find a generalization 'emeralds are green' more compelling than a generalization 'emeralds are grue', this may be due to an affective or emotional response which cannot be further explained. We immediately feel comfortable with the idea of emeralds being green and feel uncomfortable with the idea of emeralds being grue. Hookway's claim is that we rightly give in to these feelings. Because of the central role such inductive generalizations play in our epistemic practice, we cannot gain a fully reflective account of this practice. An analogous case are law-like ceteris-paribus-generalizations. It is argued that understanding such generalizations depends on a background of immediate feelings of relevance (Lange 2002). In terms of the Cartesian picture developed above, such generalizations form part of our first nature without our being able to elucidate them. We lack the explicit standards that would be required to give such a critical analysis. If such an assessment were a prerequisite of forming the respective generalizations, we would have to give up making them. Hookway does not see this impediment as a limit of rationality. He rather claims that it is rational, right and responsible to feel the way just indicated and to base inductive generalizations on these feelings. If Hookway is right, science in Cohen's sense which excludes relying on immediate affective valuations is impossible.

Furthermore, Hookway argues that requiring explicit reasons for every belief leads to an infinite regress. The same holds for evaluating the significance of the relationship between those reasons and beliefs resting on them (Hookway 2003, 88). The regress can be avoided if one can be justified despite not being able to provide a justification, if there are admissible ways of forming beliefs immediately, for example by automatically reacting to certain stimuli or 'blindly' following paths traced by emotional patterns. Hookway emphasizes that explicit standards of epistemic behaviour cannot stand alone. The best they can do is to provide rules of thumb which must be embedded in the whole network of one's immediate behavioural dispositions.

Let us summarize our discussion. We are now faced with two opposing epistemological tendencies. On the one hand, there are positions like those of Descartes, Cohen, Zagzebski and Sosa stressing the importance of reflectivity.

Their main arguments are the following: Reflectivity is necessary for a methodical epistemic activity which avoids biased or prejudiced judgement. It is a prerequisite of responsible epistemic activity. An activity which is reflectively right is better than an activity which is unreflectively right. On the other hand, there are positions like those of de Sousa and Hookway. Their main arguments are that there are indispensable factors of belief formation which must rest on immediate affective valuations. Such emotional factors do not allow explicit reflective scrutiny. Some especially relevant tasks they must fulfil are the following: Gaps left by 'pure reason' must be filled. Sound inductive generalizations must be distinguished from unsound ones. In order to avoid a justificatory regress, it must be possible to be justified without being able to provide an explicit justification. Explicit epistemic rules must be embedded in a network of emotionally guided activities.

Philosophers who attach importance to epistemic immediacy emphasize the substantial role emotions play in human practice and especially in epistemic activity. Such an epistemological consideration of the general role of emotions already meets some demands of reflectiveness. What is the difference between such an appraisal and accounts stressing the role of reflectivity? An advocate of immediacy like Hookway does not take a reflective assessment to be a prerequisite of epistemic success. Reflectivity is not even described as an epistemic aim. In contrast, as we will see in the next section, Hookway emphasizes the merits of not reflecting too much. If there is any merit in explaining the role of emotions, this merit apparently consists not so much in correcting or justifying emotions or in ameliorating our epistemic practice. Rather it consists in curing scepticism or exaggerated justificatory demands and in answering a question that may be as interesting as any other scientific question. It is not valuable or desirable as such to transform immediate affective valuations guiding epistemic behaviour into explicit reflective judgements of value. The paradigm of epistemic activity is shifted from a reflective and conscious one in accordance with explicit rules to a largely non-reflective, non-conscious and non-explicit one.

I do not want to question the indispensability of emotions. Surely acknowledging their necessity represents an important advance towards a realistic picture of epistemic activity. However, the question is what impact this insight should have on the structure of epistemic appraisals and the use of evaluative terms like 'blameless', 'good', 'right' or 'justified' in epistemology. This question will be further pursued in the next section.

4. Epistemic Immediacy and Epistemic Entitlement – the Cartesian Ideal Rejected

In the remaining sections I want to discuss the entitlements allegedly involved in our immediate affective evaluations. I will concentrate on Hookway's theory as it paradigmatically manifests the claim I want to dispute: Affective evaluations are not only an indispensable part of any epistemic activity, it is appropriate to follow them. Hookway develops these claims in his argument against scepticism regarding the epistemic role of emotions. Thus his discussion of scepticism will be resumed first:

Sceptical doubts arise from considerations leading to distrust affective evaluations. But it is appropriate to trust them, provided one is affectively integrated and one identifies emotionally with one's immediate evaluations. Then, three examples will be discussed to question Hookway's anti-sceptical argument and the claims based on it. They combine, on the one hand, affective integration with a deviant epistemic behaviour guided by affective evaluations, and affective disintegration with appropriate epistemic behaviour on the other. These paradigm cases show that affective integration is neither necessary nor sufficient for appropriate epistemic behaviour. But since there is no evidence that our situation is different, we cannot maintain that it is appropriate to follow our affective valuations.

Let us now take a closer look at Hookway's anti-sceptical argument. Hookway shows how scepticism emerges from an awareness that 'our affective valuations will lead us wrong' (Hookway 2003, 89). The result seems to be a general justificatory requirement regarding these affective valuations: 'What justifies our confidence in our practice? We might argue that unless we can give a justification of our trust in our emotional evaluations, we should feel anxiety about the ordinary beliefs that depend on them' (Hookway 2003, 90).

Yet Hookway rejects such a general demand for justifying our emotional valuations when they are challenged: So long as we confidently embrace a practice, it is appropriate to stick to it without giving a justification first: 'So long as we identify with our emotional evaluations, this quite properly produces doubt of most considerations that question them. Our confidence is untouched, and this is fully *appropriate*' (Hookway 2003, 90, my emphasis).

Hookway does not directly refute sceptical arguments. He rather doubts that they should be taken seriously. The sceptic asks questions that are not appropriate. She ignores the adequate emotional responses that would allow her to avoid questioning too much and thus must be 'cured'.

Hookway refers to 'us'. He presumably must hold that his analysis generally applies to anyone. His claim that following affective valuations is appropriate or right without further qualification implies that affective valuations as such, without any reflection taking place, carry some positive epistemic status, say positive presumption. Of course this positive presumption may be overridden by situation-specific reasons of doubt. We may express this claim as follows:

(CH) An epistemic agent is (defeasibly) justified in following her affective evaluations provided she feels confident of them.

I want to argue that only a weaker claim follows from the evidence about the role of emotions collected by Hookway. One is blameless in sometimes relying on immediate affective valuations because reflection is naturally limited. In the same sense a madman or a dreamer might blamelessly identify himself with Napoleon. We can express this weaker claim as follows:

(CD) We are blameless in following our immediate affective valuations unless a reflective assessment of them is available.

This claim is reconcilable with the evidence Hookway provides. It leads to the same results regarding our emotional evaluations unless a reflective assessment is available. If one is available, however, the pragmatic consequences of my principle (CD) differ from Hookway's principle (CH): (CD) demands that, in order to attain justification, we should develop such a reflective appraisal and act according to it.

In order to substantiate my claim, I now want to discuss three examples which cast doubt on Hookway's stronger principle (CH): They suggest that a feeling of confidence is neither necessary nor sufficient for epistemic justification. Some cases which are discussed in the literature indicate difficulties in identifying immediate patterns of epistemic behaviour being right or appropriate with their being emotionally approved. This holds even if this behaviour is granted to be externally successful to a sufficiently high degree. Abrol Fairweather provides one of them by imagining '… Conrad, the doxastic conformist. Conrad's primary cognitive goal is that a class of his beliefs largely overlap with that of Mr. Cool. If Mr. Cool believes *P*, then Conrad will believe *P*' (Fairweather 2001, 74).

This example is much more appealing in Hookway's framework than in a framework of explicit epistemic standards and motives, because it would be quite eccentric to entertain the conscious aim to believe whatever Mr Cool believes. It is not difficult to imagine Conrad being affectively so attached to Mr Cool that he habitually accepts without examination what Mr Cool believes. Now imagine further that Mr Cool is a diligent and thus very reliable epistemic agent. Conrad can be constructed to be sufficiently coherent in his affective patterns of behaviour and as sensitive to this coherence as one might wish. He is so attached to Mr Cool that his desire for true beliefs and all other emotive valuations prompting an independent epistemic behaviour have almost vanished. One could doubt the possibility of such a deviant pattern of epistemic behaviour. But one would have to explain why it is impossible. Such an explanation could involve questioning the coherence of Conrad's affective valuations. His eagerness to find the truth may conflict with his eagerness to conform to Mr Cool in possible situations in which Mr Cool's beliefs go awry. But his affective attachment to seeking the truth can be very weak. Since the strength of one's valuations must play a role in assessing their coherence, and since it would be too ambitious to require perfect coherence, the remaining incoherence may have too little importance to count. Furthermore, despite his strong attachment to Mr Cool, Conrad could cease to follow him if he were suspicious that Mr Cool's beliefs went awry. The suspicion of Mr Cool being wrong could be incompatible with the latter's nimbus. One could also question Conrad's reliability in counterfactual situations. But such situations cannot be situations in which Mr Cool is not sufficiently reliable as we claimed Mr Cool to be an excellent epistemic agent. They may be situations in which Conrad ceases to be attached to Mr Cool. But we can regard him as adhering to sound epistemic behaviour in such situations. Judging by the criterion of appropriate epistemic activity forwarded by Hookway, i.e. feeling sufficient confidence, and by Conrad's external success, his behaviour may be perfectly right. Thus Hookway is probably committed to grant Conrad a positive epistemic standing. Conrad's position does not seem worse than that of any normal epistemic agent. In contrast, I agree with Fairweather that Conrad's epistemic practice is not appropriate.

Laurence BonJour's example of a clairvoyant points in the same direction (BonJour 1985, 41). Sometimes the clairvoyant has a certain feeling which immediately prompts her to confidently and correctly believe that something will happen, similarly to us feeling confident to say that 2+2=4. There is nothing in her emotional structure that conflicts with what she feels and experiences in states of clairvoyance. One could claim that she lacks sufficient coherence in her immediate behavioural patterns. But why should an arbitrary system of immediate affective responses not be structured in a way that is sufficiently coherent with an affective attachment to clairvoyance? Her faculty may mislead her in many possible situations. But the same holds for normal perceptual faculties. Again Hookway must probably regard the epistemic practice of the clairvoyant as appropriate. In contrast, I claim, as BonJour does, that she is wrong to blindly rely on her states of clairvoyance.

My claim that Conrad and the clairvoyant are wrong can be disputed. One way to contest it is to embrace full-fledged externalism. Beliefs are justified if the mechanisms by which they are formed are reliable. This reliability does not need to be accessible. But Hookway accepts some demands of access internalism. He insists that cognitive integration should give rise to a feeling of confidence and that reasons (and perhaps background factors) enabling belief should be accessible to eventual attention (Hookway 2006, 58). Furthermore, the demands of full-fledged externalism and of epistemic rationality are at odds. In order to substantiate this claim I take the following for granted: One has the epistemic aim of believing truths and avoiding believing falsehoods in relevant fields. If one is justified in believing something in the respective fields, it is better to believe it than to withhold belief. Now assume that the way one normally acquires beliefs in a certain relevant field happens to be reliable, but one cannot forward the slightest reason for its reliability. Taking into account one's epistemic aims and externalist criteria, one should form beliefs in this way. But if one is faced with a choice whether to accept or to reject beliefs formed in this way, it is as rational to accept them as to refrain from accepting them. Since epistemic activity cannot but be guided by the requirements of rationality, it is completely unclear how it can cope with externalist demands. This problem relates to affective valuations too. If we cannot give a reason why to follow them, it is as rational not to follow them. Instead of endorsing externalism we could impose a conservativity-requirement according to which one may adhere to what one already accepts (Boghossian and Williamson 2003, 238). But the Cartesian situation of a communal epistemic practice going awry shows how arbitrary such a conservativity-requirement is.

Consider as a third example Carla the Cartesian. She entertains a set of immediate emotional evaluations which together with her natural excellences and her explicit epistemic policy allow her to be extremely reliable in forming first-order beliefs. Furthermore, she develops patterns of salience making her an outstanding and much admired epistemic agent. However, Carla entertains a second set of emotional evaluations embodying both Cartesian ideals of full reflectivity and at the same time grave doubts about the feasibility of these ideals inspired by de Sousa and Hookway. As a result, she continually suffers from an extremely distressing feeling of cognitive disintegration. Furthermore, inspired by the Cartesian ideal Carla reflectively scrutinizes her ways of belief acquisition whenever she has the opportunity to do so without neglecting everyday epistemic tasks. I feel inclined

to attribute first-order knowledge to Carla, but not to Conrad and the clairvoyant. In contrast Hookway apparently must prefer Conrad and the clairvoyant. Judging from Hookway's criterion of confidence or affective integration, Conrad and the clairvoyant have a better epistemic standing than Carla who does not feel confident. They can be granted knowledge. If Hookway regards cognitive integration as a prerequisite of knowledge, he cannot accept that Carla has knowledge. What advice could Hookway give to Carla? To suggest that she should embark on an enterprise of Cartesian doubt would be a strange concession to the Cartesian ideal. To suggest that she should simply dismiss her feelings of uneasiness would contradict their functioning as a criterion. Of course Hookway could claim that Carla's Cartesian valuations fail to track epistemic relevance. But this move may be at odds with his denial that it is appropriate to worry about emotive valuations failing to track epistemic relevance (see below).

I presume that whereas Carla's epistemic standing is probably uncomfortable but nevertheless even better than that of most people, Conrad and the clairvoyant show an epistemic behaviour that is not appropriate. Furthermore, again in contrast to Carla, their behaviour exemplifies a wrong pattern of immediate affective responses. In my opinion this is incompatible with Hookway's claim that a pattern of immediate affective responses which is embraced with sufficient confidence is right. It is also incompatible with emotional integration and external success being sufficient for appropriate epistemic behaviour.

To summarize our discussion of the three examples: While the first two paradigm cases cast doubt on Hookway's claim that affective integration is sufficient for justification, the third casts doubt on affective integration being necessary for epistemic success.

I now want to elucidate what is wrong with Conrad and the clairvoyant. Let us first look at some unsatisfying answers. Fairweather proposes that Conrad lacks the right kind of motivation, namely a proper motivation to seek the truth and avoid falsehoods (Fairweather 2001, 75). But Conrad may have a motivation for truth. This motivation could explain his beliefs, but is overridden by his motivation to conform to Mr Cool, which by over-determination explains his beliefs as well. I consider Fairweather's claim too strong anyway. One can have knowledge without being motivated to seek the truth. The same holds for alternatives like a deontological theory demanding that one should obey epistemic rules for their own sake. If Conrad in addition to his weak motivation to believe the truth has a reflective second-order knowledge of how reliable Mr Cool is, he may well have first-order knowledge. Furthermore, Fairweather's answer does not apply to the clairvoyant as she may be properly motivated to believe the truth and to avoid believing falsehoods.

There are further possibilities to explain what might be wrong with Conrad and the clairvoyant. Probably the most suggestive claim is that despite being reliable and cognitively integrated, they embrace the wrong set of affective evaluations. It may thus seem to be in the spirit of Hookway's approach to accept a weaker claim instead of (CH):

(CH*) It is appropriate or right to follow one's immediate valuations – provided they are sufficiently virtuous.

Such a virtue approach faces intricate problems. I want to focus on the following: We cannot simply take the fulfilment of the virtue proviso for granted, especially when taking into account that virtue has a success component. How can we trust our affective valuations if it is unclear whether we are virtuous? We may claim default justification or a positive presumption that our immediate valuations are virtuous. But what could be the rationale of such a presumption? We could also consider the epistemic agent's being confident or not being hesitant as an indicator of her being virtuous. But such a claim rests on already crediting her with being virtuous in her affective patterns. Hookway himself cannot accept the virtue proviso. This becomes obvious when he considers a gap between feelings of salience and external relevance. Such a gap apparently is a reason to demand that feelings of salience be properly grounded. Hookway rejects such a demand:

> [This] suggests that all we can be confident of is: If what we find epistemically salient tracks what is epistemically relevant, then we can be reliable in … obtaining knowledge … . Now I want to suggest that if justification rests on feelings towards beliefs and inferences that involve 'immediate' evaluations, the problem of closing this gap fades. (Hookway 2003, 90)

The proviso that salience must track relevance is a way to put what I called the virtue proviso. Hookway rejects this proviso. For him, we do not need to feel anxious whether this proviso is fulfilled if justification rests on immediate evaluations. Yet the issue is how justification can rest on immediate evaluations. Hookway can only recur to the claim (CH) considered above. Thus he cannot accept any condition limiting the claim that an affective valuation which is confidently endorsed yields justification.

5. Reflectivity and Epistemic Entitlement – the Cartesian Ideal Rises again

We still lack an answer to the question what is wrong with Conrad and the clairvoyant. The answer proposed in this section is the following: They lack an appropriate reflective knowledge regarding their ways of forming beliefs. Hookway's position that immediate emotional valuations yield justification is rejected. The Cartesian ideal of reflectivity as the only source of justification is revived.

Contrary to Hookway, I want to suggest that coping with the examples discussed requires a stronger reflective perspective on the epistemic character than it is given by emotional integration alone. Conrad and the clairvoyant both lack an appropriate critical awareness which could provide evidence that their ways of forming beliefs are reliable. Such a reflectiveness involves giving a sufficiently independent appraisal of reasons, motives and principles leading to one's beliefs. Conrad and the clairvoyant should proceed to critically appraising what they are immediately inclined to do. Furthermore, they should act in accordance with this assessment. Conrad should assess Mr Cool's qualities as a source of beliefs. If he still adheres to his behaviour after having gained independent reflective knowledge about Mr Cool's reliability, we might despise him. But – so I claim – we should not deny him knowledge. The same holds for the clairvoyant. She should identify her states of clairvoyance and compare them with the practice of her epistemic community and scientific evidence about being

clairvoyant. She should give an assessment of the grounds of accepting her opinions concerning the future. This assessment would allow her and her fellow citizens to properly evaluate what to make of her predictions, for example by comparing them to the products of established faculties and skills. When Conrad and the clairvoyant amass sufficient reflective knowledge confirming the reliability of their established processes of belief formation, or when they develop at least a sufficiently independent critical perspective on them, it may be epistemically appropriate to further adhere to them. If not, it is not appropriate.

One may wonder what exactly Conrad and the clairvoyant gain when they reflectively examine their epistemic practice. Imagine that both prove unable to amass sufficient reflective knowledge supporting their ways of belief formation and come to a negative assessment. In this case, they should give up the incriminated beliefs. Now we have every reason to think poorly of Conrad's epistemic skills compared to Mr Cool's. Unpredictable as the future is, a reliable faculty of clairvoyance may provide far better information about events to come than any normal faculty could. Let us presume that any assessment that leads Conrad and the clairvoyant to lose confidence in their established habits of belief formation worsens their epistemic position. Both lose their immediate unquestioned confidence, but also lose in terms of reliability or external epistemic success if they accept more sensible habits of belief formation than trusting Mr Cool or clairvoyance. If we understand rationality as depending on what is accessible to the epistemic agent, we can say that they gain in terms of rationality. It is not rational to rely on ways of belief formation without having reasons to do so even if they are in fact reliable. In contrast, it is rational to rely on ways of belief formation even if they are not reliable provided one has convincing reasons to do so.

This analysis of paradigm cases of deviant epistemic behaviour teaches us a general lesson. Hookway claims that it is appropriate to stick to our habitual patterns of epistemic behaviour. One may hesitate to embrace such a second-order claim without an independent assessment of this practice. So long as there is no such assessment, one could as well take a much less favourable attitude towards one's cognitive achievements. Consequently, one might question the claim that it is appropriate or right tout court to follow one's confidently endorsed emotive valuations. Could we not just as well be in the situation of those behaving deviantly? Hookway forbids such reasoning. As long as we feel sufficiently confident, we should not pay attention to it. But this feeling seems arbitrary. We may as well feel distressed and confused like Carla. Should the mere fact that we happen to have the former and not the latter feeling ground a wholly different epistemic standing? Why should absence of hesitation or discomfort be a criterion of good epistemic standing and not of stubbornness and intellectual laziness?

What distinguishes us from Conrad and the clairvoyant? They both follow patterns of belief formation which are eccentric. They do not conform to their community's established practice. Yet in my opinion, the epistemological problem involved in their eccentricity is that there is no established body of reflective knowledge that would allow to appraise the practice of Conrad and the clairvoyant – in contrast to sense perception, for example: We have a huge body of reflective knowledge confirming the reliability of sense perception. Anyone whose reliability is questioned could

point to this knowledge. This ability makes a practice of following sense perception right and appropriate. Of course, such knowledge presumably strongly depends on sense perception. Thus it can never provide a wholly non-circular justification. A sceptic may doubt that the assessment is sufficiently independent.

Now what about aspects of belief formation that cannot be evaluated in this way? Let us look again at the riddle of induction. Hookway stresses that we may never be able to provide an explicit rule explaining why 'emeralds are green' is a better generalization than 'emeralds are grue'. He emphasizes that we nevertheless are right that emeralds are green and not grue. Such generalizations rest on affective valuations. Perhaps we must simply rely on our feeling that 'emeralds are green' is sound, whereas 'emeralds are grue' is not, without being able to make this aspect of our activities more explicit. In contrast to sense perception, there is a gap in our reflective assessments. Hookway denies that this gap is a gap in being right or wrong. But it could threaten our entitlement to form inductive generalizations if this entitlement depended on a reflective assessment. We can point to many cases in which these generalizations have led to correct predictions. The use of inductive generalizations is embedded in a complex network of epistemic skills. Their functioning may involve both arbitrary and justifiable elements. However, even parts that are arbitrary when taken in isolation may have some positive epistemic status, simply because they are part of the use of an integrative faculty – of forming reliable predictions, say – which can be accounted for. Assume that we can give an appraisal according to which our capacities of forming inductive generalizations are sufficiently reliable. Thus we may be entitled to use these capacities, even though this use involves making blind emotion-guided decisions between, say, green and grue. But such a decision must be integrated in the use of a faculty we can account for in general. It is not right to follow an emotive valuation independently of such an embedding. Now other things being equal, it would be better to give a reason why 'green' is preferable to 'grue' without referring to a mere feeling. We can compare this situation to that of medieval philosophers and scientists regarding sense perception. They were entitled to use their perceptual faculties (as we are) because they could refer to a body of reflective knowledge that these faculties are reliable. But there was a blind spot in their assessments as they could not tell how perception functions. Or they had, like Descartes, quite erroneous opinions about the physical processes that lead from outer events to perceptual knowledge. The same may hold for the ability to provide inductive generalizations. The notable difference could be the following: We may never be able to account for the blind spots Hookway hints at. Just as in the case of sense perception, there may be another limit of reflection: Since induction is so deeply rooted in our epistemic practice, it could be impossible to provide a non-circular reflective evaluation of its achievements which does not already rest on induction.

Now when we keep in mind Sosa's proposal that reflective knowledge grows from a holistic network of animal knowledge, it is tempting to consider *all* immediate affective valuations as parts of such a holistic network. Such a network could resemble the network of scientific predictions in which inductive generalizations are embedded. Thus it could be argued that reflective knowledge serves to justify any immediate valuations playing a role in the holistic process of belief formation.

But while belief formation may be holistic, justification would surely require a more specific relationship between a set of immediate factors of belief formation and a certain body of reflective knowledge.

Perhaps a set of emotional valuations can itself be interpreted as something like a skill or faculty. But it is questionable whether there is a comprehensive body of reflective knowledge which allows to account for such a 'faculty' of emotions. Up to now, no explanation has shown that our immediate behavioural patterns with sufficient probability lead to epistemic success. Hookway considers the possibility of a naturalistic explanation (Hookway 2003, 90). But as Alvin Plantinga criticizes, such an explanation may lead to quite unsatisfying results concerning higher cognitive faculties (Plantinga 2000, 238). If such a naturalistic explanation confirming the reliability of affective valuations could be established, it would probably provide resources to justify our trust in these valuations. But as long as no naturalistic explanation has been established, it would be too rash to rely on its possibility.

The result of our discussion is that some amount of reflective knowledge or at least critical reflective awareness is a prerequisite of epistemic entitlements. Furthermore, reflective knowledge is epistemically valuable. What Conrad and the clairvoyant lack is something anyone relying solely on immediate valuations lacks. Accordingly one should reflect whenever the expected gains of reflecting are not outweighed by the opportunity costs of not using one's intellectual capacities otherwise. Yet de Sousa and Hookway argue that there are several natural limits of explicit reflection. Firstly, reflection will sometimes be too costly. We need to spend resources of time and intellectual strength on reflection that may be used to gain important first-order beliefs. Secondly, it is difficult to tell to what extent we can voluntarily control such an assessment and its implementation. We often cannot but adhere to our affective patterns of behaviour. This also holds for the decision when to reflect. Thus reflection arguably rests on affective patterns of salience resembling those it is to assess. Thirdly, there are tasks, such as the framing function of emotions, that simply cannot be further accounted for. Thus reflection probably has absolute limits. Some relevant factors of belief formation cannot be further elucidated. Fourthly, reflection may be an infinite task which never comes to a natural end. Reflectivity may be some sort of Kantian idea. Thus reflection may have relative limits, too, confining a realm to be infinitely extended.

Two final worries remain: The first concerns the alleged independence of reflective judgement. If there is no Archimedean vantage point from which emotions can be assessed, if reflective scrutiny must itself rest on emotional valuations, how can it be impartial? A headstrong sceptic will insist that any reflective endeavour is enmeshed in a vicious circle. But many will agree that even a critical appraisal which is not fully independent can have a certain justificatory value. The second worry concerns the regress problem. Like first-order knowledge claims, second-order claims need justification themselves. Claims to reflective knowledge seem to hang in the air until a third-order reflective ratification is given and so on. Yet firstly a reflective perspective might give a certain justification even if it is not yet justified itself. Secondly the regress could be stopped by a meta-reflective rating referring to reflective judgements as such. Such a meta-reflective rating would have to ground our capacities of giving the relevant reflective assessments.

6. Conclusion

We are now faced with a dilemma. On the one hand, we cannot but accept reflection to be shallow. There is no fully transparent, responsible and voluntary epistemic agency as Descartes and Cohen envisage it. On the other hand this does not mean that it is right or rational tout court to stick to affective patterns. Sometimes, if salience tracks relevance, our immediate affective valuations are virtues, sometimes they are vices. But since there often is nothing to be done about it, we are subject to epistemic luck. One might hesitate to call the outcome of such luck right or justified. Rather we may be called blameless victims of more or less favourable circumstances. However, due to the shortcomings of reflective control, it is quite doubtful whether there is anything not strongly depending on such epistemic luck. Contrary to Descartes, like Carla we must presumably accept our immediate affective patterns of behaviour. We cannot always give a concrete assessment. But in accordance with Descartes and Carla, we may be uneasy about such an acceptance. We may insist on a regulative ideal to account for all these patterns by making them explicit and scrutinize them critically. If no scarcities of time and resources were to be taken into account, reflection should have no limits except the absolute natural limits of the first kind mentioned above, for example those owing to the framing function of emotions. In spite of such scarcities and absolute limits, explicit rules and arguments form a growing network that partly replaces immediate behavioural patterns and valuations by enlightened ones, just as sketched by Sosa.

If it is to arise from reflection, justification must admit of degrees depending on how far reflection has already gone. Since there is no natural end to reflection, we probably must find some compromise between the ideal of reflectivity and other requirements of epistemic success. Since there may be no explicit rule to be stated where reflection should begin and end in view of these scarcities, it must be determined by immediate affective patterns of salience. But being blameless so far does not involve any stronger entitlement to limit reflection. It does not mean that it is right to trust one's affective patterns. Sound epistemic agency does not merely involve following paths of habit and emotion, but also trying to make transparent and explicit what one does, while being aware of one's shortcomings with respect to the Cartesian ideal of full transparency.

There are several other phenomena of epistemic immediacy which pose similar problems and which presumably are to be treated in the same way as the role of emotional valuations in an epistemic practice. Firstly there is sense perception. Some analogies have already been indicated. As Zagzebski emphasizes sense perception, too, often immediately influences our belief system without preceding reflection and perhaps without the possibility of a wholly non-circular reflective examination. But examples like clairvoyance which could also be interpreted as a sort of perceptual ability show that we cannot simply take the justificatory role of sense perception for granted. Secondly there are the notorious problems of rule following, especially regarding basic epistemic rules, for example rules of inference like *modus ponens*. We must be able to reason according to *modus ponens* before we can do it intentionally and before we can hold it up as a subject of critical scrutiny (Boghossian 2005; Pettit 2005). Furthermore, there is an analogy between inductive inference and the task of

extrapolating infinitely many applications of a rule from a finite number of paradigm cases. Such an extrapolation poses the Gerrymandering-problem that a series, say 1, 2, 3, 4 …, could be continued in infinitely many different ways. If Hookway's diagnosis holds, affective valuations may play a role in determining how to continue such a series. Thirdly: If we embrace some sort of inferential role semantics as, for example, Robert Brandom does, we must arguably take some inferential lines which are constitutive of the content of our concepts for granted. As an example of such a line take the transition from 'Paris is in the south of London' to 'Paris is not in the north of London' or from 'x is green all over' to 'x is not red all over'. We cannot critically assess all these inferential lines. But the discussion of concepts like 'boche' or 'tonk', which would license unacceptable inferences, indicates that we cannot simply take such constitutive inferential lines to be in the clear (Brandom 2000, 68–9, critically Boghossian 2003). Although Brandom apparently holds that one is by default justified in using established concepts, he at the same time acknowledges that a full reflective examination is an ideal. Such a full assessment would have to make explicit everything one commits oneself to by using these concepts (Brandom 2000, 69–70). I suggest that all these phenomena remind us of a lesson that the ancient sceptics taught us and which we are always in danger of forgetting: how precarious our epistemic standing is, how the light of reason emerges from a dark affective ground never to be fully enlightened.

References

Boghossian, Paul. 2003. 'Epistemic Analyticity. A Defense'. *Grazer Philosophische Studien* 66: 15–35.

Boghossian, Paul. 2005. 'Rules, Meaning and Intention'. *Philosophical Studies* 124: 185–97.

Boghossian, Paul; Williamson, Timothy. 2003. 'Blind Reasoning'. *Proceedings of the Aristotelian Society*, *Supplementary Volume* 77: 225–93.

BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge, Mass.: Cambridge University Press.

Brandom, Robert B. 2000. *Articulating Reasons. An Introduction into Inferentialism*. Cambridge, Mass.: Cambridge University Press.

Cohen, L. Jonathan. 1989. 'Belief and Acceptance'. *Mind* 98: 367–89.

Descartes, René. [1641] 1996. *Meditationes de Prima Philosophia*. In: Adam, Charles; Paul Tannery (eds). *Oeuvres de Descartes*. Paris: Vrin. Vol. 7.

Descartes, René. [1644] 1996. *Principia Philosophiae.* In: Adam, Charles; Paul Tannery (eds). *Oeuvres de Descartes*. Paris: Vrin. Vol. 8. 1–353.

Descartes, René. [1649] 1996. *Les Passions de l'Âme*. In: Adam, Charles; Tannery, Paul (eds). *Oeuvres de Descartes*. Paris: Vrin. Vol. 11. 293–497.

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.

Fairweather, Abrol. 2001. 'Epistemic Motivation'. In: Fairweather, Abrol; Zagzebski, Linda (eds). *Virtue Epistemology. Essays on Epistemic Virtue and Responsibility*. Oxford: Oxford University Press. 63–81.

Goodman, Nelson. 1954. *Fact, Fiction and Forecast*. London: Athlone Press.
Hookway, Christopher. 2003. 'Affective States and Epistemic Immediacy'. *Metaphilosophy* 34: 78–96.
Hookway, Christopher. 2006. 'Reasons for Beliefs, Reasoning, Virtues'. *Philosophical Studies* 130: 47–70.
Lange, Marc. 2002. 'Who Is Afraid of *Ceteris-Paribus* Laws? Or: How I Learned to Stop Worrying and Love Them'. *Erkenntnis* 57: 407–23.
Pettit, Philip. 2005. 'On Rule Following, Folk Psychology, and the Economy of Esteem. A Reply to Boghossian, Dreier and Smith'. *Philosophical Studies* 124: 233–59.
Plantinga, Alvin. 2000. *Warranted Christian Belief*. New York: Oxford University Press.
Quine, Willard V. 1960. *Word and Object*. London: Wiley.
Schouls, Peter. 1989. *Descartes and the Enlightenment*. Kingston: McGill-Queen's University Press.
Sosa, Ernest. 1997. 'How to Resolve the Pyrrhonic Problematic. Lessons from Descartes'. *Philosophical Studies* 85: 229–49.
Zagzebski, Linda. 2001. 'Must Knowers Be Agents?' In: Fairweather, Abrol; Zagzebski, Linda (eds). *Virtue Epistemology. Essays on Epistemic Virtue and Responsibility*. Oxford: Oxford University Press. 142–57.

Chapter Six

Critical Reflections on Affective Epistemology

Markus Wild*

In recent studies emotions have increasingly been regarded as inherently rational. Some philosophers have transported this tendency into the realm of epistemology. Affective states are said to play a decisive role in epistemic processes. Call this 'Affective Epistemology'. This chapter is mainly critical. It will downplay the supposed epistemological roles of emotions. A distinction between a Traditional Project and an Inclusive Project is introduced (introduction). Relying on a representational account of emotions (J. Prinz), it is shown that emotions do not play a decisive role in the Traditional Project (section 2). However, philosophers defending Affective Epistemology (C. Elgin and C. Hookway) are more concerned with the Inclusive Project. Here emotions are supposed to play a double role: on the one hand they highlight the salient aspects of a situation and on the other hand they regulate epistemic activities. These are the basic ideas of Affective Epistemology. This account faces several problems. Emotions, if not externally regulated, do not regulate inquiry, they add to epistemic self-questioning. Moreover, emotions are in many ways over-inclusive (section 3). One way of solving these problems is to bring emotions together with virtue responsibilism. As the requested version of virtue epistemology is based on the questionable notion of character (G. Harman and J. Doris), this 'marriage' is bound to fail (section 4). The final section endorses a very modest version of Affective Epistemology, and defends a distinction between knowledge and understanding.

1. Introduction: Two Epistemological Projects, Two Kinds of Virtue Epistemology

There are two large projects that are pursued in contemporary epistemology.[1] The first and more traditional project attempts an understanding of knowledge and warranted true belief and distinguishes them from other cognitive states. The nature of warrant and the structure of knowledge are the central concerns in this project; the need to answer scepticism looms large (Stroud 2000; Williams 2001). The second and more inclusive project describes and evaluates epistemic practices such as inquiry, belief

* Earlier versions of the paper were presented at the University of Zürich and the Humboldt University of Berlin. Many thanks to Simone Ungerer, Anik Waldow, Christopher Hookway, Martin Lenz, Dominik Perler, Alexander Staudacher, and Ulvi Doğuoğlu for discussion and criticism.

1 For a differentiated view on epistemological projects in philosophy see Weinberg, Nichols and Stich (2001).

formation or deliberation. Knowledge and warrant are not the central issues. While the Traditional Project focuses on cognitive *states*, the Inclusive Project is primarily concerned with epistemic *activities*. The latter is inclusive because it can incorporate epistemic agents. Virtuous traits of character (such as open-mindedness) come into view, not only beliefs but other mental states (such as emotions); not only warrant but also other epistemic qualities (such as regulation of inquiry) enter the picture. Even the primary epistemic goal of the Traditional Project is challenged: *truth*. In addition to truth there are other worthy epistemic goals, such as understanding, wisdom, solutions, power, justification, reflective equilibrium or permanent tenability (Elgin 1996, 123–7; Hookway 2003b; Kvanvig 2005). Generally speaking, epistemic activities aim at *cognitive success*.

The difference between the two projects parallels a difference in contemporary virtue epistemology. In moral philosophy the *moral* virtues can be seen as character traits, i.e. robust and fixed excellences involving both a motivational component and a reliable success component (Hursthouse 1999). Benevolence, for example, is a virtue that involves a person's motivation to promote the well-being of others and that is successful at doing so. The idea of character seems to play a fundamental role in almost all varieties of virtue ethics.[2] In virtue epistemology, this is not the case. Virtue epistemologists are interested in *intellectual* virtues, but they differ in their opinions about the very nature of these virtues (Axtell 1997; Hookway 2003b, 186–90; Baehr 2006). Virtue *responsibilists* hold that intellectual virtues have an analogous structure to moral virtues, they are excellences of epistemic character. Fair-mindedness, open-mindedness, intellectual courage and the like can be specified in terms of motivations that are somehow related to the general goal of epistemic success. For example, intellectual courage might be a virtue that involves a person's motivation to be persevering in her own inquiries and that is successful at doing so. Contrarily, virtue *reliabilists* conceive of intellectual virtues not so much as character traits, but as reliable or truth-conducive cognitive faculties or abilities. Such virtues include belief formation based on sight, hearing, memory or reasoning in approved ways. Virtue reliabilism is intended to engage in the Traditional Project, whereas responsibilism emphasizes the Inclusive Project by stressing the psychological, social, moral and political importance of our cognitive practices. Responsibilists are interested in the regulation of inquiry, and they ask for more than the mere assessment of beliefs as warranted. Both varieties of virtue epistemology, however, share the assumption that the epistemic properties of beliefs are to be defined in terms of epistemic properties of agents (Greco 2002). So they both agree on the direction a fruitful analysis is to take.

In this essay I am going to investigate the role emotions can possibly be said, and actually are said to play, in both epistemological projects. The general direction is mainly critical: I will relegate the supposed epistemological roles of emotions. Thus, in section 2, I am concerned with the endeavour of showing that emotions should not play a very important role in the Traditional Project. In a further move I will turn to

2 Virtue ethics is sometimes contrasted with utilitarianism and deontology because it is character based, rather than consequence or duty based. For a careful discussion of these contrasts see Watson (1990).

philosophers like Catherine Elgin or Christopher Hookway, and their investigation of emotions in epistemic inquiry. Their interest lies with the Inclusive Project, where emotions are supposed to play a double role: on the one hand they highlight the salient aspects of a situation, and on the other hand, they regulate epistemic activities. This is the basic idea of what I will call *Affective Epistemology*. As I will show in section 3, this account faces several problems and lacks a clear focus. One way of solving these problems is to bring emotions together with virtue responsibilism. This 'marriage', however, is bound to fail, for as I will argue in section 4 the requested version of virtue epistemology is based on a questionable notion of character. The final section will then present a modest form of Affective Epistemology which is able to stand scrutiny.

2. Emotions as Ways of Knowing

Sue is mad at Ernest because he broke little Doritt's new toy. Charles has told her so, but he was lying to her. Nothing happened, nobody broke the toy, everybody is just pulling her leg. Where did Sue's anger come from? From her belief that somebody (Ernest) did something she disapproves of (breaking the girl's toy). Sue is angry because (among other things) she knows, or at least thinks she knows, that *p*. Many emotions require belief or knowledge (e.g. being amazed, ashamed, disappointed, glad, proud, resentful, happy). In contrast, some emotions preclude belief or knowledge. Hoping or fearing that *p* preclude knowledge that *p*. You cannot know that Ernest Hemingway won the Nobel Prize and still fear or hope that Hemingway won, wins or will win the Prize. On the other hand, you cannot be glad or regretful that Hemingway won the Nobel Prize without knowing or believing that he did. This is not just a difference between forward-looking emotions (e.g. hope or fear) and backward-looking emotions (e.g. gladness or regret). Of course, Ernest can know that his plane will arrive late in Barcelona, if the plane has been scheduled for 7 p.m. but is still flying over France at 7 p.m. Under these circumstances, he cannot hope or fear that the plane will arrive late, but still be glad or regret that this will be the case. In a different situation, Ernest might hope or fear that his plane will arrive late, because he does not know whether the plane will be late or not; this however precludes his feelings of relief or regret (Gordon 1969).

These examples are meant to indicate that there are emotions *based on* knowledge or belief, and that there are emotions *precluding* knowledge or belief. This classification, though, is not exhaustive. Sometimes, emotions are *immune to* knowledge. Ernest is afraid of frogs although he of course knows that frogs are not dangerous creatures at all. Now, contrast emotions with the classical sources of knowledge found in most epistemology textbooks: perception, memory, testimony, introspection, inference or intuition. Ernest can see, remember, testify or infer that the plane was late or that his friend died in Barcelona. These are ways of knowing facts. He can know these facts in different ways. Ernest can resent, be horrified or happy about the plane's delay or about his friend's death in Barcelona (instead of, say, in the hills). There are, as it were, different ways of *emoting* facts. But none of these ways of emoting that *p* is a way of knowing that *p*. It goes without saying that

Ernest can *introspect* and find out about his mental state. By feeling sad he learns that he is sad; through further reflection about his sadness he comes to know that he deeply regrets the death of his friend. However, he comes to know that he regrets this tragic event by way of introspecting his mental states. His sadness and his regret becoming or being *conscious* is the important step in this process, not his being sad or regretful. A mental state being or becoming conscious is just a very good precondition to know one's own mind, while it is hard to know one's mental states when they are not conscious.

While emotions can be based on knowledge, preclude knowledge or be immune to it, they fail to be ways of knowing. The fact that emotions sometimes preclude knowledge or are immune to it has traditionally been a reason for treating them with caution and suspicion. Ever since Thucydides and Tacitus, anger and zeal (being immune to knowledge), fear and hope (precluding knowledge) are considered inimical to epistemic success. The point is not that emotions are, as the saying goes, traditionally irrational or antithetical to reason (this is what Elgin in this volume calls 'the standard view'). Emotions *are* related to reason by sometimes presupposing knowledge, sometimes precluding or even contradicting it. Rather the point is that emotions are not ways of knowing or states of knowing, which precludes them from the realm of knowledge.

Another reason why emotions fail to be part of the realm of knowledge is that emotions are gradual, they vary in intensity. Sue's anger permits of more or less; Ernest's sadness can be more or less intense; she can be *very* disappointed; he can be *slightly* surprised. Contrarily, knowledge of facts does not come in degrees; one cannot know more or less, very much or just a little bit. Knowledge is not gradual, it is absolute. As Fred Dretske puts it:

> When talking about people, places and topics (*things* rather than facts), it makes sense to say that one person knows something *better than* another … . But factual knowledge, the knowledge that something is so, does not admit of such comparisons … . In this respect knowledge is *absolute*. It is like being pregnant: an all or nothing affair. (Dretske 1981, 363)

You can *see* Ernest's moustache better than I do (suffering from bad eyesight), or you may *recognize* Ernest sooner than I do (having not seen him for a long time). But you and me, we both know the fact that Ernest has a moustache (by seeing him having a moustache). We know the fact that this man is Ernest (by recognizing him). You do not know any of these facts better than I do. Just because of your excellent eyesight, you do not know better than I do that Ernest has a moustache. You just see it or him better. As you know Ernest in many ways (by the way he looks, walks, talks, dresses etc.), you surely know him better than I do. But this does not alter the fact that as soon as we both recognize him, we both know that this man is Ernest. A general way of understanding the absoluteness of knowledge consists in appreciating knowledge's intimate relation to truth. If we know a fact, we have a true belief. Beliefs are not more or less true. The truth predicate does not permit a

more or less, nor does knowledge.[3] Emotions are not absolute in this sense. If Ernest *knows* that he is late, he does not know it more or less. If he *regrets* that he is late, he can regret it more or less. Emotions have a relation to truth only inasmuch as they are based on knowledge. There is no intimate relation between emotions and the concept and truth.[4]

Apparently. One might, however, object that without emotions one would not gain knowledge of certain properties of the environment. What kind of properties? In order to offer a more detailed explanation of this objection I will follow the representational theory of emotions Jesse Prinz has developed recently (Prinz 2004). In accord with the representational theory of the mind, Prinz favours the idea that emotions are mental states and that mental states are representational states. Mental representations have informational functions, which means that a mental state represents something that reliably causes it only if the state has the function of being normally caused by that thing or has the function of indicating another thing by being caused by that thing (Dretske 1995). Prinz holds that emotions are *embodied appraisals*. This means, firstly, that emotions are perceptions of patterned changes in the body and therefore are caused by bodily states. These bodily states track certain kinds of conditions in the environment of an organism. The function of emotions thus is not to represent bodily states but environmental conditions. The psychologist Richard Lazarus has offered a highly intuitive taxonomy of so-called 'core relational themes of emotions' (Lazarus 1991, 121). In virtue of their having core relational themes, emotions can qualify as appraisals. For example, Lazarus identifies 'having experienced an irrevocable loss' for sadness and 'facing an immediate, concrete, and overwhelming physical danger' for fear. Fear is caused by patterned changes in the body, but it represents danger in the environment. Based on these ideas, Prinz conceives of emotions as representations of core relational themes, and emotions do so by perceiving bodily changes. Core relational themes are 'the real contents of emotions, and bodily changes are their nominal contents' (Prinz 2004, 68). The perception of a snake, for example, causes changes in the bodily state (the heart races, the hair stands on end and so on). Fear is the perception of these changes (nominal content) and thereby represents danger (real content): it has the function of representing danger, it is set up to be set off by dangerous things. If fear is a representation, it can misrepresent. Such a misrepresentation would for instance occur with someone's fear of perfectly innocent animals such as (European) frogs. The important point is that in normally functioning organisms fear reliably represents danger, or sadness the loss of something valued. Some of these emotional responses might have been acquired in the process of evolution, others by learning and education (Prinz 2004, 66–76, ch. 5–6). Therefore, some emotional

3 According to Dretske knowledge is absolute, because the evidence one must have in order to know must be *conclusive*, or because knowledge inherits this quality form the *information* it depends on. The absolutist concept of knowledge has been challenged by Hetherington (2001), not convincingly as it seems to me.

4 For another point of view see Salmela (2006).

representations have phylogenetic functions, others have ontogenetic functions (Dretske 1995, 8–22).[5]

The upshot of this is that by emotions we gain knowledge of certain properties in the environment. It therefore seems, contrary to what has been defended so far, that emotions can be ways of knowing. According to the embodied appraisal theory, emotional states reliably represent core relational themes as their real content. Compare: The intellectual virtues of the virtue reliabilist are reliable cognitive faculties. Just as belief formation based on reliable sight or reliable memory produces knowledge about things and properties (which these faculties are set up to be set off by), so can belief formation based on emotions (i.e. based on the perception of bodily changes) reliably represent certain features in the surroundings of an organism and produce knowledge about these properties. But what kinds of properties are represented? These properties seem to be *response-dependent*, because their representation and the knowledge we gain about them depends on our emotional responses (on our bodily responses). It seems that these properties resemble colour properties or other secondary qualities. Some thing's looking red is a good reason for believing that there is something red. Likewise, some thing's eliciting fear is a good reason for believing that there is something dangerous. By our responses we classify certain things as dangerous, as disgusting, as offending and so on. Without our emotions, these properties would not exist, or so the idea goes (Elgin 1996, 161–7; see Elgin, this vol.).

Yet, we have to be careful. Being dangerous, of course, is a *relational* property. Something can be dangerous only to some creature or other, depending on the creature in question. The cat is not dangerous to Ernest, but it is dangerous to the mouse. But being dangerous to a creature does not depend on being *represented* by that creature as being dangerous. A hungry snake is dangerous to a mouse, because the presence of a hungry snake means immediate, concrete and overwhelming physical danger for the mouse. This holds whether or not the mouse is frightened by the snake, or represents it in any other way. So, being dangerous is not a response-dependent property.[6]

Moreover, the knowledge we hopefully gain by being afraid or experiencing fear (knowledge about something being dangerous) could be replaced. Consider Frank Jackson's Knowledge Argument: Mary has spent all her life in a black-and-white room and has never seen any colours. *Ex hypothesi* she knows all the physical facts

5 This is not the place to generalize the above points taken from Prinz. I will, therefore, side-step the question whether all emotions can be characterized in representational terms, and whether all emotions can be characterized in representational terms referring to core relational themes. The reason for my reluctance lies in Paul Griffiths's well argued case that the expression 'emotion' covers too many different phenomena (Griffiths 1997).

6 As Stephen Mulhall puts it: 'So, for example, any adequate account of the fearfulness of certain objects must invoke certain subjective states, certain facts about human beings and their responses. However, it must also invoke the object of fear … . Whether something is really fearful is in an important sense an objective question – the fact that we can find some things fearful when they do not merit that response (e.g. house spiders) shows this; and insofar as our capacity to fear things permits us to discriminate the genuinely fearful from the non-fearful, then that affective response reveals something about the world.' (Mulhall 1996, 198)

about colour. One day Mary leaves her black-and-white room and sees a red tomato. In this very moment she learns something she has not yet known while being captured in the room: what it is like to see red. Yet, what she knows now is not a physical fact: *ex hypothesi* she has already acquired knowledge of physical facts about colours. Now imagine Ernest living in a room inhibiting all fear. He never experiences fear. However, he learns about all the things, creatures and situations being dangerous for human beings. One day, Ernest leaves his fear inhibiting room, he sees a huge snake and is horrified to death. By now he knows something which he has not known while spending time in the room: what it is like to experience fear. The conclusion is that the features reliably represented by emotions can be replaced by other kinds of knowledge, save the subjective knowledge that fear (or any other emotion) feels like *that*. Therefore it seems a little hasty to hold that without our emotions these properties would not exist.

The parallel with purely response-dependent properties is misleading in another way. A properly functioning visual system of an organism is able to detect colours and other visual properties in the surroundings as soon as it gets in contact with surfaces. Given normal conditions and good eyesight, Ernest will see the colours of the delicious apples in front of him. Now imagine that unbeknownst to Ernest the apples are poisoned, which entails immediate, concrete and overwhelming physical danger. Of course, Ernest does not react by experiencing fear, because the apples look red, round and so on, but not dangerous. While tasting the fruit he will not experience fear. It is only by believing or knowing that the apples are poisoned, or by perceiving something dangerous that bodily reactions, leading to the emotion of fear, would occur. This example shows that perhaps it is true that certain patterned bodily changes reliably indicate danger; yet apparently, they are not reliable indicators of the presence of dangerous things. And this marks a crucial difference to the reliability of properly functioning visual systems which do reliably indicate colour properties: they are reliable indicators of the presence of so-and-so coloured surfaces. Hence, even though in some respects emotions might be seen as ways of knowing, they are so only in a very restricted sense.

Emotions are based on knowledge, preclude knowledge or are immune to it. However, they are ways of knowing in the following sense: Given a certain account of what emotions are, they can be viewed as being reliable cognitive faculties indicating certain features of a creature's surroundings and thereby giving rise to knowledge about these features. These features should not be understood as being response-dependent properties in the sense colour properties are sometimes said to be. Emotions can be integrated in the Traditional Project as ways of knowing, they fit in the general frame of virtue reliabilism. But as such they do not by themselves contribute to the project of understanding the nature and structure of knowledge.

3. Principles of, and Problems for, Affective Epistemology

When it comes to the Inclusive Project, emotions apparently have a more illustrious standing, because they have two characteristics which seem to render them useful for epistemic activities. The first characteristic has been pointed out by Ronald de

Sousa (1987, 195–6). Emotions highlight patterns of salience, thereby fixing and regulating our attention. 'Emotions are sources of salience' (Elgin 1996, 149). The second characteristic stems from a comparison of emotions with other mental states. Unlike the attitudes of higher cognitive and motivational states, emotional attitudes have charged valences. Where amusement and joy are intrinsically positive and sadness and fear intrinsically negative, belief or intention are neutral (Gunther 2004). Emotions seem to possess intrinsic evaluative force. These characteristics have a promising appeal for epistemologists interested in the Inclusive Project: emotions can be viewed as *guiding our attention to salient features of epistemic situations*, and because emotions are intrinsically evaluative they could even pick out *the right kinds of salient features and thereby regulate inquiries*. This is the basic idea of Affective Epistemology to be found in the writings of Elgin and Hookway. In this section I want to raise a number of problems relating to Affective Epistemology. These problems mainly concern Hookway's demanding version of Affective Epistemology. (I take up the discussion of Elgin's modest version of Affective Epistemology in the final section.)

Consider the following cases: The salesman has argued convincingly for his product, nonetheless Sue has bad feelings about this affair, and rightly so. Despite his enthusiasm Ernest takes the uneasy feeling of his guide as a good reason for discontinuing the lion hunt.[7] Here is a more detailed example. In the following passage from *The High Window* private detective Philip Marlowe, while searching the house of a suspect, stumbles across a photograph. The photograph depicts the late Horace Bright leaning out of a window and yelling.

> There I was holding the photograph and looking at it. And so far as I could see it didn't mean a thing. I knew it had to. I just didn't know why. But I kept on looking at it. And in a little while something was wrong. It was a very small thing, but it was vital. The position of the man's hands, lined against the corner of the wall where it was cut out to make the window frame. The hands were not holding anything, they were not touching anything. It was the inside of his wrists that lined against the angle of the bricks. The hands were in air. The man was not leaning. He was falling. (Chandler 1988, ch. 29)

Marlowe feels that the photograph has a meaning and he feels that something is wrong. These affective states motivate and instruct further examination: Do not look at the obvious, mind the details! As a result Marlowe gains the new belief that the man is falling. With this belief the uneasy feeling vanishes and the examination comes to an end. The general idea is this: affective states *motivate*, *direct* and *terminate* everyday, professional or scientific inquiry. It seems obvious that affective states can be part of a *motivation*. *Direction* is a variant of de Sousa's idea that emotions are determinate patterns of salience among objects of attention, lines of inquiry and inferential strategies.[8] Importantly, Marlowe's bad feelings are not themselves

7 Because Elgin and Hookway tend to extend the scope of what is customarily called an emotion and include feelings, I will sometimes use the expression 'affective states'.

8 A lot of things produce patterns of salience. Take drinks or drugs. Take being left after 15 years of marriage and walking through the empty house. Take becoming blind or diseased. Take going through the city with a disabled person or with little Doritt.

reasons or justifications for his new belief, they just direct his search. However, in the presence of that belief the initial affective state vanishes and Marlowe knows for sure that the man is falling. In this way affective states contribute to the fixation of belief, thereby *terminating* the inquiry: they 'ensure that our reflections stop at the right place' (Hookway 2002, 257). Let us call this the Regulation Principle: *Affective states regulate (motivate, direct and terminate) epistemic activities* (Hookway 2000, 156–9; Hookway 2002, 253–9; Hookway 2003a, 84–5). Which are the relevant affective states and how do they regulate (i.e. motivate, direct and terminate) inquiry? I think that affective states, neither in their role as motivators, nor as directors or terminators, contribute to a large extent to our epistemic activities. They are just minor characters in the drama. To show this I will pursue the ingredients of the Regulation Principle – namely motivation, direction and termination of inquiry. Let us start with motivation and go back to Marlowe's case for illustration.

Marlowe gains a new insight (the man is falling) and abandons an old belief (the man is leaning out the window). Looking at the photograph and having a bad feeling, he first enters a state of doubt. By way of feeling doubt an epistemic actor may withdraw epistemic credit from some of his beliefs. This actual feeling of doubt can be characterized as 'epistemic anxiety' (Hookway 1998; Hookway 2000, 154–6). Epistemic anxiety is a motive for distrusting one's beliefs, and for refraining from them without the support of further inquiry or examination. It is, however, *one* motive among others and it would not motivate further inquiry and examination unless there were additional motives. (Doubt, as Pyrrhonists are well aware, is a good reason to stop further investigation in the first place.) Marlowe would not examine the photograph any further without the general aim of solving a crime and his somehow old-fashioned sense of professional loyalty. Moreover, there is the question *which* beliefs are to be discredited or distrusted. Doubt should not afflict all of his beliefs, but only those relevant to a specific *field* of his investigation. Hookway seems to overlook this point, for he emphasizes at various places that emotions or affective evaluations *spread* through our whole cognitive system, without delimitation of specific fields. The advantage of his position seems to be that states like doubt can be important to all of our epistemic activities. But consider the consequences! Sometimes emotions just 'skew the epistemic landscape' (to borrow an expression from Peter Goldie; see p. 159). Anger, boldness, cowardice, despair, envy, fear or gratitude may distort our epistemic activities. It is obvious that emotions (or hypotheses or beliefs, for that matter) can do that. If you are in the right emotional disposition, and if there are no undue influences, there will be no problems. Imagine Sue after one year in her new job. Sue is desperate as she does not like it. Why is it so? What is going wrong? Is it Sue or is it the job? What shall she do? Well, some rational self-control will help! Just check the quality of the emotional disposition you are in. What worries me is the question whether Sue, being in bad emotional shape, actually can exercise rational self-control. The exercise of self-control will itself be affected by her bad emotional disposition. Even worse, emotions can go unnoticed, change into another emotion, resist identification, or continue to resonate long after they have gone. These are no idle objections, such as: imagine an evil scientist stimulating your frontal lobe to the effect that you find yourself in a constant, distorting bad emotional state. If Sue is indeed in the wrong emotional disposition, she is possibly unable to

exercise rational self-control, because the very same emotional disposition skews her epistemic landscape. If the emotional disposition is indeed essential to Sue's practice of epistemic evaluation, and if some distorting emotion *spreads* all over her belief system, and if the very same disposition keeps her *motivated* to move towards epistemic morasses, she may be unable to exercise rational self-control in a principled way. If affective states actually have this ability to spread doubts all over a belief system, better look for help! It therefore seems much more reasonable to engage in the attempt of demarcating a field of inquiry where emotions are allowed to play a minor part.

This brings me to my last point concerning the motivation part of the Regulation Principle. Suppose Marlowe is drunk, ill, under the influence of drugs, jealous or deeply depressed. In any of these circumstances he might have reason to refrain from his feeling of doubt and preserve his old beliefs instead, beliefs which were generated under more favourable conditions. Doubts are to afflict beliefs relevant to a specific epistemic field only if the epistemic actor takes himself to be in normal conditions. What these considerations show is that whether or not an affective state like doubt is able to motivate inquiry depends on three general factors: The presence of additional motives to start or continue inquiry, the delimitation of an epistemic field and the identification of normal conditions of the epistemic actor.

I now turn my attention to the direction part of the Regulation Principle. Emotions are said to be determinate patterns of salience among available information, inferential strategies and so on. One interpretation of this claim runs as follows: They render salient a relevant set of information or strategies among available alternatives. Emotions prevent us from getting lost in endless explorations of a potentially infinite epistemic landscape by providing us with a kind of search strategy for each kind of problem we must solve by calling up the relevant information. In this function they avoid one variant of the frame problem: Given an enormous amount of information, and some task to be fulfilled using some pieces of information, which piece of information is relevant for the task? The frame problem is a research problem. Emotions actively illuminate epistemic space, choose between alternatives and passively reflect what is really and rightly important. Viewing this complex task one may now wonder how emotions are able to accomplish this. I will pick out two of all the relevant points here.

The first point, the decision problem, has many solutions. Emotional salience is just one possibility among many others. The following procedures also narrow epistemic search space, help to choose between alternatives and have time-saving effects: tossing a coin; asking friends, colleagues or experts; building up reliable responsive reactions by training; fixing primary interests and goals, taking means just good enough relative to these goals and interests ('satisficing'); using fast and frugal heuristics such as 'never change a winning team', 'better safe than sorry' or 'take the first'. And of course, there is room for reasoned judgement, reflection, giving and asking oneself or others for reasons. The frame problem is sometimes construed as if it were reasonable to go through all thinkable or possible alternatives, to consider all the available information and strategies. To be sure, this would be very irrational. The delimitation of the epistemic field alone constrains search space. More importantly, emotions are not by themselves decisive. In fact, they are part of the *material* to take

into consideration when confronted with the problem of limiting epistemic space or choosing between alternatives, but they are not necessarily parts of the decision *procedure*. Some set of information and strategies might feel right, while there still might be reasons available that prevent us from opting for them. Sometimes emotions and other affective states just interfere with other decision making strategies or with reasoned judgement. They take over control. In such cases we decide and act without further reflection. Emotions take over because they momentarily deactivate other dispositions or judgements.[9] The second point, whether or not emotions are calling up the *right or relevant* kind of information and strategies, is more important. The general idea amounts to this: Emotions determine patterns of salience, and *if* salience tracks epistemic relevance, epistemic success is secured – at least this is the major claim of Affective Epistemology. What could possibly be a non-circular justification of the conditional claim that salience tracks relevance? Does salience track relevance in the first place? Is this process reliable? I will address these questions by discussing the termination part of the Regulation Principle.

Remember: Marlowe accepts a new belief. One reason for accepting it is the way the man on the photograph, especially his hands, looks like. Another reason is the vanishing of doubt and something we might describe as a 'feeling of knowing' (Mangan 2001). This feeling of knowing Hookway calls *immediacy* (Hookway 2003a). Epistemic immediacy endows inferences, beliefs, doubts or questions with the feeling of being 'compelling' or 'good', they stand out as relevant, they are salient. Such immediacy is 'unreflective' (it strikes you as just the right thing) but evaluative (Hookway 2003a). At this point the two characteristics of emotions mentioned at the beginning of this section meet again: emotions seem to have intrinsic evaluative force and they pass this force on to the features they mark out as salient. And it is because of this that emotions and other affective states can be perceived as *articulating* evaluative, epistemic standards in a non-reflective way; they *express* the standards of an epistemic actor (Hookway 2000; Hookway 2003a). The basic idea, the *Articulation Principle*, runs as follows: *Emotions articulate epistemic standards, because they highlight the relevant epistemic features, because they are intrinsically evaluative and because they have some phenomenal quality (such as immediacy).* As emotions articulate epistemic standards, they regulate inquiry. The Articulation Principle forms the basis of the Regulation Principle. What is this supposed to mean?

In the Traditional Project scepticism and the problem of justificatory regress loom large. The regress problem is part of the Agrippan Trilemma[10]: either the justificatory chain goes on for ever (regress), or it stops with a belief which stands

9 As Jon Elster puts it rather sharply: 'The emotion serves as a functional equivalent for the rational faculties it suspends, by inducing the very behaviour that is rationally required and that reason, if left undisturbed, could have come up with by itself. The emotions do solve problems – but problems that are to some extent of their own making. The capacity for the emotions to supplement and enhance rationality would not exist if they did not also undermine it.' (Elster 1999, 291.)

10 Named after the ancient philosopher Agrippa who seems to have been the first to make devastating use of it (see Williams 2001).

with no further justification (dogmatism), or one goes round in circles and returns to the very belief that one intended to justify (*petitio*). As the Agrippan Trilemma offers three options, there are three answers: infinitism, foundationalism and coherentism. One could be tempted to think that the feeling of immediacy might be an answer to the sceptical problem in the sense of the second part of the trilemma: there is immediate justification without any need of further grounding; this immediate justification consists in the feeling of immediacy, which is intrinsically evaluative and articulates epistemic standards (Hookway 1993). This perspective might provide us with a fairly good psychological explanation for why we actually stop somewhere: we cannot go on for ever, and something feeling just right or compelling or relevant is a good place to stop. We can trust our intuitive feelings of simplicity or immediacy. Obviously this will not answer scepticism (McGinn 2003).[11] The sceptic will just retort: Well, of course, *if* the feeling of immediacy is reliable, and *if* emotions actually do articulate epistemic standards, and *if* these standards can lead to epistemic success! But how do you know? The Affective Epistemologist could answer: On behalf of my feeling of immediacy! Which is, of course, question-begging. Or he could answer: By our epistemic success! Which is no answer at all, because it is the very possibility of epistemic success that is challenged. As far as I am concerned, the only possible answer could be: By our epistemic standards! This answer, however, cannot rely on affective states articulating these standards.

The conditioned answer to the sceptic is just a variant of the answer we found in discussing the direction part of the Regulation Principle: If salience indeed tracks relevance, epistemic success is rewarded. How do emotions and other affective states evaluate and regulate our epistemic activities? Why do they track what is relevant? According to the Articulation Principle, the answer is that emotions articulate epistemic standards. Of course, it just might be the case that emotions are intrinsically linked to a mentally represented set of standards (norms, goals and values) regulating our epistemic activities. This set of standards, though, needs to be exposed to a process of epistemic evaluation. And it is plain to see that those emotions linked to the set in question cannot form the basis of the required evaluation, because they depend upon the norms, goals and values we seek to evaluate.

I have argued in this section that emotions articulate epistemic standards only in a very limited sense, by pointing out the following problems for Affective Epistemology:

- Emotions skew the epistemic landscape or take over control against considered judgement. If not externally regulated, they add to epistemic self-questioning and loss of self-control.
- Emotions depend upon the presence of additional motives to start or continue inquiry, the delimitation of an epistemic field and the identification of normal conditions of the epistemic actor.
- Aspects of a situation might become salient by other means than emotions.

11 One might think that the burden of proof lies with the one who questions intuitive simplicity or immediacy. This misses the point. Answering scepticism by pointing to epistemic feelings just invites the question of how such feelings possibly can stop justificatory regress.

Emotions by themselves do not guarantee the salience of epistemic relevant features.

- Emotions cannot justify the conditional implicit in the Articulation Principle.

One way to solve all of these problems is the 'marriage' of emotions with virtue epistemology. Before discussing this solution in section 4, I briefly want to turn to a quite general problem for Affective Epistemology.

In Affective Epistemology we find emotions, feelings and affective states. Proponents of Affective Epistemology like Elgin or Hookway are not particularly interested in the concept of emotion itself. It seems sufficient that emotions are neither understood as pure states of feeling nor as pure cognitive states (Elgin 1996, 147–9). It nevertheless seems to me that emotions are somewhat *over-inclusive* in Affective Epistemology. Let me explain.

According to the Regulation Principle emotions motivate, direct and terminate epistemic activities, and according to the Articulation Principle they are patterns of salience, solving a version of the frame problem. They are intrinsically evaluative, and they posses some relevant phenomenal quality: they regulate inquiry, evaluate activities and articulate standards. Affective states do a lot of epistemological work. This is comparable to a movie with several main characters played by one single actor. Remember Peter Sellers playing four roles in Stanley Kubrick's fabulous *Dr. Strangelove*. True, he plays most of the main characters in Kubrick's story, but it would not be telling to call all of the main characters 'Sellers'. The first potential problem with over-inclusiveness is just this: There are too many fundamental roles for the affective states to play.[12]

Talk of emotions, feelings and affective states includes a great deal of things, amongst them being the feeling of immediacy. Inferences, for example, are said to posses the phenomenal feeling of immediacy in the sense that they are compelling (Hookway 2003a, 81–3). But do inferences feel like anything? Is it like anything to infer? Is the phenomenal feel of an inference a property of the inference in the same way the red-round-sprinkled-look of a strawberry is a property of the *phenomenal content* of a conscious perception? It is strange enough that both philosophers and scientists often think it is obscure what work consciousness does in our cognitive lives. To me it seems rather clear that conscious experience of the world and of one's mind makes it possible for us to think about properties and objects, contents and processes. In the greater part of the philosophical literature on consciousness, four classes of mental states are credited with phenomenal content: sense perceptions (like seeing), proprioceptions (like pain), emotions (like surprise) and moods (like being depressed). The propositional attitudes are not included. In addition to these four classes some psychologists think there is a kind of *non-sensory* consciousness

12 The problem of over-inclusiveness is not that emotions are wholly unable to play any of these roles. Rather, the point is that proponents of Affective Epistemology should differentiate between different mental states (such as affections, emotions, consciousness, intuitions, moods and so on) in order to evaluate more clearly the distinctive roles different mental states play. Affective Epistemology, as it stands, seems to lack a clear focus.

constituting a fifth class, comparable to what William James used to call the 'fringe of consciousness'.

According to James the most important feelings among these fringes are the mere feelings of harmony and discord, of a right and a wrong direction in thought (James 1890). Psychologists talk of a feeling of knowing, of familiarity or of rightness. People suffering from Capgras (a disorder that makes a person believe that a close acquaintance has been replaced by an identical looking swindler or machine or what have you) or from OCD (obsessive compulsive disorder) exhibit the lack of such fringe feelings. OCD patients, for example, have to check over and over whether they have locked the door. OCD may result from a sort of disconnection of the feeling of rightness in relation to recent memory of having taken the necessary precautions. The cognitive function of fringe consciousness is quite obvious. Most of our everyday thinking, feeling and acting operate outside conscious awareness. Fringe consciousness represents this non-conscious processes and provides our conscious reasoning with necessary context information. This fringe is something like the bridge between non-conscious and conscious cognition. It prevents conscious processing from cognitive overload and clears the space of conscious processing for its proper tasks like attention to new information or salient features. More specifically, the feeling of rightness functions in consciousness to signal how well the conscious and non-conscious domains of human cognition are integrated. Rightness, therefore, is the core feeling of positive evaluation and of coherence (Mangan 2001). It is obvious, I think, that this point of view perfectly well fits many of the things Affective Epistemology tell us. Besides fringe consciousness, one may be inclined to point to *intuitions* as a source of knowledge having a special phenomenal quality, or to something like Heideggerian *moods* (see Mulhall 1996). Thus, the second potential problem with over-inclusiveness is this: It seems that there are further mental states worthwhile of careful consideration (like fringe consciousness, intuitions or moods), covered under the general talk of emotions and affective states.

There is yet another sense of over-inclusiveness. How useful is the category of emotion at all? Paul Griffiths argued that the expression 'emotion' covers too many different phenomena (Griffiths 1997). Some emotions are more like sneezes, others are like 60 years of marriage, some spread across cultures and species, others are highly determined by cultural factors, some are phenomenally salient, some are very diffuse or have no phenomenal feel at all. Nevertheless we, the gullible folk, see a distinct and unitary kind where in fact there is a heterogeneous category. Griffiths claims that emotion research shows that three very different sorts of states fall under the extension of the folk concept: affect programs, higher cognitive states and social constructions. He suggests that these are radically different kinds for which no unified theory is likely, and that this is a good reason for scepticism about the theoretical utility of our concept of emotion. The third potential problem with over-inclusiveness is: Even if we manage to limit emotion's cognitive tasks, and even if we differentiate emotions from other relevant mental phenomena (like fringe consciousness or intuitions), there is still a question whether emotion is a uniform and useful theoretical concept.

As it stands, Affective Epistemology seems to lack a clear focus. These potential problems notwithstanding, I will now turn to the marriage of emotions with virtue epistemology.

4. A Critique of Pure Character

Virtue responsibilists recognize that their approach to ethics is concerned with patterns of motivation, emotion and reasoning that lead us to call someone a person of a certain character. Virtues, be they moral or intellectual, are traits of character (Hookway 2000, 152). Just as for moral virtues, intellectual virtues can be seen as robust and fixed excellences of epistemic character. Taking seriously the role character traits possess can help to understand how the involvement of affective states can regulate inquiry, if affective states are seen as articulations of the epistemic standards embodied in our virtuous character (Hookway 2003a). An inquirer does not have to provide a non-circular vindication for the normative standards which guide him, because this guidance is just part of his virtuous character. Although the virtuous person in question may not be able to provide a justification of the epistemic success he achieves, the merits can nevertheless be attributed to the virtuous person. In this sense the person is responsible for his epistemic success (Hookway 2003b). The general idea is this: epistemic success depends upon emotional salience tracking relevance, which in turn depends upon virtuous epistemic character (Hookway 2003b, 200). Let us call this the *Epistemic Character Thesis*. It implies that an epistemic actor having the right kind of epistemic virtues will reliably be motivated and directed by the right kind of affective states. Emotions will not skew the epistemic landscape or take over control; epistemically relevant aspects of a situation will become emotionally salient; the virtuous character can justify the important conditional implicit in the Articulation Principle. I am not going to elaborate upon these implications and then put them into doubt. I will scrutinize the central notion in question in the Epistemic Character Thesis: character.

What is the notion of character involved in this discussion? A character can be decomposed in certain character traits. Character traits are broad based dispositions that help to explain what they are dispositions to do. They are to be distinguished from other broad dispositions such as possession of certain knowledge or skills or innate abilities because they involve more than simply having relevant skills or knowledge: a person with the relevant character trait has a robust disposition to use relevant skills and pieces of knowledge in the relevant way. Traits of character are supposed to be robust and fixed in the sense that they are relatively long-lasting and are or *would* be exhibited in a variety of circumstances. Finally, we ordinarily suppose that character traits, and not just features of a situation, help to explain some of the things a person does. The upshot of this is that a (morally or intellectually) virtuous person would not behave in (morally or intellectually) irresponsible or vicious ways. This at least is what we all expect her to do. Character traits are

counterfactually robust dispositions.[13] With this characterization of virtue ethics and epistemological virtue responsibilism, a particular part of moral psychology, namely the attribution of character traits, comes into sight. If someone says that you have a virtuous character, he presupposes that you have a character and he takes it for granted that there are counterfactually robust and fixed dispositions. If there is no such thing, the Epistemic Character Thesis has no foundation.

Inspired by social psychology, both Gilbert Harman (1999) and John Doris (2002), insist that the conception of character presupposed by virtue ethics is empirically inadequate. The very idea of a character trait has been challenged by empirical research. Experiments in social psychology indicate that, in their actions, people are more vulnerable to changing elements of a certain situation than dependent upon character traits. Take the infamous Milgram experiments: subjects defer to authority if put under some pressure. Thus they can be led step by step to administer very severe electrical shocks to another person. Or a more everyday example: Whether someone in a waiting room will go to the aid of another person who seems to have fallen off a ladder in the next room may depend on whether there is another person in the waiting room who seems unconcerned with the apparent fall (Latané and Darley 1970).[14] Such experiments suggest that subjects do not exhibit counterfactually robust virtuous dispositions. A virtuous subject would not have tried to give electric shocks to others. However, it needs to be noted that those experiments do not by themselves question the assumption that subjects have virtuous character traits. What they actually do illustrate is the tendency of observers to infer wrongly that actions are due to character traits rather than to aspects of the situation. Psychologists refer to this tendency as the fundamental attribution error: observers often wrongly attribute character traits to subjects on the basis of inadequate behavioural evidence.[15]

---

13 Compare the following description of character: 'For Aristotle, as for us, the term has to do with a person's enduring traits; that is, with the attitudes, sensibilities and beliefs that affect how a person sees, acts and indeed lives. As permanent states, this will not explain merely why someone acted this way *now*, but why someone can be *counted on* to act in certain ways.' (Sherman 1989, 1.)

14 More Examples: subjects were five times more likely to help an injured man who had dropped some books when ambient noise was at normal levels than when a power lawnmower was running nearby. Passers-by not in a hurry were six times more likely to help an unfortunate person who appeared to be in significant distress than were passers-by in a hurry. People who had just found a dime were 22 times more likely to help a woman who had dropped some papers than those who did not find a dime. College student's role playing in a simulated prison rapidly descended to grave forms of barbarism (taken from Doris and Stich 2005).

15 A textbook's summary: 'Our notion of traits as broad and stable dispositions that manifest themselves to the same extent in a variety of situations cannot hold water. … In conclusion, it appears that we are truly quite consistent in our behavior within each situation, and it is quite appropriate to expect such consistency in others. But we run into trouble when we expect this consistency to extend to other situations as well. Even slight variations in the features of a situation can lead to dramatic shifts in people's behavior.' (Kunda 1999, 443 and 499.)

Moreover, in referring to traits as the causes for behaviour a problem with consistency emerges: if someone possesses a certain virtue (being courageous, being observant) he is expected to *consistently* behave courageously or in an observant way when it is ethically or epistemically appropriate to do so – if one adopts the globalist moral psychologist's perspective, characterized by the idea that behaviour is governed by robust traits (Doris 2002, 22–3). If behaviour is typically governed by robust traits, systematic observation could but reveal behavioural consistency. But according to social psychology this is not the case. This suggests the falsehood of the globalist moral psychologist's perspective, namely that behaviour is typically governed by robust traits (Doris 2005). The argument thus reveals that the conception of traits, as outlined above, is radically empirically unsupported. To this end, Doris adds (2002, 62–4) that we are fragmented selves with local traits. Thinking of ourselves as having global traits which unify our practical reasoning would be a mistake. Therefore one is to evaluate people not in terms of robust character traits but rather in terms of local, situation-specific traits.[16] These results fit nicely with a common observation: Ernest may be very observant when taking notes for his new novel and he might be most courageous in the face of physical danger. However, he is remarkably unobservant when the emotional needs of his friends are at stake and he is a coward when it comes to confronting personal deficiencies. Someone being courageous might be expected to be honest, too; someone being observant might be expected to be open-minded. Nevertheless, Ernest tells a lot of lies or exaggerates most of the time; and despite being observant he is very conservative and stubborn. In general one might expect that character will be evaluatively integrated, i.e. traits with associated evaluative valences are expected to co-occur in personality. The lack of uniformity in behaviour in different situations adduced from the empirical literature, however, undermines expectations of integrated character structures.

Doris's main argument for situationism can now be formulated with one eye on the empirical results concerning compassionate behaviour: If compassionate behaviour is typically governed by robust traits of character, systematic observation will reveal behavioural consistency across situations strongly conducive to compassion. However, according to social psychology this is not the case. So, behaviour is not typically governed by robust traits and people's compassionate behaviour will typically be inconsistent across situations. According to the argument behaviour is not typically governed by the traits that figure prominently in virtue ethics. The same argument, I suggest, can be applied to responsibilist virtue epistemology: if there is no evidence for character traits governing behaviour, the Epistemic Character Thesis cannot hold. This seems to be an empirical question.[17] Considering

16 *Pace* Montmarquet (2003) the point of Harman's and Doris's critique is not that character traits do not exist. For more detailed discussions concerning the critique of character see Annas (2005).

17 Example: Employers are convinced that useful information can be gained from interviewing potential employees. However, for the most part, interviews simply add noise to the decision process. Empirical studies indicate that decisions made on information available apart from an interview are more reliable than decisions made when an interview is added (Ross and Nisbett 1991, 136–8).

the parallel of moral and epistemic virtues as traits of character, the attack on the former affects the latter. Of course, the results in social psychology do not establish that there are no traits of character at all. They rather undermine confidence in such traits' explanatory power and show that it is not obvious that there are such traits. Therefore, Affective Epistemology should not rely on the notion of character to solve the bundle of problems mentioned in section 3.

Perhaps our characterization of a character trait is a bit too tight? Virtue ethics requires a robust notion of character, but not as something fixed and permanent, holding over non-normal situations (like the one produced in Milgram's experiments). Virtues are not rigid habits, which lead to overconfidence in our ability to see what is salient in particular situations. Virtues might be viewed as dispositions to act on reasons, they are developed through reflection and reasoning, and they are exercised by deliberating and in making decisions. This, of course, requires openness and sensitivity to situations. Such a less restrictive notion of character would be compatible with the results of social psychology (Annas 2005). It would stress the relevance of reason and reflection: an epistemic agent is reliable and trustworthy because he develops and exercises reflection and deliberation. Despite these positive implications Affective Epistemology still faces troubles. First, epistemic immediacy endows inferences, beliefs, doubts or questions with the feeling of being 'compelling' or 'good'. As we have seen, such immediacy is 'unreflective'. Second, emotions are supposed to be intrinsically evaluative, but, as we have seen, only because they are linked to a set of norms, goals and values, regulating our epistemic activities. But this set requires a process of epistemic evaluation. The emotions do not form the basis of this process, but reflection and deliberation. Third, according to the Epistemic Character Thesis epistemic success depends upon emotional salience tracking relevance, which in turn depends upon character traits. But in the light of the revisited notion of character and virtue, epistemic success depends upon reflection and reasoning, exercised by deliberating and decision making. Of course, emotions are part of the material taken into consideration when confronted with epistemological problems. But this does not entail that they are necessarily part of the process of consideration itself. Finally, concerning openness and sensitivity to situations, we have already indicated that it seems much more reasonable to demarcate a field of inquiry in which emotions play a minor part.

All in all this seems to suffice to support the conclusion that the problems for Affective Epistemology set forth in section 3 cannot be solved by the marriage of the emotions with such a dubitable candidate as character.[18]

5. Modest Affective Epistemology

Unlike Hookway, Elgin is not referring to virtuous character. Her account of Affective Epistemology is less demanding and avoids most of the problems I have indicated in section 3, especially the charge of over-inclusiveness. In Elgin's view emotions *can* be sources of reliable information that sometimes affect the epistemic tenability of a

18 For a situation-sensitive account of the notion of character see Upton (2005).

system of beliefs (Elgin 1996, 147). For one thing, according to embodied appraisal theory at least some emotions reliably represent core relational themes as their real content. Given the qualifications limiting the properties emotions can represent, as suggested in the first section, the very fact that emotional responses are reliable indicators of how things stand bestows some degree of epistemic dignity upon them. Furthermore, emotions might shatter somebody's beliefs relevant to a specific field. Feeling suspicious that Ernest is really telling me the truth, the tenability of my beliefs concerning Ernest's trustworthiness has to be revised. Doubting that the scientist really could have been carrying out so many experiments since he left for the industrial laboratory makes Sue suspicious of the putative results. The point of Elgin's discussion is that although emotions may affect the *initial* tenability of belief systems, they do not confer *full* tenability on it (Elgin 1996, 150). They are only imperfect indicators of initial tenability (Elgin 1996, 159). Thus, modest Affective Epistemology is characterizable as the position that sees emotions as indicators of core relational themes and thereby as ways of knowing about certain properties (see section 2). These might be properties of belief systems, indicating something about its initial tenability. The main difference between this modest and the more demanding form of Affective Epistemology criticized in sections 3 and 4 consists in this: The modest form is not intended to regulate inquiry and articulate epistemic standards, and it does not rely upon the notion of character. I have no quarrels with the modest point of view. Modesty, to be sure, might be in itself an epistemic quality, however, not a virtuous trait of character.

Nevertheless, it seems to me that Elgin's version of the modest view is unduly inimical to the Traditional Project by her repeated effort to *replace* knowledge with understanding. However, there are different pictures involved in the juxtaposition of understanding and knowledge. Let me just point out three pictures.

A first picture can be found in the work of Barry Stroud. According to Stroud, epistemology is the project of understanding human knowledge in general, or more precisely: 'We want to be able to apply what we find out about knowledge to ourselves, and so to explain how our own knowledge is possible' (Stroud 2000, 112). What we need is a good reason for accepting some account of knowledge as adequate and applicable to ourselves. In other words, in order to understand knowledge I need to know that my theory of knowledge does really apply to myself, the knower; I need to know that the world exists in a certain way, a way in accordance with my account of what knowledge is. As Stroud has it, epistemology is intimately tied to self-understanding (Stroud 2000, ch. 8 and 10).

A second picture recommends a shift of perspective from knowledge to understanding. According to Elgin, the Traditional Project is truth-centred: knowledge is directed at objective facts and expressed in true propositions or sentences. Knowledge, therefore, comes in discrete bits, it is granular. This conception seems to run counter to scientific and other epistemic practices. Science, for example, is holistic. There is no sentence by sentence verification. Scientific theories face the tribunal of experience as a corporate body, not in discrete bits (Elgin 2006). As epistemologists, she continues, we had better be in a position to appreciate the cognitive achievements of science and other epistemic practices. Epistemology's emphasis on factive knowledge, however, seems to prevent this. Elgin holds that

science uses idealizations and models which do not mirror the facts. She defends, therefore, a more generous, flexible conception of epistemology that accommodates science and reflects our epistemic practices beyond which restrictive factive orientation is called for (Elgin 2007). In this spirit she suggests 'understanding' to be the better alternative to denote our epistemic achievements. Understanding can grow and advance (Elgin 1996, 123ff.), which means that our capacities for 'recognition, reasoning, representation, and explanation become better and more refined' (Elgin 2006, 215). As Elgin has it, understanding (the Inclusive Project) replaces knowledge (the Traditional Project).[19]

It is not easy to understand the concept of understanding. One could deal with this by opposing understanding to knowledge. In the third picture, knowledge is one thing and understanding another. As I have already pointed out in section 2, knowledge is an absolute concept, not allowing for degrees. Understanding, on the contrary, is a matter of degree. Sue and I, we both know that Hemingway wrote *For Whom the Bell Tolls*, we both know this fact and have a true belief, but being a literary critic and a historian of the Spanish Civil War Sue has a better understanding of this fact. Why? Her belief that Hemingway wrote *For Whom the Bell Tolls* relates, as it may be, to other beliefs about war and fiction or about the role of intellectuals in the Spanish Civil War. According to the definition advanced above, understanding is not directed at a single proposition; or more precisely, the understanding of a proposition is not primarily directed at the proposition itself, rather one understands the proposition in question as part of and because of one's understanding of a network of other beliefs. Knowledge, on the other hand, is directed at a single proposition. Knowledge aims at truth, while what is understood is not always believed or established as true, of course. Knowledge in principle is absolute, piecemeal and factive, whereas understanding is gradual, interconnected and reflective. Now, think of knowledge in terms of virtue reliabilism: intellectual virtues are reliable or truth-conducive cognitive faculties, including belief formation based on sight, hearing, memory or reasoning in approved ways. These cognitive faculties are widespread among humans and other animals. The epistemic properties of beliefs are to be defined in terms of the reliable intellectual virtues of an agent. Understanding, on the other hand, is both a matter of the use the agent makes of these virtues while inquiring into an epistemic field, and a matter of the explanations or reasons the agent finds, asks for or gives for the things she believes to know. Knowledge involves beliefs about one's surroundings, one's past or one's own mind, and such beliefs are direct responses to the surroundings, the past or the mind. As we have seen, these responses might involve emotions as appraisals. The notion of understanding places these responses in a wider whole, including, for example, 'recognition, reasoning, representation and explanation' (and, if possible at all, Stroudian self-understanding). A good example of the interrelation between

19 This picture goes back, of course, to Elgin's and Nelson Goodman's collaboration: 'We want here to sketch an alternative epistemology … . Epistemology, as we conceive it, comprehends understanding or cognition in all its modes – including perception, depiction, and emotion, as well as description. … Finally, knowledge, plagued with certainty and uncertainty alike, gives way in our reconception to *understanding*.' (Goodman and Elgin 1988, 4–5, 161).

knowledge and understanding is Marlowe looking at the photograph. He sees all the things he needs to know but he does not understand them. He keeps looking at it. And in a little while he understands what he knows by seeing it with his eyes: the man is falling. In sum, affective states such as moods, emotions or feelings, might play a role in the growth of understanding the way Elgin thinks they do. There is, however, no reason to replace knowledge with understanding.

References

Annas, Julia. 2005. 'Comments on John Doris' Lack of Character'. *Philosophy and Phenomenological Research* 71: 636–42.

Axtell, Guy. 1997. 'Recent Work in Virtue Epistemology'. *American Philosophical Quarterly* 34: 1–27.

Baehr, Jason. 2006. 'Character, Reliability, and Virtue Epistemology'. *Philosophical Quarterly* 56: 193–212.

Chandler, Raymond. 1988. *The High Window*. New York: Random House.

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.

Doris, John. 2002. *Lack of Character*. Cambridge: Cambridge University Press.

Doris, John. 2005. 'Précis of Lack of Character'. *Philosophy and Phenomenological Research* 71: 632–5.

Doris, John; Stich, Stephen. 2005. 'As a Matter of Fact. Empirical Perspectives on Ethics'. In: Jackson, Frank; Smith, Michael (eds). *The Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press. 114–53.

Dretske, Fred. 1981. 'The Pragmatic Dimension of Knowledge'. *Philosophical Studies* 40: 363–78.

Dretske, Fred. 1995. *Naturalizing the Mind*. Cambridge, Mass.: MIT Press.

Elgin, Catherine Z. 1996. *Considered Judgment*. Princeton: Princeton University Press.

Elgin, Catherine Z. 2006. 'From Knowledge to Understanding'. In: Hetherington, Stephen (ed.). *Epistemology Futures*. Oxford: Clarendon Press. 199–225.

Elgin, Catherine Z. 2007. 'Understanding the Facts'. *Philosophical Studies* 132: 33–42.

Elster, Jon. 1999. *Alchemies of the Mind. Rationality and the Emotions*. Cambridge: Cambridge University Press.

Goodman, Nelson; Elgin, Catherine Z. 1988. *Reconceptions in Philosophy and Other Arts and Sciences.* London: Routledge.

Gordon, Robert M. 1969. 'Emotions and Knowledge'. *Journal of Philosophy* 66: 408–13.

Greco, John. 2002. 'Virtues in Epistemology'. In: Moser, Paul K. (ed.). *The Oxford Handbook of Epistemology*. Oxford: Oxford University Press. 287–315.

Griffiths, Paul E. 1997. *What Emotions Really Are. The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Gunther, York H. 2004. 'The Phenomenology and Intentionality of Emotions'. *Philosophical Studies* 117: 43–55.

Harman, Gilbert. 1999. 'Moral Philosophy Meets Moral Psychology. Virtue Ethics and the Fundamental Attribution Error'. *Proceedings of the Aristotelian Society* 99: 315–31.

Hetherington, Stephen. 2001. *Good Knowledge, Bad Knowledge. On Two Dogmas of Epistemology.* Oxford: Clarendon Press.

Hookway, Christopher. 1993. 'Mimicking Foundationalism'. *European Journal of Philosophy* 1: 156–74.

Hookway, Christopher. 1998. 'Doubt. Affective States and the Regulation of Inquiry'. In: Misak, Cheryl (ed.). *Pragmatism*. *Canadian Journal of Philosophy,* suppl. vol. 24: 203–25.

Hookway, Christopher. 2000. 'Regulating Inquiry. Virtue, Doubt, and Sentiment'. In: Axtell, Guy (ed.). *Knowledge, Belief, and Character. Readings in Virtue Epistemology*. Lanham, Md.: Rowman and Littlefield. 149–60.

Hookway, Christopher. 2002. 'Emotions and Epistemic Evaluation'. In: Carruthers, Peter et al. (ed.). *The Cognitive Basis of Science*. Cambridge: Cambridge University Press. 251–62.

Hookway, Christopher. 2003a. 'Affective States and Epistemic Immediacy'. *Metaphilosophy* 34: 78–96.

Hookway, Christopher. 2003b. 'How to Be a Virtue Epistemologist'. In: DePaul, Michael R.; Zagzebski, Linda (eds). *Intellectual Virtue. Perspectives from Ethics and Epistemology*. Oxford: Clarendon Press. 183–202.

Hursthouse, Rosalinde. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.

James, William. 1890. *The Principles of Psychology*. Vol. 1. New York: Henry Holt.

Kunda, Ziva. 1999. *Social Cognition. Making Sense of People*. Cambridge, Mass.: MIT Press.

Kvanvig, Jonathan. 2005. 'Truth Is Not the Primary Epistemic Goal'. In: Sosa, Ernest; Steup, Matthias (eds). *Contemporary Debates in Epistemology*. Malden: Blackwell. 285–95.

Latané, Bibb; Darley, John M. 1970. *The Unresponsive Bystander. Why Doesn't He Help?* New York: Appleton-Century-Crofts.

Lazarus, Richard S. 1991. *Emotion and Adaption*. New York: Oxford University Press.

McGinn, Marie. 2003. 'Reply to Hookway'. *Metaphilosophy* 34: 97–105.

Mangan, Bruce. 2001. 'Sensation's Ghost. The Non-Sensory "Fringe" of Consciousness'. *Psyche* 7/18. http://psyche.cs.monash.edu.au/v7/psyche-7-18-mangan.html.

Montmarquet, James. 2003. 'Moral Character and Social Science Research'. *Philosophy* 78: 355–68.

Mulhall, Stephen. 1996. 'Can There Be an Epistemology of Moods?' In: Anthony O'Hear (ed.). *Verstehen and Humane Understanding.* (*Royal Institute of Philosophy Supplement* 41). Cambridge: Cambridge University Press. 191–210.

Prinz, Jesse J. 2004. *Gut Reactions. A Perceptual Theory of Emotion*. New York: Oxford University Press.

Ross, Lee D.; Nisbett, Richard E. 1991. *The Person and the Situation. Perspectives of Social Psychology*. New York: McGraw-Hill.

This page intentionally left blank

Chapter Seven

Misleading Emotions

Peter Goldie

Over recent years there has been an optimistic trend in empirical and evolutionary psychology, and also more widely amongst philosophers of a broadly 'cognitivist' bent, which emphasizes the usefulness of emotion in picking up saliences in the environment, and enabling quick and effective action with little or no conscious deliberation. This optimism, I believe, deserves to be tempered with some realism. The emotions can systematically mislead us, and they can do so in ways that are systematically hard to detect and correct. This is especially the case in respect of those emotional capacities which evolved in environments that differ in important respects from the environment in which we now live.

1. Introduction

It is an orthodox view these days, and a correct one I think, that our emotional capacities and dispositions enable us to perceive certain things in our environment as salient, and to respond emotionally to those things in ways that are advantageous for us (see, for example, Elgin, this vol.; Tanesini, this vol.; de Sousa 1987; Griffiths 1997; Greenspan 2006). Roughly speaking, we are better off, as individuals and as a species, in having such capacities and dispositions. For example, when we are in a dangerous situation, we are able to perceive things as dangerous, and to respond appropriately with fear and with evasive action, doing so at a speed and with a lack of conscious deliberation that we would not be able to achieve without our capacity for fear. It may well be that, as Joseph LeDoux (1998) has argued, we have a 'high road' and a 'low road' for processing this kind of information – a more complex, slower route, and what is sometimes called a 'quick and dirty' or 'fast and frugal' route. This idea is now taken very seriously by empirical psychologists in the form of what is often called 'dual process' theory. The two processes have various names, but I will call them 'intuitive thinking' and 'deliberative thinking'.[1] They are meant to complement each other. Intuitive thinking, involving emotion and imagination, operates fast, does not involve conscious thinking and plays a vital epistemic role in a world in which energy resources are limited and speed of response is of the essence. Deliberative thinking, in contrast, is more 'cool', involves conscious deliberation and has as one of its functions operating as a check or balance on intuitive thinking.

1 See, for example, Haidt (2001); Haidt (2007); Greene et al. (2001); Gigerenzer (2000). I discuss some related issues in Goldie (2007).

But both processes are cognitive; the old idea, if there ever was such an idea, of the emotions as mere irrational or arational urges is long gone.

I willingly accept the idea that emotions and intuitive thinking can and do play this role of helping us to manage our way through the world under constraints of finite time and energy resources. But I want to add a more pessimistic note about the role of deliberative thinking as check and balance. The emotions, I will argue, can and often do *systematically* mislead us, and this gives rise to some interesting and practically challenging epistemic issues which, as will emerge, have some connections to the so-called heuristics and biases tradition in empirical psychology.[2] The 'systematically' is important here. Of course there is nothing fundamentally problematic with the occasional misleading emotion – for example, with the occasional fear felt towards something that is not really dangerous. It might get you into difficulties at the time, but it does not throw into question the role of emotion in general, any more than the occasional false belief throws into question the role of our belief-forming capacity in general. The concern that I want to raise about emotion here goes deeper than that. Furthermore, emotions not only systematically mislead us, but they also do so in ways that can be very hard to detect – and to correct. So we cannot always safely rely on epistemic checks and balances on emotion and intuitive thinking from the slower and 'cooler' processes of thought involved in deliberative thinking. Even if the old idea of the emotions as irrational or arational is long gone, we should avoid a recoil into the idea of the emotions as thoroughly ordered, and nicely and reliably aligned with reason. On the contrary, they can both undermine reason and disguise the fact that this is what they are doing.

2. Misleading Saliences

In general, our emotional capacities and dispositions 'filter' our perception of our surroundings (Wollheim 1999; Roberts 2003; Döring, this vol.), and to that extent emotional salience and emotional arousal are generated according to what those capacities and dispositions are. So, for example, a man who is disposed to find cows frightening will see them as frightening, dangerous and liable to stampede in ways that the rest of us would not when confronted with a herd of cows, and he will feel fear where we would not. Salience and arousal go together.

Salience and emotional arousal can be distorted because of irrational emotional dispositions, and I dare say that a morbid fear of cows is one such. But I am not concerned here with phobias, obsessions and the like. I am more interested in those cases where there is nothing irrational about the emotional disposition as such, but where it can still be systematically misleading. To see what I mean, let us start with David Hume, who made the point in relation to our moral sentiments. We will see that the point turns out to be quite general.

'Morality', Hume famously said, 'is more properly felt than judged of' (Hume [1739/40], 470).[3] But he made it clear that not *any* feelings are relevant

2 See, for example, Nisbett and Ross (1980), and Kahneman, Slovic and Tversky (1982).

3 Throughout I have modernized Hume's spelling.

in distinguishing moral good and evil. If we are to engage in moral thought and discourse, two kinds of correction, using reason and imagination, are required. First, in judging the morality of an action we must put to one side the feelings we have which result from the contingencies of our own particular relations with the protagonists. For example, the fact that the Good Samaritan happens to be my son should not influence my moral judgement of his action, even though his kindness and generosity are more salient for me, and even though I accordingly feel more admiration, just *because* he is my son (Hume [1739/40], 472). This is not to say that I should necessarily avoid having the feelings proper to my particular relationship with my son, but rather that these feelings should be kept apart from my feelings about the morality of his action. Indeed, as Hume observed, these two kinds of feeling can even be contrary: 'The good qualities of an enemy are hurtful to us; but may still command our esteem and respect … as when the fortifications of a city belonging to the enemy are esteemed beautiful upon account of their strength, though we could wish that they were entirely destroyed' (Hume [1739/40], 586–7). So the first thing we have to take into account and adjust for is the particularity of relationships: the same act of kindness done by a friend, an enemy or a complete stranger should be judged to have the same moral properties even though the salience of the action's moral properties, and our moral feelings in response, are different.

Secondly, our moral sentiments vary depending on the contingent proximity or remoteness of the object of our sentiments, and we need to correct for these variations in sentiment too when arriving at moral judgements of virtue and vice (Hume [1739/40], 581). What is nearer is more salient than what is farther away, and thus has a greater effect on our sentiments. Hume draws this analogy: our judgement of the moral esteem of someone distant should be corrected by reason, just as we should judge that an approaching object is not really getting any larger even though it might appear to be getting larger (Hume [1777], 227–8).

So reason has an important role here, to correct for the biases to our moral judgements that can arise from these two influences on our moral sentiments. (In this sense one can see him, perhaps rather controversially, as a kind of proto-dual-process theorist; more of this later.) If our moral judgements and discourse were slavishly to follow our sentiments, as Hume says, 'it is impossible we could ever converse together on reasonable terms', and '[i]n order, therefore, to prevent these continual *contradictions*, and arrive at a more *stable* judgement of things, we fix on some *steady* and *general point of view*; and always, in our thought, place ourselves in them, whatever may be our present situation' (Hume [1739/40], 581–2). Hume made the point many times over, both in the *Treatise* and in the *Enquiry*, that this requirement, to take a steady and general point of view, is essential to our moral discourse.[4]

4 'It is impossible men could ever agree in their sentiments and judgements, unless they chose some common point of view' (Hume [1739/40], 591). 'The intercourse of sentiments … makes us form some general inalterable standard' (Hume [1739/40], 603; Hume [1777], 229). 'General language, therefore, being formed for general use, must be moulded on some general view' (Hume [1777], 228). 'The notion of morals implies some sentiment common to all mankind, which recommends the same object to general approbation, and makes every man, or most men, agree in the same opinion or decision concerning it' (Hume [1777], 272).

But it is at this point that Hume reveals his pessimism about the corrective power of reason. We might manage to make moral 'pronouncements' based on our reasoned adjustments to our sentiments, in line with the general requirement to take a steady and general point of view, but nevertheless the heart seldom follows the head: 'I do not feel the same lively pleasure from the virtues of a person who lived in Greece two thousand years ago that I feel from the virtues of a familiar friend and acquaintance' (Hume [1739/40], 581). Whilst reason might enable me to judge that my friend and the distant Greek are equally virtuous, my close relationship with my friend, and his proximity, still have their distorting influence on my passion: the general principles for correction by reason 'are not altogether efficacious, nor do our passions often correspond entirely to the present theory' (Hume [1739/40], 583). If our passions do not entirely fall into line with correction by reason, then, one might think, it is likely that our motivations will also follow our heart and not our head – that they will follow the direction of our sentiments and not that of our judgements or discourse. I will return to this important point shortly.

The question that I now want to address is whether Hume's idea of how sentiments or passions can bias judgement generalizes beyond the moral domain. Do the same biases arise elsewhere too? According to the work of the psychologists Richard Nisbett and Lee Ross, and of Daniel Kahneman and Amos Tversky, founders of the heuristics and biases tradition, indeed they do. We can immediately see the connections between their work and Hume's moral philosophy. Let us consider Nisbett and Ross, who, in Chapter 3 of their *Human Inference: Strategies and Shortcomings of Social Judgment* (Nisbett and Ross 1980), discuss how emotions can have misleading or distorting effects on our inferences according to what they call the 'vividness criterion'. Information is vivid, they say, when it is 'likely to attract and hold our attention and to excite the imagination to the extent that it is (a) emotionally interesting, (b) concrete and imagery-provoking, and (c) proximate in a sensory, temporal or spatial way' (Nisbett and Ross 1980, 45).

Regarding (a), emotional interest, events tend of course to have more emotional interest if we are ourselves directly involved. And, as Nisbett and Ross add, 'events also are more interesting when they happen to people we know than when they happen to people we do not know, and they are more interesting when they happen to people about whom we have strong feelings than when they happen to people about whom we have neutral feelings' (Nisbett and Ross 1980, 46). Regarding (b), a concrete and imagery-provoking event will be one which, roughly, has more emotional content in virtue of the way what happened is grasped; the more emotional detail there is available, the more salience and impact it will have. For example, a short report in the newspapers of a flood in the Philippines with five thousand dead will have less emotional detail than a report on television with close-up footage of the devastation and of the drowned bodies. And discussing the proximity of information, (c), Nisbett and Ross give this example: 'The news that a bank in one's neighbourhood has been robbed just an hour ago is more vivid than the news that a bank on the other side of town was robbed last week' (Nisbett and Ross 1980, 49). These three elements of the vividness criterion are typically found together, even if they are conceptually distinct (Nisbett and Ross 1980, 45).

Nisbett and Ross go on to argue that the vividness of information affects inferences. (The epigraph to the chapter is a quotation from Bertrand Russell: 'Popular induction depends upon the emotional interest of the instances, not upon their number'.) When information is what they call 'pallid', as dry statistical evidence is, it tends to be given less inferential weight than the evidence of one's own eyes or evidence that is otherwise more salient. The robbery in one's own neighbourhood an hour ago, being more vivid that the robbery last week on the other side of town, 'is likely to have a greater impact on one's views of the seriousness of the crime problems in one's city or the need for stiffer prison sentences for bank robbers' (Nisbett and Ross 1980, 50). And the vivid television footage of the flood in the Philippines is likely to make one more concerned about the inadequacy of the flood defence systems than is the pale newspaper report.

A direct line-by-line comparison of Hume's account with that of Nisbett and Ross would not be fruitful without a lengthy prior examination of their philosophies of mind and of action. Nevertheless, it seems clear that what their accounts have in common is the idea that our judgements and inferences about empirical facts can be systematically biased by emotion in cases where emotions pick up on certain saliences that are not relevant for the judgement or inference. The question then is how these biases affect motivation.

Sometimes, of course, vividness of information can have a beneficial effect on motivation, getting one to do something that one ought to have done anyway, but never got around to. Nisbett and Ross give a nice example of how, in 1974, the much-reported and much-discussed mastectomies performed on Mrs Ford and Mrs Rockefeller led to a mass of visits for cancer checkups, whereas in the past widely disseminated statistics about the risks of breast cancer had produced nothing like such a significant effect on behaviour. But saliences do often mislead one into error, leading one to attach too much weight to certain considerations and thus leading one to act wrongly–ethically or prudentially.

Towards the end of their chapter on misleading saliences and the vividness criterion, Nisbett and Ross accept, as I do on my own behalf, that they have chosen to emphasize the 'serious inferential and behavioural costs' associated with emotional saliences, the reason for this pessimism being, simply, that '[t]he vividness of information is correlated only modestly, at best, with its evidential value' (Nisbett and Ross 1980, 60). But they then go on to make some speculative remarks about the reasons why we are subject to making these errors in reasoning and behaviour as a result of emotional saliences. This will help to lead us towards the next thing that I want to discuss – how misleading emotional saliences arise through environmental mismatch. This is what they say:

> During all but the most recent moments of our evolutionary history, dangers and opportunities have been relatively concrete and vivid. … Now, however, our world has come to have pressing dangers which are complex and abstract matters … , dangers best described by abstract and often statistical information. (Nisbett and Ross 1980, 60)

The suggestion, then, is that what used to be advantageous is now not so, because the environment has changed in respect of its increasing complexity, and in respect

of our increased knowledge of its complexity. We will see that the environment has changed not only in this way, but in others too.

3. Environmental Mismatch

Let us begin with the notion, recently developed by Gerd Gigerenzer and his collaborators, of 'bounded rationality', and see how it relates to the heuristics and biases tradition of Nisbett and Ross and Kahneman and Tversky. The central idea of bounded rationality, emerging from the discipline of evolutionary psychology, is that humans, under the constraints of time and limited information, do not typically reason using the accepted standards of decision theory. We act on the strength of a collection of rules and heuristics embodied in what they call an 'adaptive toolbox', which has the following features:

> First, it [is] a collection of rules or heuristics rather than … a general-purpose decision-making algorithm … . Second, these heuristics are fast, frugal, and computationally cheap rather than consistent, coherent, and general. Third, these heuristics are adapted to particular environments, past or present, physical or social. … Fourth, the bundle of heuristics in the adaptive toolbox is orchestrated by some mechanism reflecting the importance of conflicting motivations and goals. (Gigerenzer and Selten 2001, 9)

Emotions, they say, can themselves be heuristics – let us call them 'emotion-based heuristics'. Gigerenzer's example is disgust, which operates only within the domain of things we choose to ingest, and which serves the adaptive function of preventing poisoning (Gigerenzer 2001, 42). If the waiter in the restaurant presents us with a green steak on the plate, we will feel disgust towards it and refuse to eat it, even if we are assured that it is harmless colouring.[5]

We can see that there is agreement between the recent work of Gigerenzer and colleagues on bounded rationality, and that of Nisbett and Ross and Kahneman and Tversky in the heuristics and biases tradition, in this respect: they agree that our emotion-based heuristics have a very important role in our intuitive responses to the world around us, and as a result in the ways we act. However, whereas the earlier researchers put the emphasis on the errors and biases that can arise, Gigerenzer and colleagues say that 'bounded rationality is not an inferior form of rationality' (Gigerenzer and Selten 2001, 6). Peter Todd, writing in the same volume, registers the disagreement by commenting that 'the basic message of [the heuristics and biases] research program … is that humans use heuristics at their peril, more often than not making errors of judgment and inaccurate decisions. … In contrast, the vision of ecological rationality emphasizes that humans use specific simple heuristics because they *enable* adaptive behaviour, by exploiting the structure of information in natural decision environments. Simplicity is a virtue, rather than a curse' (Todd 2001, 52–3). So the rationality wars, as they are sometimes called, are not over the

5 See also Todd and Gigerenzer (2000, 740 and 741) where the precise role of emotion is discussed as one of the 'challenges that remain'. Adam Morton (2000) discusses the possibility that a heuristic can also play the role of 'orchestration'.

role of emotion in our reasoning, nor are they over the emergence of our emotions from our evolutionary past. Rather, the rationality wars are, at least in part, over whether or not heuristics, including emotion-based heuristics, in general serve our interests *in the environment in which we now live*: in other words, whether we should be pessimistic or optimistic about the epistemic role of heuristics, and thus, we can add, of emotion-based heuristics in particular.[6]

The answer to this question will depend, in part, on whether or not there is a match between, on the one hand, the actual environment in which we now live and, on the other hand, the related domain-specific emotion-based heuristic. 'Heuristics that are matched to particular environments allow agents to be *ecologically rational*, making adaptive decisions that combine accuracy with speed and frugality. The degree to which a match exists … determines how accurate a heuristic is' (Gigerenzer 2001, 46). I will argue, in respect of at least three types of emotion, that there is a *systematic* environmental mismatch, that this mismatch *systematically* leads to wrong intuitive thinking, and thus to wrong motives and wrong actions, and furthermore, that this mismatch is *systematically* not easy to detect or correct through reason, through deliberative thinking. To repeat what I said at the outset, the concern I want to raise is not the bare possibility of what are called 'false positives'. These can be no bad thing: better to jump at a stick on the path, fearfully taking it to be a snake, than not to jump at a snake on the path, blithely taking it to be a stick. The concern goes much deeper than false positives.

Let us start with male aggression. The anthropologist Daniel Fessler tells of a case from his fieldwork in south-western Sumatra which illustrates well what I have in mind. Late one night a man, Rustram, and his girlfriend hired a minibus to take them home. In a senseless dispute over the fare (the discrepancy was tiny by any standard), a fight began and the driver and his friend stabbed Rustram fourteen times and left him for dead by the roadside with his girlfriend. Discussing what happened, some of the villagers blamed it on possession by the Devil, but another wisely said it was because of *malu*: 'No one wants to be *malu* in front of a girl' (Fessler 2001, 194). Literature, history and empirical psychology are replete with such examples, and we have all witnessed or been involved in incidents of men getting wildly angry and out of control; something like it happens most nights in most big cities, so often over some trivial matter, such as who was first in the queue at the bar.

Malu is an emotion which is something like a simpler version of shame, and its converse, *bangga*, is something like a simpler version of pride – Fessler calls them Protoshame and Protopride. They are based on what he calls a three-point logic, in which the other 'is viewed not as a target for intersubjectivity, but merely as a

6 J.-L. Bermúdez (2000) discusses an important issue, which I will not directly address: what the status of bounded rationality is in normative epistemology, as contrasted to descriptive epistemology; in other words, whether or not Gigerenzer and other proponents of bounded rationality are claiming that this is the way we *ought* to reason, as well as claiming that this is the way we, in fact, do reason. For an irenic discussion of the rationality wars see Samuels, Stich and Bishop (2002), and Samuels and Stich (2003).

feature of the social world' (Fessler 2001, 197), and can thus be emotions which are homologous with emotions in other animals.[7]

Although Protoshame and Protopride (unusually, Fessler remarks, English does not have a term for them) are thoroughly dysfunctional for human beings in today's environment, they were not always so. Fessler speculates on their adaptiveness:

> Because Protoshame, an aversive emotion, is elicited by subordination, while Protopride, a rewarding emotion, is elicited by dominance, individuals capable of experiencing these emotions would have been motivated to seek out higher rank … [and hence] would have greater reproductive success than those who lacked these traits. (Fessler 2001, 198)

We need not be committed to this particular explanation of why Protopride or Protoshame are selectionally advantageous, of why they are 'adaptations' or products of natural selection which evolved for adaptive reasons over the species' evolutionary past (Barkow, Cosmides and Tooby 1992). All we need is the idea that they play a role in the emotion-based heuristic concerned with male aggression towards other males, and that in an ancestral environment they were adaptive – for some reason or other. But today, of course, we have an environmental mismatch. Not only are violent disputes between males over trivial matters often ethically problematic, with the dominant man taking advantage of his physical superiority. This kind of aggression also now characteristically fails to serve the purpose that it did: disputes get out of control with no one the winner in the end, often also against the best interests not only of the individuals concerned but also of the wider community. At the time of the dispute the protagonists will do everything in their power to win the day, sacrificing much else of what they value when they are not in the grip of the emotion: their health, their liberty, their dignity and much else besides.[8] Leaving all this to one side, there is also the terrible damage that can be done to someone's psyche through sublimation of his aggression against other men into some other activity.[9]

The second kind of domain-specific emotion-based heuristic which has an environmental mismatch is fear and mistrust of strangers – xenophobia. The xenophobe tends immediately and unreflectively to react adversely to those who are 'not like us', and to treat them with suspicion or ever worse. The 'fast and frugal' nature of such responses has some quite robust support from the well-known Harvard implicit association tests for race, in which we are found to respond much more quickly to terms such as 'good' and 'nice' when juxtaposed with white faces than we do when they are juxtaposed with black ones.[10]

---

7 The three-point logic for Protoshame is (1) the other is assessed as more important than the self; (2) the self must interact with the other in which (1) will be salient; and (3) as a result the self feels Protoshame or *malu* (Fessler 2001, 195–6).

8 One might argue that what we have in such cases is not irrationality but simply radically shifting preferences (see Elster 1999b), but this is surely rescuing rationality at the cost of common sense. See also Fessler (2001, 208–9).

9 These emotions can also be economically damaging, with males competing pointlessly to have the 'best' positional good such as the biggest house or the fastest car.

10 See https://implicit.harvard.edu/implicit/.

Jim Hopkins has argued that the tendency to what he calls 'ingroup cohesion and outgroup conflict' appears in individual motivations as 'two sides of the same evolutionary coin' – once adaptive, now dysfunctional, and, however much 'we moderns think ourselves guarded against such modes of thought, the roots remain active' (Hopkins 2004, 225). The evidence for this is all around us, in conflicts as far apart as local gang wars in South London and tribal disputes in remote parts of Africa.

Finally, there is male sexual jealousy, which, like male aggression, is supposed to be pan-cultural and shared with other animals. Let us assume, as a number of evolutionary psychologists have argued, that the capacity for male sexual jealousy is selectionally advantageous (see for example Wilson and Daley 1992); again the details of the explanation do not concern us here. This would then explain why we now have this domain-specific emotion-based heuristic, and it is, indeed, a very clear example of the kind of fast and frugal thinking that Gigerenzer and colleagues have in mind.[11] It involves an amazing capacity to pick up the saliences of a possible or imagined transgression: the slightest sign, 'trifles light as air', can be enough. The mismatch that is at issue here is, as with male aggression, in part prudential but most of all it is an ethical one: we now know that male sexual jealousy, with its inbuilt capacity to get violently out of control, is wrong (see Taylor 1988; for a nuance on Taylor see Goldie 2000). In Patrick Marber's play, *Closer*, the doctor Larry blames his behaviour on the fact that he is just a 'caveman'; but this is not even an excuse, let alone a justification.

With this discussion in mind, we can now return to our earlier discussion of morality, and see indications here too of environmental mismatch in the influence on moral sentiments of the particularity of our relationships and of the proximity of the object of our moral attitudes. Examining these connections in the light of recent work in moral psychology on dual process thinking will reveal why environmental mismatch in general, whether in the domain of morality or elsewhere, gives rise to specifically *epistemic* concerns and difficulties.[12]

In a number of important papers, Paul Slovic, writing in the heuristics and biases tradition, has argued that a principal reason why we fail to act when we become aware of genocide is because of failure to be properly *affected*: 'the statistics of mass murder or genocide, no matter how large the numbers, fail to convey the true meaning of such atrocities. The reported numbers of deaths represent dry statistics … that fail to spark emotion or feeling and thus fail to motivate action' (Slovic 2007, 79). Slovic's explanation of this is highly persuasive. He argues that what he calls 'System 1 thinking', or what I call 'intuitive thinking', is significantly affected by the vividness criterion. Intuitive thinking not only involves emotion but also imagination, in particular perceptual imagination, which in turn affects emotion, motivation and action so that, in accordance with the vividness criterion, the amount

11 Gigerenzer makes it clear that not all heuristics are innate, so new tools can be added to the toolbox through learning. Nevertheless, '[for] some important adaptive tasks … there would be strong selective advantages in coming into the world with at least some heuristics already wired into the nervous system' (Todd and Gigerenzer 2000, 768).

12 I am grateful to the editors for pressing me to say more about what follows.

by which one is affected emotionally by the suffering of others typically does not increase proportionately with the numbers of people involved. 'People are much more willing to aid identified individuals than unidentified or statistical victims', as Slovic (2007, 88) says.

Slovic's explanation is that there is an environmental mismatch (although he does not use the term): 'System 1 [intuitive] thinking evolved to protect individuals and their small family and community groups from present, visible, immediate dangers. This affective system did not evolve to help us respond to distant, mass murder' (Slovic 2007, 84). Slovic adduces some surprising evidence to support the claim that we have a diminishing sensitivity to human suffering – what he calls 'psychophysical numbing' – and that this begins to have its effects at a very early point on the scale of increasing numbers. In one study, participants contributed much more to a $300,000 fund when that fund was said to be for the treatment of just one child than when it was said to be for the treatment of eight children (Slovic 2007). This result emerged even though the members of the group of eight children were individually identified; even here more distress and compassion is felt towards the identified single child. 'Left to its own devices,' says Slovic, 'System 1 [intuitive thinking] will likely favor individual victims and sensational stories that are closer to home and easier to imagine. It will be distracted by images that produce strong, though erroneous, feelings, like percentages as opposed to actual numbers' (Slovic 2007, 91). Thus our intuitive thinking in the moral domain, and our moral motivations and actions, seem to ignore, or somehow not to be properly influenced by, what should be a plain moral truth: individual lives matter equally, whether that life is considered by itself or as one amongst a million, and whether that life is vividly salient or not.

4. Can we Correct for our Misleading Emotions?

Slovic says, following Kahneman (2003), that 'one of the important functions of System 2 [deliberative thinking] is to monitor the quality of mental operations and overt behaviors produced by System 1' (Slovic 2007, 91). The point is really a general one, extending beyond the moral domain. If we assume that humans do have, broadly speaking, dual processes of thinking, intuitive and deliberative, then Slovic is surely right that at least one of the roles of deliberative thinking is a 'monitoring' one – to act as a kind of epistemic check and balance on fast and frugal intuitive thinking. The question now arises, in relation to the systematically misleading emotions that I have been considering – male aggressiveness, xenophobia and male sexual jealousy, as well as morality – whether deliberative thinking is up to the job.

In the domain of morality, Slovic was doubtful of the motivating powers of deliberative thinking. Whatever we might judge, our actions seem to follow our intuitions. We are, as Slovic puts it, in the 'twilight between knowing and not knowing' (Slovic 2007, 82, citing Power 2003, 505). The parallel with Hume's ideas are quite striking. Hume too, as we have seen, thought that the role of reason was in part to correct for the misleading influences of our sentiments (and this was why I chose to describe him as a proto dual-process theorist). Yet he, like Slovic, was pessimistic

about reason's corrective powers, and about whether ultimately motivation and action will follow the heart or the head. As Hume put it with characteristic wit, 'though the heart does not always take part with those general notions, or regulate its love and hatred by them, yet are they sufficient for discourse, and serve all our purposes in company, in the pulpit, on the theatre, and in the schools' (Hume [1739/40], 603).

The epistemic task of unemotional deliberative thinking, to monitor and correct fast and frugal emotional-based intuitive thinking, has a place wherever the two kinds of thinking potentially clash. The task is not only for deliberative thinking to be able to sort out the good reasons from the bad, the right perspectives from the biases, and so on. It seems that, in the moral cases which concerned Hume and Slovic, it is capable of this – at least for the purposes of moral discourse. The task is also to turn these good reasons into good *motivating* reasons; and it is in just this sense, it seems, that deliberative thinking, at least in the moral domain, fails. We are indeed in the twilight between knowing and not knowing when we say, as we so often do, how terrible is the genocide reported on the latest news, and then press on to worry only about the single murder in our own home town. Categorizing the failure of deliberative thinking in this second task as a kind of 'knowing-but-not-knowing' helps one to see both that the failure is an epistemic one, and also to see that there is an important connection between the failure and weakness of the will. One might put it like this: if we *really* knew how terrible the genocide was, then we would try to do something about it.

In respect of the three other domains that I have been considering, sexual jealousy, aggression and xenophobia, the epistemic difficulties take different forms. To correct for a misleading emotion, we must first recognize it as such – that is, we must recognize that we are experiencing an emotion, and we must recognize that we are being misled by it. This can be particularly worrying, because of the propensity of misleading emotion to 'mask' its misleadingness. There are at least two ways in which it can do this.[13]

The first way in which emotions and intuitive thinking can mask their misleadingness is manifested in the emotion-based heuristic of sexual jealousy. The emotion *skews the epistemic landscape*. Trifles light as air *seem* like irrefutable evidence of transgression. And evidence which we might otherwise, through cool and calm deliberative thinking, take to count *against* our emotion we now ignore, or even take to be confirmatory of our suspicions. Leontes in Shakespeare's *A Winter's Tale* not only completely ignored what he knew to be the indisputable evidence of the oracle of Apollo (knowing-but-not-knowing again); he also saw the protests of his servant Camillo at the unfoundedness of his jealousy as confirming evidence of its justification: the servant whom he used to trust completely must now be a 'false villain', working against him and on the side of his supposed rival. (In this respect, and in others too, the epistemology of sexual jealousy is similar to that of paranoia.) One alarming thing that this example brings out is the doubtful epistemic benefit of talking to a friend, or getting advice from someone wise, in order that what they tell you should serve as a corrective mechanism. For this is

13 I will not consider here Jon Elster's (1999a) fascinating discussion of the alchemies of emotion.

just what Leontes did, and yet he turned against Camillo and ignored the oracle of Apollo. Whilst the case of Leontes may be an extreme one, I think the phenomenon is quite widespread. We so easily find the well-intentioned corrective remarks of our friends as *parti pris* or as not really properly informed (Goldie 2004).

Xenophobia too skews the epistemic landscape. The xenophobe, who believes that much of his country's troubles are due to immigrants, is faced with reliable statistics showing that this is far from the case, and indeed that immigrants have done much good for the economy. But, with his xenophobia already in place, he ignores the reliable data, and instead latches on to the vividness of the stories in his *Daily Mail* of the latest crime by blacks against property or person. Such newspapers both thrive on, and exacerbate, our only partially recognized fears. We might suggest that the xenophobe buy a newspaper which we consider to be more reliable, but why should he? After all, so far as he can tell, it is only the *Daily Mail* that has the courage to say how bad things really are, and our telling him otherwise just shows how little we know.

So the first epistemic difficulty that we face in trying to correct for our misleading emotions is that our emotionally-grounded fast and frugal intuitive thinking can systematically skew the epistemic landscape, and that we can systematically fail to recognize this. There is no contrary 'check' from cool and calm deliberative thinking because reason has *already* been undermined by emotion without our conscious awareness: the pale nonemotional evidence is *already* discounted and the vivid emotional evidence is *already* given too much weight. As Jonathan Haidt puts it, 'the reasoning process is more like a lawyer defending a client than a judge or scientist seeking truth' (Haidt 2001, 820).

The second epistemic difficulty is different. Here we might well recognize the emotion as such, and even that we are being misled by it. But now what has happened is that our preferences are skewed in the heat of the moment – the emotion has *skewed the preferential landscape*. Even if we hear the voice of cool and calm deliberative thinking telling us to stop this silliness, we are still often motivationally overwhelmed by the passion. This might sound like the old idea of reason versus the passions, but this is not so. The idea is not that passion necessarily lacks rational authority altogether; the idea rather is that the motivating power of passion often exceeds its rational authority when we are in the grip of an emotion, and the emotion itself serves to mask this fact.

Even knowing that we are in the grip of an emotion, we can still often fail adequately to recognize the error of our ways, with all the focus of our mind on the salience of the object of our passion; only too late, after the passion is spent, do we realize our mistake. This kind of misleadingness is manifested in male aggressiveness. A nice example is from F. Scott Fitzgerald's *Tender is the Night*. Dick Diver, after a night of drinking, starts a fight with a taxi driver, is arrested and then, in the police station, just as he is about to get away with a small fine, he acts stupidly, with aggression. The consequences, as he would have foreseen without the blinding passion, are terrible:

> The captain stood up.
>
> 'Écoute!' he cried portentously. 'Vous êtes saoûl. Vous avez battu le chauffeur. Comme ci, comme ça.' He struck the air excitedly with right hand and left, 'C'est bon que je vous donne la liberté. Payez ce qu'il a dit – cento lire. Va au Quirinal.'
>
> Raging with humiliation, Dick stared back at him.
>
> 'All right.' He turned blindly to the door – before him, leering and nodding, was the man who had brought him to the police station. 'I'll go home,' he shouted, 'but first I'll fix this baby.'
>
> He walked past the staring carabinieri and up to the grinning face, hit it with a smashing left beside the jaw. The man dropped to the floor.
>
> For a moment he stood over him in savage triumph – but even as a first pang of doubt shot through him the world reeled; he was clubbed down, and fists and boots beat on him in a savage tattoo. He felt his nose break like a shingle and his eyes jerk as if they had snapped back on a rubber band into his head. A rib splintered under a stamping heel. Momentarily he lost consciousness, regained it as he was raised to a sitting position and his wrists jerked together with handcuffs. He struggled automatically. The plainclothes lieutenant whom he had knocked down, stood dabbing his jaw with a handkerchief and looking into it for blood; he came over to Dick, poised himself, drew back his arm and smashed him to the floor.[14]

As a practical corrective to these epistemic difficulties, it is not sufficient simply to think about, and perhaps to 'mouth', the error of our ways. What matters is the excessive motivating power of the preferences, and it is not enough to counteract this just to recognize that their power exceeds their authority. This is nicely evidenced by Aristotle's example of the emotionally weak-willed person: whilst indulging himself in food and drink, he recites the verses of Empedocles, which are all about the mistakes of excessive indulgence in food and drink. Again we are in the land of knowing-but-not-knowing.[15]

Another practical epistemic corrective to the skewing of the preferential landscape might be consciously to try to turn one's attention away from the salience of the object of temptation. But this has obvious and familiar difficulties: one has to focus on the very thing that one is set on ignoring, such as, in Dick Diver's case, the grinning, leering, nodding face of the man who had brought him to the police station.

A third epistemic corrective that one might deploy in deliberative thinking, which often has more promise of success, is not to think to oneself of the error of one's ways (the first corrective), nor to try to reduce the salience of the temptation (the second corrective), but to increase the salience of the paler information, and thus to increase its emotional import and motivating power. In effect, this involves recruiting

---

14 F. Scott Fitzgerald's *Tender is the Night*, chapter XXII of book 2; first published Charles Scribner's Sons 1934, now in various editions.

15 See Aristotle's *Nicomachean Ethics* Book VII, and for discussion Broadie (1991). Also see Lovibond (2002) for a more Platonic conception.

emotion, imagination and vividness on to the side of deliberative thinking. After all, there is no reason why these powerful battalions should only be deployable on the side of non-conscious intuitive thinking. Thus, for example, one might intentionally try hard vividly to imagine the bad consequences of one's action before doing it – for example imagining being seen in the act by the person you are betraying, or imagining hearing the shopkeeper's shout of 'Stop thief!' as you walk away from the scene of your petty crime. In this way the countervailing considerations become salient too, in spite of their distance, and this can then have the effect of balancing out the saliences somewhat; emotion is deployed on both sides, not only on the side of intuitive thinking. In Dick Diver's case, he might have tried vividly to imagine what would happen to him if he hit that man in that police station, and this would have given rise to fear at what he imagined, to set against the vivid fury and Protoshame that he felt at the humiliating remarks of the police captain. However, this of course required foresight and, as we have just seen, Dick realized only too late the implications of what he had done – the damage was already done – the emotions has already done their misleading work.

And this, of course, points to a fourth and very familiar corrective in which deliberative thinking can play a role: stop and think; count to ten; bite your tongue; take a deep breath; sleep on it. All these familiar everyday admonitions are implicit acknowledgements of the power of emotion, of quick and dirty intuitive thinking, to mislead us to over-hasty action.

There is nothing in principle against any of these correctives, and some or all of them surely can be effective on occasion. But there remains the central concern that emotion-based intuitive thinking does its dirty work, so to speak, before deliberative thinking comes on the scene, so that when deliberative thinking does arrive, the epistemic landscape and the preferential landscape have *already* been skewed and the dirty work *already* covered up. To adapt Haidt's metaphor, it is as if the defence lawyer, before he takes on the case, already believes that his obviously guilty client is innocent.

5. Conclusion

Over the last few years there has been this optimistic trend in emotion research, exemplified here in the work of Gigerenzer and his colleagues, and also more widely amongst philosophers of a 'cognitivist' bent, which emphasizes the usefulness of emotion in picking up saliences in the environment, and enabling quick and effective action with little or no conscious deliberation. This optimism, I believe, deserves to be tempered with some healthy realism. So too does optimism about the reliable power of deliberative thinking to correct any distortions that might arise through the influence of emotion. So when Peter Todd, writing about fast and frugal heuristics, says that 'humans are uniquely able to set aside such mental shortcuts and engage in extensive cogitation, calculation, and planning' (Todd 2001, 54), we might agree with him, so far as it goes. But I have been trying to make the realistic – or pessimistic – point that if Todd's remark, and others like it, is taken to suggest that 'setting aside' our heuristics to engage in cogitation is something we can *reliably and effectively*

do, then this optimism is open to doubt when the emotion-based heuristic has an environmental mismatch. Emotions such as these can systematically mislead us, and they can do so in ways that are systematically hard to detect, and systematically hard to correct for.

References

Barkow, Jerome H.; Cosmides, Leda; Tooby, John (eds). 1992. *The Adapted Mind. Evolution, Psychology, and the Generation of Culture*. Oxford: Oxford University Press.

Bermúdez, José-Luis. 2000. 'Rationality, Logic, and Fast and Frugal Heuristics. Review of Gigerenzer et al.' *Behavioral and Brain Sciences* 23: 744–5.

Broadie, Sarah. 1991. *Ethics with Aristotle*. Oxford: Oxford University Press.

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.

Elster, Jon. 1999a. *Alchemies of the Mind. Rationality and the Emotions*. Cambridge: Cambridge University Press.

Elster, Jon. 1999b. *Strong Feelings. Emotion, Addiction, and Human Behavior*. Cambridge, Mass.: MIT Press.

Fessler, Daniel. 2001. 'Emotions and Cost–Benefit Assessment. The Role of Shame and Self-Esteem in Risk Taking'. In: Gigerenzer, Gerd; Selten, Reinhard (eds). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, Mass.: MIT Press. 191–214.

Gigerenzer, Gerd. 2000. *Adaptive Thinking*. Oxford: Oxford University Press.

Gigerenzer, Gerd. 2001. 'The Adaptive Toolbox'. In: Gigerenzer, Gerd; Selten, Reinhard (eds). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, Mass.: MIT Press. 37–50.

Gigerenzer, Gerd; Selten, Reinhard. 2001. 'Rethinking Rationality'. In: Gigerenzer, Gerd; Selten, Reinhard (eds). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, Mass.: MIT Press. 1–12.

Goldie, Peter. 2000. *The Emotions. A Philosophical Exploration*. Oxford: Clarendon Press.

Goldie, Peter. 2004. 'Emotion, Reason, and Virtue'. In: Evans, Dylan; Cruse, Pierre (eds). *Emotion, Evolution, and Rationality*. Oxford: Oxford University Press. 249–67.

Goldie, Peter. 2007. 'Seeing What Is the Kind Thing to Do. Perception and Emotion in Morality'. *Dialectica* 61: 347–61.

Greene, Joshua D.; Sommerville, R. Brian; Nystrom, Leigh E.; Darley, John M.; Cohen, Jonathan D. 2001. 'An fMRI Investigation of Emotional Engagement in Moral Judgement'. *Science* 293: 2105–8.

Greenspan, Patricia S. 2006. 'Practical Reasoning and Emotion'. In: Mele, Alfred R.; Rawling, Piers (eds). *The Oxford Handbook of Rationality*. New York: Oxford University Press. 206–21.

Griffiths, Paul E. 1997. *What Emotions Really Are. The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Haidt, Jonathan. 2001. 'The Emotional Dog and Its Rational Tail. A Social Intuitionist Approach to Moral Judgement'. *Psychological Review* 108: 814–34.

Haidt, Jonathan. 2007. 'The New Synthesis in Moral Psychology'. *Science* 316: 998–1002.

Hopkins, Jim. 2004. 'Conscience and Conflict. Darwin, Freud, and the Origins of Human Aggression'. In: Evans, Dylan; Cruse, Pierre (eds). *Emotion, Evolution, and Rationality*. Oxford: Oxford University Press. 225–48.

Hume, David. [1739/40] 1978. *A Treatise of Human Nature*. Ed. Lewis A. Selby-Bigge and Peter H. Nidditch. Oxford: Oxford University Press.

Hume, David. [1777] 1975. *An Enquiry Concerning the Principles of Morals*. Ed. Lewis A. Selby-Bigge. Oxford: Oxford University Press.

Kahneman, Daniel. 2003. 'A Perspective on Judgement and Choice. Mapping Bounded Rationality'. *American Psychologist* 58: 697–720.

Kahneman, Daniel; Slovic, Paul; Tversky, Amos. 1982. *Judgment under Uncertainty. Heuristics and Biases*. Cambridge: Cambridge University Press.

LeDoux, Joseph E. 1998. *The Emotional Brain. The Mysterious Underpinnings of Emotional Life*. London: Weidenfeld and Nicolson.

Lovibond, Sarah. 2002. *Ethical Formation*. Cambridge, Mass.: Harvard University Press.

Morton, Adam. 2000. 'Heuristics all the Way up?' *Behavioral and Brain Sciences* 23: 758–9.

Nisbett, Richard E.; Ross, Lee D. 1980. *Human Inference. Strategies and Shortcomings of Social Judgment*. Englewood Cliffs: Prentice-Hall.

Power, Samantha. 2003. *A Problem from Hell. America and the Age of Genocide*. New York: Harper Perennial.

Roberts, Robert C. 2003. *Emotions. An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.

Samuels, Richard; Stich, Stephen. 2003. 'Rationality and Psychology'. In: Mele, Alfred R. (ed.). *The Oxford Handbook of Rationality*. New York: Oxford University Press. 279–300.

Samuels, Richard; Stich, Stephen; Bishop, Michael. 2002. 'Ending the Rationality Wars. How to Make Disputes about Human Rationality Disappear'. In: Elio, Renee (ed.). *Common Sense, Reasoning and Rationality*. New York: Oxford University Press. 236–68.

Slovic, Paul. 2007. '"If I Look at the Mass I Will Never Act". Psychic Numbing and Genocide'. *Judgment and Decision Making* 2: 79–95. http://journal.sjdm.org/jdm7303a.pdf.

Taylor, Gabrielle. 1998. 'Envy and Jealousy. Emotions and Vices'. *Midwest Studies in Philosophy* 13: 233–49.

Todd, Peter M. 2001. 'Fast and Frugal Heuristics for Environmentally Bounded Minds'. In: Gigerenzer, Gerd; Selten, Reinhard (eds). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, Mass.: MIT Press. 51–70.

Todd, Peter M.; Gigerenzer, Gerd. 2000. 'Précis of *Simple Heuristics That Make Us Smart*'. *Behavioral and Brain Sciences* 23: 727–41.

Wilson, Margo; Daley, Martin. 1992. 'The Man Who Mistook His Wife for a Chattel'. In: Barkow, Jerome H.; Cosmides, Leda; Tooby, John (eds). *The Adapted Mind.*

Evolution, Psychology, and the Generation of Culture. Oxford: Oxford University Press. 289–322.

Wollheim, Richard. 1999. *On the Emotions*. New Haven, Conn.: Yale University Press.

This page intentionally left blank

Chapter Eight

How Cognition Meets Emotion: Beliefs, Desires and Feelings as Neural Activity

Paul Thagard*

This chapter argues that beliefs, desires and emotions should be construed naturalistically in terms of brain mechanisms rather than as propositional attitudes. Mental states are patterns of neural activity, not relations between dubious entities such as selves and propositions. From this perspective, it becomes easy to see how cognition and emotion are intertwined, and hence how emotions can be integral to epistemology. Emotions are relevant to epistemology as frequent contributors to the growth of knowledge, as sometime impediments to knowledge acquisition, and as components of knowledge about persons and morality. I give a theoretically rich and empirically supported account of the neurophysiological interconnections of cognition and emotion, and discuss the epistemological significance of this naturalistic, materialist reconstrual of cognitions and emotions.

Deep appreciation of the relevance of emotion to epistemology requires a rich account of how emotional mental states such as happiness, sadness and desire interact with cognitive states such as belief and doubt. Analytic philosophy since Gottlob Frege and Bertrand Russell has assumed that such mental states are propositional attitudes, which are relations between a self and a proposition, an abstract entity constituting the meaning of a sentence. This chapter shows the explanatory defects of the doctrine of propositional attitudes, and proposes instead that beliefs, desires and emotions should be construed naturalistically using current understanding of brain mechanisms. Mental states are patterns of neural activity, not relations between dubious entities such as selves and propositions. From this perspective, it becomes easy to see how cognition and emotion are intertwined, and hence how emotions can be integral to epistemology.

I begin by reviewing some of the ways in which emotions are relevant to epistemology: as frequent contributors to the growth of knowledge, as sometime impediments to knowledge acquisition and as components of knowledge about persons and morality. I then argue that propositional attitudes do not exist, because the selves and the propositions that they purportedly relate do not exist. Thus the doctrine

* I am grateful to Ulvi Doğuoğlu and Benoit Hardy-Vallée for comments on an earlier draft. The Natural Sciences and Engineering Research Council of Canada provided research support.

of propositional attitudes is as useless for epistemology as it is for explaining human action. I argue for an alternative construal of mental states as patterns of neural activity, and describe how it is possible to give a theoretically rich and empirically supported account of the neurophysiological interconnections of cognition and emotion. Finally, I discuss the epistemological significance of this naturalistic, materialist reconstrual of cognitions and emotions.

1. Connections Between Epistemology and Emotions

There are many ways in which emotions and epistemology are mutually relevant, but I will merely review some connections that I have explored at much greater length elsewhere (Thagard 2006b). Traditional analytic epistemology has largely ignored the role of emotions in knowledge, but from a broader perspective they are clearly relevant. In particular, understanding the development of scientific knowledge requires noticing a positive contribution of emotional thinking. Emotions such as interest, curiosity, wonder and surprise are inextricable from the cognitive processes of scientific investigation, guiding researchers to generate important questions and to try to produce acceptable answers to them. The search for empirical and theoretical success can be accompanied by episodes of hope and happiness, but inevitable impediments can also lead to negative emotions such as worry, fear, frustration, anger and disappointment. Positive emotions provide the motivational fuel to conduct the difficult work that is crucial to any scientific investigation. Both positive and negative emotions provide signposts about current progress and directions to pursue. Important breakthroughs are marked by elation, whereas setbacks can prompt disappointment and even despair. Philosophical thinking is also often driven by emotions, ranging from wonder at the marvels of the universe to angst and despair about the human condition.

In addition to their crucial role in helping to direct investigation, emotions are also involved in the acceptance and rejection of beliefs. Belief-revision is a matter of explanatory coherence: you accept or reject beliefs on the basis of how well they fit with your full set of observations and explanations (Harman 1986; Thagard 2000). The process of assessing explanatory coherence operates unconsciously and in parallel, so we have no conscious access to it. All that comes to consciousness is a feeling that a belief seems to fit, which is part of a positive emotion that things make sense. On the other hand, if a belief does not fit, we get a negative feeling about it or about the overall state of our belief system. Coherence and acceptance feel good, whereas incoherence and doubt are irritating. Hence emotions are relevant to the development of knowledge with respect to acceptance as well as investigation and discovery. The emotional character of belief acceptance and rejection is especially evident in ethics, where utterance of an appealing principle, such as that it is moral to help the poor, generates positive emotion. In contrast, utterance of a dubious principle, such as that it is moral to torture suspected criminals, may be met with negative emotions, perhaps even anger or disgust.

The contributions of emotions to epistemology are often positive, as in their encouragement of investigation and coherence, but there are also ways in which

emotions can skew beliefs. Thagard (2007c) discusses the 'affective afflictions', which are systematic ways in which belief fixation is biased by emotions. These include wishful thinking and motivated inference, in which people's inferences are shaped by their personal goals rather than consideration of all the evidence. Another affective affliction is self-deception, which also involves people being unable to see why they are making the inferences that they do. Conflicts of interest in government and business often distort inferences through motivated inference and self-deception.

A third way in which emotions are relevant to epistemology concerns knowledge about the emotional states of ourselves and others. Such knowledge is crucial to the functioning of human organizations from families to governments, all of which depend on emotional intelligence (Goleman 1995). Some of this knowledge is constituted by explicit beliefs, for example when I infer that a friend is angry about something. But knowledge about the emotions of others can also be implicit, as when I use empathy to appreciate non-verbally my friend's emotional state by perceiving the physical manifestations of distress (Thagard, 2007b). An account of the interrelations of cognition and emotion should illuminate the nature of cognitions about emotions, as well as the contributions of emotions to scientific thinking and belief acceptance. I will now argue that traditional analytic epistemology based on propositional attitudes is unsuited for understanding the epistemological significance of emotions.

2. Against Propositional Attitudes

I will not attempt to review the vast philosophical literature on propositional attitudes (see e.g. Fitch 2005; Gale 1967; Iacona 2003; King 2006; Richard 1990; Salmon and Soames 1988). There are different theories about what propositions are, but the most common view is that they must be invoked as the meanings of sentences. When two sentences are synonymous, it is because they express the same proposition, which is an abstract entity independent of any uttered sentence. Mental states such as belief and desire are described by sentences with *that* clauses: I believe that today is Monday, and I desire that I will have a productive week. Hence my belief is a propositional attitude, a relation between me and the proposition that today is Monday. When we say that someone believes or desires that *P*, a proposition is the reference of the *that* clause. Emotions are also propositional attitudes, as when I am happy that I have written several pages today, but frustrated that I have not had time to organize my course for next term. My actions are explained by my propositional attitudes such as beliefs, desires and emotions. For example, when I believe that an action is the best way to satisfy a strong desire, I perform that action.

Part of the appeal of the propositional attitude doctrine is that it fits well with our ordinary way of talking: we do utter sentences about beliefs, desires and emotions that contain *that* clauses. But there is no reason to assume that this way of talking is scientifically or metaphysically justified. People commonly talk as if the sun rises, as if what goes up must come down and as if whatever happens is God's will. What evidence is there that mental states are propositional attitudes?

From a scientific perspective, there is no evidence and there cannot be, because abstract entities cannot contribute to explanations. My argument has the following structure:

(1) Theoretical entities are justified only by inference to the best explanation.
(2) Explanation requires a causal link between entities and what they explain.
(3) Abstract entities have no causal links with observable phenomena.
(4) Therefore, belief in abstract entities is not justified.

Each of these steps requires amplification and defence.

In science, we have good grounds for believing in the existence of theoretical entities such as atoms, electrons, forces, viruses and genes. Justification does not come from direct observation, the way we are justified in believing in the existence of dogs and tables, because theoretical entities are not directly observable. Fortunately, human cognition has the power and flexibility to go beyond our limited perceptual apparatus and find good reasons for postulating things that we cannot perceive. The form of this reasoning is *inference to the best explanation:* we infer that an entity exists because the hypothesis that it exists is part of the best explanation of what we know through observation and experiments (see Harman 1986; Lipton 2004; Thagard 1988; Thagard 1992). There is ample historical evidence that inference to the best explanation is ubiquitous in science and everyday life. This kind of inference is highly fallible, but still provides us with an often reliable way of going beyond the information given to our senses (Thagard, 2007a).

But what is explanation? If explanation is a logical relation, as maintained by the hypothetico-deductive model of Hempel (1965), then there might be hope for the doctrine of propositional attitudes. We might be able to generate deductive explanations something like the following:

(1) Whenever a person *P* has a desire that *D* and a belief *B* that an action *A* is the best way to accomplish *D*, then *P* does *A*.

(2) Paul desires to go to the baseball game and believes that buying a ticket is the best way to go.

So

(3) Paul buys a ticket.

Unfortunately, there are myriad reasons for doubting that deduction is necessary or sufficient for explanation (Salmon 1989). It is not necessary, because there are good explanations in biology and other fields, for example employing Darwin's theory of evolution by natural selection, that are not deductions. That deduction is not sufficient for explanation is shown by myriad examples where deduction is not explanatory, for example when one explains a man's non-pregnancy by deducing it from the facts that he took his wife's birth-control pills and that people taking birth-control pills do not get pregnant.

An alternative account, more consistent with scientific practice, is that explanations are representations of causal mechanisms (Salmon 1984; Bechtel and Abrahamsen 2005). A mechanism is a system of objects related to each other in various ways including part–whole and spatial contiguity, such that the properties of the parts and the relations between them produce regular changes in the system. For example, modern medicine explains diseases in terms of the interactions of parts of the body such as organs and cells with external influences such as bacteria. The high fever and other symptoms of influenza are explained by causal interactions among viruses and body cells including the immune system. These interactions cause the symptoms of the disease, not in some abstract deductive sense, but in the physical sense that they make the symptoms happen (Woodward 2004).

Now the problem with propositional attitudes becomes evident. Because propositions are abstract entities, they cannot be part of any physical mechanism, and so they cannot make anything happen. Hence hypotheses about propositions cannot be part of any explanation of observable facts, let alone part of the best explanation. Therefore, because inference to the best explanation is the only way to infer the existence of theoretical entities, there is no way that the existence of propositions can be justified. We must accordingly conclude that propositions, construed as abstract entities constituting the meaning of sentences, do not exist.

It follows immediately that propositional *attitudes* do not exist, because there are no propositions for a self to have a relation to. Moreover, the notion of a self is metaphysically dubious, redolent as it is of the Cartesian soul. I have argued elsewhere that the hypothesis of the existence of a non-material soul is not part of the best explanation of human behaviour (Thagard 2000, ch. 4). Perhaps there is some other more plausible conception of the self that could stand in some relation to abstract propositions, but I do not know what it would be. A better course is to develop a more empirically supportable conception of the self as part of a neurophysiological account of cognition and emotion (Metzinger 2003).

My argument would be undercut if there were some way to justify the existence of propositions besides inference to the best explanation, perhaps by direct acquaintance. A proponent of direct acquaintance would have to explain, however, how it is that people manage to have this kind of direct Platonic connection with abstract entities. Intuition should be psychologically explainable, not utterly mysterious. An alternative attempt to save propositional attitudes would be to have a much looser notion of explanation, perhaps some vague intuitive notion of making sense. But this would run counter to our best examples of successful explanations, those found in science. Even if it could be argued that propositions and propositional attitudes furnish some sort of explanation of our mental life and semantic capabilities, it would remain to be shown that they are part of the best explanation, and I will try to provide a better account below.

Another way of attempting to undercut my explanation-based argument against propositional attitudes would be to invoke the view that explanations are just answers to questions (van Fraassen 1980). It could then be argued that hypotheses about propositions and propositional attitudes answer many questions about meaning and mental states, so we should count them as explanatory. My first response to this argument is that for scientific and philosophical purposes it is crucial to spell

out what counts as an answer, and there is ample evidence that in contemporary biology and psychology the appropriate kinds of answers are provided by causal mechanisms (Bechtel and Abrahamsen 2005). My second response is that the invocation of van Fraassen's pragmatic approach to explanation should give scant support to proponents of the existence of propositions and propositional attitudes, because the point of that approach is to avoid making commitments to the existence of non-observable entities. From van Fraassen's perspective, the most one could say is that theories about propositions are empirically adequate. I will argue below that there are alternative theories of mental states that are much more successful at answering empirical questions.

My rejection of propositions and propositional attitudes does not involve denial of the existence of beliefs and desires. I have given reasons to doubt the propositional-attitude understanding of mental states, but do not deny that there are mental states that are something like beliefs and desires. The folk psychology of mental states may be approximately true, even if the philosophical understanding of folk psychology as involving propositional attitudes is false. I am far from the first to challenge the explanatory usefulness of construing propositions as abstract entities: see the nominalism of Ockham (Panaccio 2004), the behaviourism of Quine (1960) and the eliminative materialism of Churchland (1989). My main contribution here will be positive rather than negative, to describe a neurophysiological account that integrates cognition and emotion.

3. Beliefs as Neural Activity

In current cognitive psychology, the term *proposition* is used to refer, not to an abstract entity, but to a sentence-like mental representation (e.g. Anderson 1983; Kintsch 1998). Such representations are structured, consisting of simpler representations of objects and concepts. For example, the proposition that Jennifer loves Vince is a mental representation constructed from representations for Jennifer, Vince and the relation *loves.* In the early days of cognitive psychology, it was tempting to think of mental representations as akin to those in natural language, part of a language of thought. Only recently has it become evident how mental representations are at bottom neural.

It is still fashionable to deride the notion of a grandmother neuron, a nerve cell that fires just when you are thinking of our grandmother. But Quiroga et al. (2005) report the finding in human brains of individual neurons tuned to fire when a person sees pictures of well-known personalities such as Jennifer Aniston. Such neurons were found by recording the electrical activity of individual neurons in patients undergoing brain surgery for epilepsy. No one would claim, however, that any one neuron is *the* representation of Aniston, because other neurons must fire to correspond to aspects of her other than facial appearance, and even the particular neurons that fire when her face is presented occasionally fire when similar actors are presented.

Accordingly, neural representations are best thought of, not as individual neurons, but as groups of neurons, often called neural populations. The approximately 100 billion neurons in the brain are organized into large areas such as the prefrontal cortex,

which are in turn organized into populations of neurons with many interconnections with each other. Representations in neural populations are highly distributed, in that the firing patterns in a population may represent numerous objects and concepts, and each object and concept may be represented by many neurons in the neural population. As an approximation, we can describe each neuron as having a firing rate characterized by how frequently it fires given a stimulus. Taking this rate as a number, we can represent the rates of firing of a population of 100 neurons by a list of 100 numbers, which is called a 100-dimensional vector. For example, the vector (0.5, 0.8, 0.1, …) represents a population of neurons whose response to a stimulus involves firing 0.5 of the maximum rate, 0.8 of the maximum rate and so on.

However, the encoding power of neurons is not limited to rates of firing, but can also involve patterns of firing and not firing, also known as *spike codes.* Consider the following simple patterns of neural firing: (1) fire, rest, fire, rest; (2) fire, fire, rest, rest. Both patterns involve firing half the time, but they are clearly different in that (1) alternates firing and resting whereas (2) fires twice and then rests twice. There is neurological evidence that brains encode information using spike codes, which can also be shown to be more computationally powerful than only using rates of firing (Rieke et al. 1997; Maass and Bishop 1999). It is therefore plausible to conclude that neural populations represent objects by patterns of firing that employ spike codes. A neural population can represent an object when exposure to stimuli adjusts the excitatory and inhibitory connections between its neurons in such a way that the population is tuned to the object, in that its neurons exhibit a specific pattern of firing in response to the object (see Eliasmith and Anderson 2003). The same population of neurons may also represent a similar object by having a different pattern of firing in response to it.

Concepts that represent classes of objects can be encoded in the same way. A neural population is tuned to a class of objects if it has a pattern of firing that occurs when presented with one of the objects in the class. Hence the neural representation for a concept is structurally the same as the neural representation for an object, although the pattern of firing has to be flexible enough to respond to an indefinite number of similar objects. Of course, the neural pattern for an object has to be flexible too, since the object is not always presented in the same way. For example, the neural population for Jennifer Aniston has to be able to respond to different pictures as well as verbal and auditory inputs. Tuning of a neural population to represent a concept or object is not simply a matter of tying sensory experience to neural firing. If you have a neural representation of Jennifer Anniston, it should have a pattern of firing that occurs when you hear about her or imagine her, not just when you see a picture. Hence the pattern of firing in a neural population that constitutes the representation of an object or concept should be responsive to a variety of connected populations, not just ones that involve sensory stimuli. A neural theory of representation does not have to be behaviourist.

If objects and concepts have neural representations, what about beliefs such as that Jennifer loves Vince? How can the brain combine patterns of firing in neural populations for Jennifer, loves and Vince into something that represents the state of Jennifer loving Vince? Fodor and Pylyshyn (1988) criticized early connectionist (artificial neural network) models for their representational inadequacy: it was not

clear how a neural representation could capture the crucial difference between Jennifer loving Vince and Vince loving Jennifer. However, there are now a number of effective techniques for encoding complex relational information in neural networks, including the tensor products of Smolensky (1990) and the holographic reduced representations of Plate (2003); see also Eliasmith and Thagard (2001) and Smolensky and Legendre (2006). I will avoid the highly technical details, but describe the general approach.

Suppose that we have patterns of firing, not only for Jennifer, Vince and *loves*, but also for the agent role and recipient role of the relation *loves*. We can then generate a new pattern of firing that is tuned to the combination of the pattern for *loves* with two other combinations: the pattern for Jennifer integrated with the pattern for love-agent and the pattern for Vince integrated with the pattern for love-recipient. The resulting pattern of firing is then a representation that Jennifer loves Vince. It is still unknown exactly how the brain encodes objects and relations and combines them into neural patterns as complex as sentences, but tensor products and holographic reduced representations provide two examples of mathematical techniques that show how it is possible to build complex representations out of building blocks as simple as neurons. These techniques are adequate for capturing additional syntactic complexity, making possible the encoding of sentences such as 'Because Jennifer loves Vince, Jennifer struck the photographer' and even 'Jennifer believes that Vince loves her.'

The feasibility of construction of sentences out of neural activity legitimizes the claim that beliefs are neural activity: my belief that Jennifer loves Vince is a pattern of firing in my neural populations. I can still have this belief even if I am not at the moment thinking about them, because the synaptic connections between the neurons in the relevant populations are such that they will generate the relevant firing pattern when required, for example if I am asked whether Jennifer loves Vince. Thus occurrent beliefs are patterns of neural firing, but dispositional beliefs are structures consisting of neural connections that lead to the patterns of firing that constitute occurrent beliefs.

But what makes that pattern of firing the particular belief *that* Jennifer loves Vince? The answer is partly in the combinatorics by which the pattern of firing for this belief gets built out of the patterns of firing for the relevant objects and concepts. We need an account of how the neural population succeeds in representing Jennifer, which presupposes a theory of neurosemantics. Such theories have been offered by Eliasmith (2005) and Ryder (2004). I think that the core of the answer is that the neural activity that represents Jennifer has a complex kind of causal correlation with Jennifer herself, enabling the neural population to represent Jennifer. Defending this account of neurosemantics would distract this paper from its main goal, to show how cognition relates to emotion. Hence I now turn to an account of desires.

4. Desires as Neural Activity

To develop a neural account of desires, we can begin with the reward-based theory of desire proposed by Schroeder (2004, 131): 'To have an intrinsic (positive) desire that *P* is to use the capacity to perceptually or cognitively represent that *P*

to constitute *P* as a reward.' Schroeder's view is superior to previous philosophical views that tie desire to behavioural dispositions or the experience of pleasure, but it is exceedingly linguistic (Thagard 2006a). He only discusses desires *that* something be the case, which ignores human and animal examples where the object of desire is a thing, as when Jennifer desires a beer. Schroeder plausibly ties his theory to rapidly increasing understanding of the nature of reward in humans and motivation in humans and other animals, but restriction of desires to propositional contents is biologically implausible. It also has the difficulties associated with the notion of a proposition that I described in the above section on propositional attitudes.

Fortunately, there is an available alternative consistent with Schroeder's basic claims about the relation of desire to reward. Thagard (2006a, 153) rephrased Schroeder while avoiding the *that* clause: 'To have an intrinsic (positive) desire for Y is to use the capacity to perceptually or cognitively represent Y to constitute Y as a reward.' Y can still be a state of affairs if the desire is best described as a desire that something be the case, but it can also be a thing. Now we can put to work the account of neural representation that was sketched in the last section. Things and states of affairs that concern them are represented occurrently by patterns of neural activity, i.e. patterns of firing in neural populations, and dispositionally by neural structures (neurons and their synaptic connections) that generate those patterns of neural activities. Thus a desire is a pattern of neural activity that ties a representation of a thing or state of affairs with a representation of reward. Understanding reward requires a neurocomputational theory of reward processing that describes how brain areas such as the nucleus accumbens and the amygdala attach positive and negative evaluations to representations. I will now sketch such a theory as part of a neural theory of emotions.

5. Emotions as Neural Activity

The neural theories of belief and desire presented here complement the neural theory of emotions that I have presented in detail elsewhere and will here only sketch (see Thagard and Aubie, forthcoming). Philosophical and psychological theories of emotion have fallen into two camps, cognitive and somatic. Cognitive theories view emotions as appraisals, for example when you are happy that you are going to a good restaurant because it will help satisfy your goal of eating well. The view that emotions are cognitive judgements about a person's situation has been advocated by philosophers such as Nussbaum (2001) and by psychologists such as Oatley (1992), and Scherer, Schorr and Johnstone (2001). In contrast, somatic theories that understand emotions as perceptions of bodily states have been advocated by psychologists such as James (1884), by philosophers such as Prinz (2004) and by neuroscientists such as Damasio (1994). Proponents of cognitive and somatic theories have offered many competing arguments, but I will not review them here because I think that the theories are complementary.

How they complement each other is shown in the EMOCON model that I have proposed as part of a neurocomputational theory of emotional consciousness

(Thagard and Aubie, forthcoming). Figure 8.1 sketches interrelations among some of the brain areas that are most important for emotional experience. These include:

- *Thalamus*: processes sensory information for transmission to cortical areas.
- *Amygdala*: is involved in fear and other emotional evaluations.
- *Insula*: represents bodily states, especially ones linked to disgust and pain.
- *Dopamine system*: includes the nucleus accumbens, the ventral tegmental area and other circuits that process positive rewards.
- *Anterior cingulate*: is part of the cortex with functions such as error detection and modulation of emotional responses.
- *Dorsolateral prefrontal cortex*: is part of the cortex at the front/top/sides, involved in executive processes and working memory.
- *Orbitofrontal prefrontal cortex*:is part of the cortex at the lower front, involved in processing rewards.
- *Ventromedial prefrontal cortex*: part of the cortex at the bottom-middle, providing connections between the cortex and the amygdala.

This is by no means an exhaustive list of brain areas involved in emotion, but does include some of the main areas relevant to somatic and cognitive processing.

In Figure 8.1, the arrows represent some of the main causal influences among different brain areas, indicating that neural activity in one area causes neural activity in another. For example, suppose you are presented with the external stimulus of a car skidding towards you. This stimulus affects your external sensors – eyes and ears – which send information to your thalamus. Neural activity in the thalamus affects, by means of synaptic connections, neural activity in the prefrontal cortex, but the thalamus also has direct effects on the amygdala and on bodily states such as heart rate. Internal sensors transfer information about somatic states to the amygdala and insula, which also interact with the prefrontal cortex, which at the same time is conducting an appraisal of the significance of the stimulus, including assessment of possible rewards and threats.

Figure 8.1 should not be viewed as a serial flow chart, but rather as a causal map of how neural activity spreads in parallel through different brain areas. Your emotional response to the skidding car is the overall state of the brain that is produced by the parallel pathways of cognitive appraisal, involving the prefrontal cortex, anterior cingulate and dopamine system, and somatic perception, involving the amygdala and insula.

The EMOCON model shows how to integrate somatic and cognitive theories of emotion. The model includes somatic perception, with areas such as the amygdala and insula representing bodily states by receiving information from internal sensors. Again by representation I mean neural activity that is tuned in such a way that it causally correlates with something else, in this case bodily states such as blood pressure, heart rate, and blood levels of cortisol and glucose. However, human emotion is not just perception of bodily states, as it also involves appraisal performed by brain areas known to be involved in higher-level cognition, such as the dorsolateral prefrontal cortex. The involvement of cognitive appraisal in emotions is especially evident with the social emotions such as shame, guilt, embarrassment and

pride, all of which require a representation of how you are doing with respect to your place in a social group. But even basic emotions such as happiness involve appraisal of the degree to which one's goals are being accomplished. Thus the EMOCON model, based on well-known brain areas and connections, provides a way of seeing how emotions can be both cognitive appraisals and somatic perceptions.

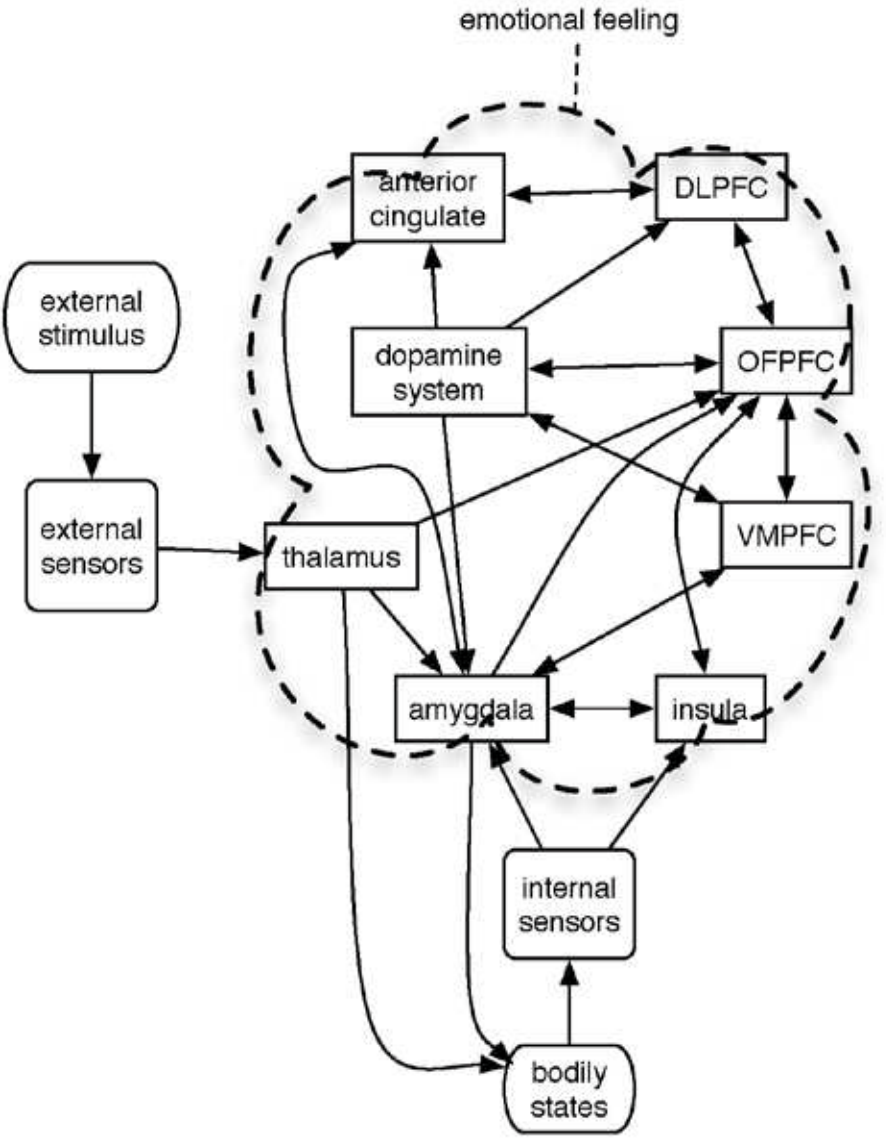

Figure 8.1 The EMOCON model of emotional consciousness, from Thagard and Aubie (forthcoming). Abbreviations are PFC for prefrontal cortex, DL for dorsolateral, OF for orbitofrontal, and VM for ventromedial.

I argue elsewhere that the EMOCON model can explain numerous aspects of emotion, including our ability to differentiate many emotions, our integration of cognition and emotion, and emotional change (Thagard and Aubie, forthcoming). My primary concern here is understanding the role of emotions in epistemology. Emotions are patterns of neural activity involving multiple brain areas that perform both cognitive appraisal and bodily perception. This activity depends on extensive feedback connections among the relevant brain areas. Emotions involve representations of bodily states and sensory inputs, and in sophisticated brains such as those possessed by humans they also involve representations of the world and of persons, including the person who has the emotion. This account of emotions as neural activity fits perfectly with my previous account of beliefs and desires as patterns of neural activity: the same kinds of representation underlie emotional states as underlie cognitive beliefs. Thus the common currency of beliefs, desires and emotions is neural activity, not the abstract propositions assumed in the traditional philosophical account. As for beliefs, emotions have both occurrent and dispositional forms, where the latter understands them as neural structures that generate neural activity when stimulated. For example, Jennifer's loving Vince may be either neural activity that occurs when she is thinking about him, or it may be neural structures (neurons and synaptic connections) that dispose her towards that neural activity even when she is not thinking about him.

As we saw for desires, the objects of emotional states can be representations of objects and concepts as well as states of affairs. I may fear *that* the skidding car will hit me, but it is just as psychologically plausible to say that I fear the car itself or that I fear dying. On the positive side, I can feel happy about Jennifer or about actors, as well as happy about the fact that Jennifer is an actor. My account of emotions as neural activity is compatible with associating them with representations of objects and concepts as well as states of affairs, which is another advantage over the narrow view that emotions are propositional attitudes.

6. Implications for Epistemology

The neural-activity view of mental states makes possible a deeper understanding of the positive and negative contributions of emotion to the development of knowledge. The accounts of beliefs, desires, emotions and consciousness sketched above make it clear that from a neural perspective there is no sharp distinction between cognition and emotion. Cognition involves a panoply of kinds of representations, not just sentence-like ones, and all of these can have associated emotions. For example, Fazio (2001) reviews evidence for the automatic activation of emotional attitudes attached to concepts. The emotional nature of concepts is most evident with extreme examples such as *cockroach* and *ice cream,* but other concepts can have more moderate emotional associations. Representations of objects and visual scenes are also often associated with emotional reactions. The brain has many interconnections, shown roughly in Figure 8.1, between areas of the brain associated with emotional processing and areas associated with cognitive processing. Thus from a neural perspective, cognition and emotion are highly interrelated.

Hence it is not surprising that even the most high-level cognitive processes, such as the discovery and evaluation of scientific hypotheses, are highly emotional. Ambitious scientists have intense goals, which are not abstract aims but rather desires associated with reward-related parts of the brain that also are involved in positive emotions. Beliefs associated with the prospect of accomplishing cognitive and personal goals naturally have positive emotions associated with them. Failure to achieve these goals activates negative emotions involving feelings of sadness, worry or even anger. Because mental representations generally have an affective component, and this component unavoidably influences how the representations are processed, emotions are highly relevant to epistemic progress.

From the perspective of traditional epistemology, considering the role of emotions in the development of knowledge might seem to conflate the descriptive with the normative. But naturalistic epistemology sees the need to base normative considerations in part on how human thinking actually works. Quine writes:

> Naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures. For me, normative epistemology is a branch of engineering. It is the technology of truth-seeking, or, in more cautiously epistemological terms, prediction. Like any technology, it makes free use of whatever scientific findings may suit its purpose. It draws upon mathematics in computing standard deviation and probable error and in scouting the gambler's fallacy. It draws upon experimental psychology in exposing perceptual illusions, and upon cognitive psychology in scouting wishful thinking. It draws upon neurology and physics, in a general way, in discounting testimony from occult or parapsychological sources. There is no question here of ultimate value, as in morals; it is a matter of efficacy for an ulterior end, truth or prediction. The normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed. (Quine 1986, 664–5)

Because cognitive psychology and neuroscience provide ample evidence of the role of emotions in cognition, naturalistic epistemology cannot afford to ignore how emotions affect the development of knowledge.

Even acceptance of hypotheses, not just their discovery, has an emotional component. This component would be avoidable if people were Bayesian engines, making decisions about acceptance and rejection based on mathematical probabilities. But we rarely have the information that would allow us to be Bayesians: in scientific, legal and other real life situations, the relevant conditional probabilities just are not known and we lack the resources to acquire them (Thagard 2000; Thagard 2004). Instead, we rely on a less rigorous but still effective method of accepting hypotheses if they cohere with the overall best explanation we can give. Simple neural networks can be used to represent both the maximization of coherence and the emotional inputs and outputs that can influence this process (Thagard 2000; Thagard 2006b).

Thagard and Aubie (forthcoming) have shown how coherence can be computed in more neurally realistic networks that better approximate to how brains work. We describe NECO (Neural Engineering Coherence), a model that uses populations of spiking neurons to represent the acceptability of hypotheses and evidence. This model is distributed in that acceptability of a hypothesis is represented by the activity of a whole group of neurons, and each neuron is involved in representing

the acceptability of multiple hypotheses. NECO also models part of the emotional side of hypothesis evaluation, in that it ties the neural representation of a hypothesis to activity in populations of neurons intended to correspond to human emotional areas such as the amygdala. Thus inference to the best explanation, implemented as a neural process of computing explanatory coherence, can be understood at the neural level as closely tied to emotional processing. Theory evaluation can have emotional inputs consisting of the positive and negative attitudes associated with the concepts and beliefs that constitute hypotheses. It can also have emotional outputs, including the appropriately good feeling we get when we achieve coherence and the annoyingly bad feeling we get when a proposed theory does not fit with our other beliefs.

There is also a negative side to the close connection in this model between cognitive and emotional coherence. When we feel good about how well our beliefs fit together, there is no way for us to tell whether in fact the coherence is really a matter of the goodness of fit of hypotheses with the evidence, or instead a matter of goodness of fit of hypotheses with our personal goals. Am I accepting a theory because it fits best with the evidence or because it is my own theory and thereby fits with my personal goal of becoming famous? The inability of individual scientists to tell introspectively whether their judgement is objective shows the need for naturalistic epistemology to be social. If I have the goal of achieving reliable knowledge not biased by my inevitable personal goals, then I ought to submit my ideas to peer review by other people who do not have the same personal goals. Only then can I be confident that my theory evaluation is not emotionally skewed by personal desires that are unavoidably as much a part of the neural assessment of explanatory coherence as are considerations of evidence.

Thus my neural account of mental states can explain the interactions of beliefs and emotions in situations where emotional contributions are negative as well as where they are positive. If we adopt Quine's engineering role for naturalistic epistemology, we have to work with the natural system that evolution has given us, which has emotions intertwined with cognitions. Just as civic engineers cannot build ideal bridges over ideal rivers, naturalistic epistemologists cannot prescribe ideal inference patterns for ideal representations such as propositions. Instead, we need to consider various cognitive and social means to amplify the positive epistemic effects of emotion, such as providing motivation and energy for scientific progress. We also need to develop cognitive and social means to diminish the negative epistemic effects, such as motivated inference. For both ameliorative projects, it is crucial to understand the underlying psychological terrain, which consists of many different kinds of representations all entangled with emotional processing.

The entanglement of cognition and emotion is even more important for establishing norms for practical reasoning, which includes both decision making and ethical judgement. There is rapidly expanding evidence that both ordinary and ethical decisions are made by processes that have a large and ineliminable emotional component. Hence moral epistemology, concerned with the justification of ethical claims, can also benefit from the deeper kind of understanding of cognitive-emotional interactions that neural theories can provide.

Even theoretical reason has a practical side, in that no intelligent agent aims only to accumulate a vast number of trivial truths. It would be pointless to accumulate a potentially infinite number of truths such as that our solar system has fewer than 12 planets, fewer than 13 planets, fewer than 14 planets and so on. Rather, scientists and people in general aim to accumulate *important* truths, ones that are relevant to their goals, which may include explanation as well as practical goals such as survival and success. A scientist, for example, may put enormous effort into pursuing a theory because of a combination of its potential for explanatory importance and promotion of the scientist's career. In animal brains, the value of actual and potential states of affairs is assessed by the emotional system, operating with the kinds of interactions shown in the EMOCON model. Hence emotions are crucial to epistemology in that they can help to guide us towards the acquisition of important truths.

7. Conclusion

I have tried to show how construing beliefs, desires and emotions as neural activity illuminates the relevance of emotions to epistemology. I argued that the traditional philosophical theory of propositional attitudes blocks rather than enhances understanding of mental states: the propositions that it invokes as abstract entities fail to explain interactions of beliefs and emotions. A traditionalist could respond that without propositions we have no way of talking about what is shared by two people who have the same beliefs. But the fact that my son Dan and I have the same height does not mean that there is an abstract thing – height – that we both share. Similarly, when he and I have the same belief that Waterloo is located in Ontario, it is not because we both have a relation to some dubious abstract entity, the proposition that Waterloo is in Ontario. Rather, the sameness of belief is the result of systematic similarities between our patterns of neural activity. These patterns should not be expected to be exactly the same, given that our life experiences have been different. But our experiences of Waterloo and Ontario have been sufficiently similar that we can estimate that there is enough in common between our neural representations of Waterloo and Ontario to attribute to us approximately the same belief. Height is approximate too, as Dan and I are not the same height if you worry about millimetres. Prescientifically, weight was thought of as a property of objects, but Newton reconceived weight as a relation between objects that gravitationally attract each other. Similarly, we need to reconceive meanings as relations that neural representations have to each other and the world, not as things.

Given the complexity of brains, there is no prospect for giving strict conditions for identifying a particular belief, desire or emotion in one person, let alone saying precisely what constitutes the same mental state in different persons. Current brain scanning technology only permits identification of the joint activity of many thousands of neurons, and there is no immediate prospect of being able to say exactly what neural structures constitute my belief that Waterloo is in Ontario. But the neural-activity view of mental states does not imply solipsism, because there is ample scientific evidence that people have very similar brains that undergo very similar causal interactions with the world. Further advances in experimental and

theoretical neuroscience can be expected to fill in many gaps in our understanding of the nature of neural representations, especially through computer models that simulate important kinds of inference.

Translation is also approximate. Philosophers have claimed that *Snow is white* and *La neige est blanche* express the same proposition and hence can constitute the same belief. From the neural perspective, the attribution of belief across languages is again approximate. We should not expect an exact correspondence between concepts of snow held by people in Canada and France, but just something close enough that we can attribute approximately the same belief to them. Similarly, the vaunted role of propositions as bearers of truth needs to give way to a looser relation between neural representations and the world. This relation will be more complex than the strict binary division of true and false, but such approximation is all that is needed for the epistemological purpose of marking a difference between belief and knowledge. Replacing strict notions of sameness of belief with the more approximate notions appropriate to neural activity also enables us to evade the notorious puzzles about propositional attitudes raised by Frege and Kripke. It becomes not all mysterious how someone can believe that Venus is the evening star but not believe that Venus is the morning star.

There are thus philosophical as well as scientific advantages to replacing the doctrine of propositional attitudes by a view of mental states as patterns of neural activity. Only recently has there been enough experimental evidence and theoretical progress to make plausible the claim that beliefs, desires and feelings are neural processes. A crucial part of this progress was the development of plausible ideas about how neural representations corresponding to beliefs can be built out of neural representations of objects and concepts. It now becomes possible to describe in neural terms the interactions of beliefs with desires and emotional feelings, and to use the resulting physical system to explain how the development of knowledge depends on emotional as well as cognitive processes. Because cognition meets emotion in the brain, emotions can be integral to epistemology.

References

Anderson, John R. 1983. *The Architecture of Cognition*. Cambridge, Mass.: Harvard University Press.

Bechtel, William; Abrahamsen, Adele A. 2005. 'Explanation. A Mechanistic Alternative'. *Studies in History and Philosophy of Biology and Biomedical Sciences* 36. 421–41.

Churchland, Paul M. 1989. *A Neurocomputational Perspective*. Cambridge, Mass.: MIT Press.

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

Eliasmith, Chris. 2005. 'Neurosemantics and Categories'. In: Cohen, Henri; Lefebvre, Claire (eds). *Handbook of Categorization in Cognitive Science*. Amsterdam: Elsevier. 1035–54.

Eliasmith, Chris; Anderson, Charles H. 2003. *Neural Engineering. Computation, Representation and Dynamics in Neurobiological Systems*. Cambridge, Mass.: MIT Press.

Eliasmith, Chris; Thagard, Paul. 2001. 'Integrating Structure and Meaning. A Distributed Model of Analogical Mapping'. *Cognitive Science* 25: 245–86.

Fazio, Russell H. 2001. 'On the Automatic Activation of Associated Evaluations. An Overview'. *Cognition and Emotion* 15: 115–41.

Fitch, Greg. 2005. 'Singular Propositions'. In: Zalta, Edward N. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/sum2005/entries/propositions-singular/.

Fodor, Jerry; Pylyshyn, Zenon. 1988. 'Connectionism and Cognitive Architecture. A Critical Analysis'. *Cognition* 28: 3–81.

Gale, Richard M. 1967. 'Propositions, Judgments, Sentences, and Statements'. In: Edwards, Paul. (ed.). *Encyclopedia of Philosophy*. New York: Macmillan. Vol. 6. 494–505.

Goleman, Daniel. 1995. *Emotional Intelligence*. New York: Bantam.

Harman, Gilbert. 1986. *Change in View. Principles of Reasoning*. Cambridge, Mass.: MIT Press/Bradford Books.

Hempel, Carl G. 1965. *Aspects of Scientific Explanation*. New York: Free Press.

Iacona, Andrea. 2003. 'Are There Propositions?' *Erkenntnis* 58: 325–51.

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

King, Jeffrey C. 2006. 'Structured Propositions'. In: Zalta, Edward N. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/fall2006/entries/propositions-structured/.

Kintsch, Walter. 1998. *Comprehension. A Paradigm for Cognition*. Cambridge: Cambridge University Press.

Lipton, Peter. 2004. *Inference to the Best Explanation*. London: Routledge.

Maass, Wolfgang; Bishop, Christopher M. (eds). 1999. *Pulsed Neural Networks*. Cambridge, Mass.: MIT Press.

Metzinger, Thomas. 2003. *Being No One. The Self-Model Theory of Subjectivity*. Cambridge, Mass.: MIT Press.

Nussbaum, Martha C. 2001. *Upheavals of Thought. The Intelligence of Emotions*. Cambridge: Cambridge University Press.

Oatley, Keith. 1992. *Best Laid Schemes. The Psychology of Emotions*. Cambridge: Cambridge University Press.

Panaccio, Claude. 2004. *Ockham on Concepts*. Aldershot: Ashgate.

Plate, Tony. 2003. *Holographic Reduced Representations*. Stanford: CSLI.

Prinz, Jesse J. 2004. *Gut Reactions. A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Quine, Willard V. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.

Quine, Willard V. 1986. 'Reply to Morton White'. In: Hahn, Lewis E.; Schilpp, Paul A. (eds). *The Philosophy of W.V. Quine*. La Salle: Open Court. 663–5.

Quiroga, Robert Q.; Reddy, Leila; Kreiman, Gabriel; Koch, Christof; Fried, Itzhak. 2005. 'Invariant Visual Representation by Single Neurons in the Human Brain'. *Nature* 435: 1102–7.

Richard, Mark. 1990. *Propositional Attitudes. An Essay on Thoughts and How We Ascribe Them*. Cambridge: Cambridge University Press.

Rieke, Fred; Warland, David; de Ruyter van Steveninick, Rob; Bialek, William. 1997. *Spikes. Exploring the Neural Code*. Cambridge, Mass.: MIT Press.

Ryder, Dan. 2004. 'SINBAD Neurosemantics. A Theory of Mental Representation'. *Mind & Language* 19: 211–40.

Salmon, Nathan; Soames, Scott (eds). 1988. *Propositions and Attitudes*. Oxford: Oxford University Press.

Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Salmon, Wesley C. 1989. 'Four Decades of Scientific Explanation'. In: Kitcher, Philip; Salmon, Wesley C. (eds). *Scientific Explanation*. Minneapolis: University of Minnesota Press. 3–219.

Scherer, Klaus R.; Schorr, Angela; Johnstone, Tom. 2001. *Appraisal Processes in Emotion*. New York: Oxford University Press.

Schroeder, Timothy. 2004. *Three Faces of Desire*. Oxford: Oxford University Press.

Smolensky, Paul. 1990. 'Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems'. *Artificial Intelligence* 46: 159–217.

Smolensky, Paul; Legendre, Géraldine. 2006. *The Harmonic Mind*. Cambridge, Mass.: MIT Press.

Thagard, Paul. 1988. *Computational Philosophy of Science*. Cambridge, Mass.: MIT Press/Bradford Books.

Thagard, Paul. 1992. *Conceptual Revolutions*. Princeton: Princeton University Press.

Thagard, Paul. 2000. *Coherence in Thought and Action*. Cambridge, Mass.: MIT Press.

Thagard, Paul. 2004. 'Causal Inference in Legal Decision Making. Explanatory Coherence vs. Bayesian Networks'. *Applied Artificial Intelligence* 18: 231–49.

Thagard, Paul. 2006a. 'Desires Are Not Propositional Attitudes'. *Dialogue: Canadian Philosophical Review* 45: 151–6.

Thagard, Paul. 2006b. *Hot Thought. Mechanisms and Applications of Emotional Cognition*. Cambridge, Mass.: MIT Press.

Thagard, Paul. 2007a. 'Coherence, Truth, and the Development of Scientific Knowledge'. *Philosophy of Science* 74, 28–47.

Thagard, Paul. 2007b. 'I Feel Your Pain. Mirror Neurons, Empathy, and Moral Motivation'. *Journal of Cognitive Science* 8, 109–36.

Thagard, Paul. 2007c. 'The Moral Psychology of Conflicts of Interest. Insights from Affective Neuroscience'. *Journal of Applied Philosophy* 24, 367–80.

Thagard, Paul; Aubie, Brandon. Forthcoming. 'Emotional Consciousness. A Neural Model of How Cognitive Appraisal and Somatic Perception Interact to Produce Qualitative Experience'. *Consciousness and Cognition*.

van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Clarendon Press.

Woodward, James. 2004. *Making Things Happen. A Theory of Causal Explanation*. Oxford: Oxford University Press.

Chapter Nine

Epistemic Feelings

Ronald de Sousa*

Somewhere along the course of evolution, and at some time in any one of us on the way from zygote to adult, some forms of detection became beliefs, and some tropisms turned into deliberate desires. Two transitions are involved: from functional responses to intentional ones, and from non-conscious processes to conscious ones that presuppose language and are powered by neocortical resources. Unconscious and functional mental processes remain and constitute an 'intuitive' system that collaborates uneasily with the conscious intentionality of the 'analytic' system. Emotions bridge these divides: in particular, specific feelings affect inference, cognition and metacognition. In what follows, after a brief reminder of the crucial role of emotions to rational thought and action in general, I first look at how fear affects belief. I then narrow my focus to some examples of what I shall refer to as epistemic feelings. These include specialized variants of fear and greed; and feelings of doubt, certainty, knowing and familiarity. I shall also describe some surprising recent findings about the influence of oxytocin on trust and about the direct influence of social conformity on perception and belief.

1. Emotions and Rational Deliberation

It is no longer controversial that emotions play a crucial role in the process of rational deliberation. Many of the ways that emotions condition our reasoning both about what to think and about what to do follow, as I have argued elsewhere, from the hypothesis that 'emotions are species of determinate patterns of salience among objects of attention, lines of inquiry, and inferential strategies' (de Sousa 1987, 196). Catherine Elgin's chapter in the present volume presents a number of vivid illustrations of that thesis in its more positive aspects. Less desirable consequences also follow, for the reasons lucidly elaborated, also in this volume, by Peter Goldie. A further consequence of the emotions' grip on our patterns of attention is that they help to tame the combinatorial explosion theoretically confronting the perfectly rational deliberating agent. As far as pure reason is concerned, every decision problem affords a virtually infinite set of plausible aims and an unlimited number of possible means. How is one to encompass that virtually infinite space and select among possible ends and means? This problem is related to the so-called 'updating' or 'frame problem' of artificial intelligence (Pylyshyn 1987). What has become recognized as a specifically philosophical version of the problem is that a purely rational being

* This paper has benefited as a result of many judicious comments by the editors, for which I am most grateful.

would have to deliberate about when to stop deliberating, and decide what to ignore in the process of making a decision. That would threaten an intractable regress of deliberation at every turn. Curiously, however, we are not generally plagued by that particular source of indecision in practice. Emotional constraints on salience and attention come close to simulating, for the duration of any emotional episode, the informational encapsulation of our sense organs; the result is that we generally do not spend any time making the second-order decision as to what to include in our decision procedures.[1]

I shall say no more here about the general topic of the influence of emotions on our practical and intellectual agendas, as it derives from the control of salience. But ordinary emotions sometimes affect cognition in different and more direct ways. In the next section, I consider certain peculiarities of the way our beliefs are affected by the emotion of fear. It is the first item in a cabinet of curiosities of which the present paper is intended to offer a brief guided tour. All exhibits featured in this cabinet involve 'epistemic feelings', of which central examples are feelings of knowing, of doubt, of certainty and of familiarity. Some more familiar emotions such as fear, greed or trust, or varieties thereof, function like epistemic feelings in affecting conviction, inference or cognitive strategies more or less directly. Where they function in this way, these emotions too will count as epistemic feelings. Feelings, in the sense intended here, differ from full-fledged emotions in two ways. Unlike emotions, they can be attributed at a subpersonal level, whereas emotions are typically attributed only to persons. A closely related point is that full-fledged emotions are more complex than the sort of feelings I shall be concerned with. Nevertheless, the latter share four points of resemblance with more robust emotions. First, they resemble standard cases of calm emotions in involving *evaluative appraisals*. Those appraisals bear on restricted domains of epistemic values, specifically on the quality of one's knowledge, on the extent of what one has learnt, and how much confidence can be placed in what one believes. But other emotions also bear on specific domains: fear evaluates risk, indignation assesses social or moral conduct and so forth. Second, the feelings in question also resemble full-fledged emotions in respect of the 'Janus-faced' character of the information they provide. They tell us something both about the subject and about something to which the subject is responding. Third, many characterizations of emotion insist on their link to agency and 'action tendencies' (Frijda 2004); epistemic feelings, as we shall see, also play a role in the guidance of (intellectual) activity. The feeling of knowing, for example, acts to stimulate or discourage us from further racking one's brains in attempted recovery. And fourth, both feelings and emotions have a characteristic phenomenology, but are sometimes legitimately attributed despite the lack of any identifiable conscious awareness.

1 This thesis has been criticized by Dylan Evans, who has labelled it the 'search hypothesis of emotions', on the grounds that neither the concept of emotions nor the statement of the search hypothesis can be delineated with sufficient precision for the hypothesis to be anything but vacuous (Evans 2004). In the context of the hypothesis about salience just recalled in the text, however, it seems plain that emotions, by narrowing the focus of attention, drastically reduce the search space for any given problem.

2. Risk, Fear and the Philebus Principle

Besides steamrolling attention and motivating self-deception, fear can play a rather different role. This is what is illustrated in the first item in my cabinet of curiosities. In some cases fear acts not merely to change the subjective probabilities of certain prospects, as seems required by its biological function of signalling danger. It acts in other ways as well, with some perverse results.

It seems plausible to speculate that the function of fear is not only to alert one to danger as a red light might do, but to track the extent of the risk envisaged by means either of its intensity or of its influence on the agent's preference rankings. Fear could then be described as an instinctive *measure of risk*. As we all know but seldom remember, however, it is not a realistic one. Terrorist attacks in the US since January 2001 have taken the lives of some 3500 persons, mostly on 11 September 2001. In the same six-year period, road accidents have killed about seventy-five times as many, as have guns; and iatrogenic diseases or medical errors have killed more than the last two together.[2] Conservatively, then, Americans are about three hundred times more likely to die of gun wounds, traffic accident or medical error as from a terrorist attack. In theory, many people know this; yet no politician could get elected by advocating billions of dollars of deficit spending for a 'war' against those far greater dangers. Terrorists ably bank on those discrepancies, achieving huge effects at remarkably little cost. This seems difficult to explain on the basis of the usual economic assumptions about the rationality of normal agents.

One clue to understanding it is that while fear may be a response to perceived risk, it has its own intrinsic experiential quality, including, for most people not currently sitting in a movie theatre, a strong negative valence. So to the classical Bayesian calculation of the expected desirability of a given prospect, which takes account of the probabilities and desirabilities of negative and positive outcomes, we should add the intrinsic desirability of the emotion that results from the contemplation of the prospective event in question. Since that emotion is itself a response to the expected desirability of that same event, the calculation has a self-referential character somewhat reminiscent of Hofstadter's Law: *It always takes longer than you expect, even when you take into account Hofstadter's law* (Hofstadter 1980, 152). Though fear is a measure of the risk contemplated, it is also a disvalue that must be added to the prospect of that risk. And unlike the negative outcome envisaged – say, a terrorist attack – it represents an actual disvalue, certain to be endured.

Consider another familiar example. In strictly Bayesian terms, one might represent the expected desirability of flying as opposed to driving to some destination as involving cost, time, and the undesirability and level of risk of an accident leading to delay or death. On the basis of those factors and reasonable estimates of their probabilities, flying is definitely the better option.[3] Adding in the factor of fear,

2 A pro-gun physician wishing to place gun ownership on a par with medical succour estimates the number of deaths due to medical error at 100,000 a year (Fackler 2000).

3 About ten times safer per mile driven or flown overall, and five hundred times safer than driving if you fly only in large commercial airliners. See http://philip.greenspun.com/flying/safety.

however, may significantly lower the expected desirability of flying in relation to driving, providing fear of flying is sufficiently higher than fear of driving.

This may provide a partial explanation of the fact that decisions taken in relation to certain prospects, such as flying or travelling to places threatened with acts of terrorism, are disproportionate to the dangers involved. But it does not follow that such responses are rational. Unless one is in the grip of a postulate holding that all general tendencies to thought, feeling and action bequeathed to us by evolution are infallible, one will recognize these cases as paradigms of emotion-induced irrationality.

We can be more specific about the type of irrationality involved. It consists, I suggest, in a violation of what I have proposed to call the *Philebus Principle* (de Sousa 2000). The name of this principle derives from the controversial suggestion made by Plato in his dialogue of that name that pleasures can be true or false (Plato 1997b). Plato argues that this holds not derivatively, in virtue of being based on false propositions, but literally insofar as pleasures can themselves *represent* other pleasures correctly or incorrectly. In the context of the present argument, the relevant example concerns pleasures of anticipation. These, Plato claims, should be proportionate in their valence and intensity to the pleasures that the anticipated events will yield. Here's how the principle might be formulated:

(PP) A pleasure taken in anticipation of X should be proportional to the pleasure that will be actually yielded when experiencing X.

Of course, it is always possible that one might be disappointed by the reality of a seemingly bright prospect. But Plato's principle does seem to capture a legitimate principle of rationality, insofar as pleasures of anticipation act also to motivate our responses. A merely random connection between – putting it now more generally – emotions of anticipation and the anticipated emotions would be seriously disruptive of all rational planning. Still, one might ask, if the Philebus principle is a principle of rationality, what *sort* of rationality is it? The rationale just given for it is practical, yet from the point of view of natural selection, one can see why it would sometimes pay to violate it. It might be simply inefficient always to calibrate levels of fear exactly with apprehended risk. When a risk is very low, it is better not to clutter one's emotional landscape with any response at all. We do not bother to consider the risk of a tsunami on every ocean beach, though it is presumably not zero. Conversely, some people just do not gamble: gambling and not gambling, if the bet is fair, are by definition equivalent in their expected desirability. It is therefore plausible to attribute the decision to bet or not to bet to an emotional attitude that belongs to a different 'channel' than the bare calculation of expected utility. An example might be a common decision procedure that takes account only of thresholds and ignores gradations between them. Emotional attitudes may determine the thresholds, but only in some regions will these attitudes be felt as distinct degrees of emotion. Thus when a danger rises above a certain threshold, we sometimes decline to 'take the risk' at all rather than make any attempt to compute it accurately. Only in the middle zone do we experience fear, but even there, the fear is not finely calibrated. We might, however, make a judgement on where to locate the threshold: is the risk worth

assessing? Is fear an appropriate emotion in the circumstances? Such judgements will be based not on fear but rather on a 'feeling of rightness' of which more in a moment.

An additional anomaly in the way that fear bears on the evaluation of future prospects is brought out by two economists in a study of responses to terrorism. Gary Becker and Yona Rubinstein point out that:

> [An] exogenous shock to the underlying probabilities affects agents' choices via two different channels: (i) the *risk* channel: a change in the underlying probabilities keeping (marginal) utility in each state constant; (ii) the *fear* channel: a change in the underlying probabilities *also* determines agents' optimal choice by affecting the expected utility from consumption in each state. (Becker and Rubinstein 2004, 4)

In their scheme, then, emotion has a distorting effect by virtue of its double life, directly affecting both parameters – desirability and subjective probability – in the Bayesian formula. As a measure of risk, fear should affect only probability: its causal influence on value or desirability seems to constitute a kind of leakage that cannot be rationally defended. We shall encounter below (in section 5) a converse case in which a feeling which seems relevant only to desirability appears to have a direct causal influence on subjective probability.

3. Epistemic Feelings

Historical Antecedents

I now focus more narrowly on some feelings that enter into the epistemic processes of inquiry, knowledge and metacognition. Some of these feelings have attracted attention only quite recently. Nevertheless they can be introduced by recalling three particularly historic moments in the history of philosophy.

The first two moments both occur in Plato's *Meno*, which contain two significant allusions to specifically epistemic emotions. First, Meno experiences *despair* or *dejection* brought on by the apparent impossibility of learning (Plato 1997a). We cannot learn anything new, it seems, because unless we already know the object of our quest we will not recognize it when we see it. Socrates urges him to view his dejection as a good thing. Awareness of our ignorance is the beginning of wisdom, providing it is used to spur inquiry. Second, when Meno's slave boy recognizes the correctness of Socrates' demonstration of the way to double the square, Plato diagnoses the slave boy's response as arising from his recollection. The slave boy presumably experiences a *feeling of familiarity*, characteristic of recollection as opposed to novel encounters or his own clueless previous guesses.

The third historic historical moment is that of Descartes' *cogito*. When Descartes noticed that 'I exist' must be true at least whenever I'm thinking about it, he wondered by what trick he had achieved the peculiar *certainty* of that conclusion (Descartes [1641]). His answer was that the thought of his existence was clear and distinct, and he inferred that the truth of anything else was as good as guaranteed if it too was

clear and distinct. As far as I'm aware, this is the first explicit use of something like a criterial *feeling of rightness* to justify a knowledge claim.

Unfortunately, Descartes was mistaken in thinking that this feeling of rightness, which he calls 'clarity and distinctness', is a sufficient criterion of truth. Or perhaps he merely failed to notice that the sufficiency of that feeling as a criterion cannot itself claim clarity and distinctness. Rather, as Hookway remarks, the mysterious 'immediacy' of the conviction that I exist is merely a reflection of 'the paucity of our epistemic self-knowledge' (Hookway, this vol., p. 55).

The Two-Track Mind

In truth, the deficiency of epistemic knowledge is not exceptional among forms of self-knowledge. Cognitive science has explored innumerable ways in which we remain 'strangers to ourselves' (Wilson 2002). We know amazingly little about the motives for which we do what we do or believe what we believe, and much of what we think we know we have confabulated (Hirstein 2005). From an evolutionary point of view, there are perfectly good 'design' reasons for this: for the 'workspace' of consciousness (Baars 1997) is very narrow and would cause a distressing bottleneck if the processing of our thoughts and intentions all had to pass through it. This creates something of a dilemma, for if the epistemic machinations of the unconscious go on unmonitored, they may well result, from the point of view of conscious planning, in time wasted in irrelevancies. On the other hand, requiring that every proposition that figures as a premise in an argument itself be grounded in a sound argument is a recipe for vicious regress.

If this no longer strikes us as a paradox, it is because it is now something of a commonplace that mental life involves two relatively distinct levels of mental processing (Strack and Deutsch 2004). Though details vary among some two dozen different versions of the view (Stanovich 2004, 34–6), most agree on a number of features of the two systems: the first or 'primary' system is characterized by 'automaticity, modularity and heuristic processing'. I'll refer to it as the '*intuitive*' system, ignoring the multiple opportunities for misreadings afforded by that term. The second, '*analytic*' system, tends to be 'rule-based, often language based, computationally expensive' (Stanovich 2004, 36). Typically, though not invariably, the processes and sometimes the behavioural manifestations of the intuitive system occur without being monitored in consciousness, whereas (again, typically rather than invariably) the analytic system seems to operate largely in the full light of consciousness.

The place of the emotions in this classification is ambiguous. On the one hand, emotions are traditionally held to be 'primitive', hasty, relatively automatic and often impervious to reason, all of which fits the label 'intuitive' rather comfortably. Recent brain science concurs, insofar as the regions of the brain associated with what are often referred to as basic emotions[4] – the limbic system, the amygdala – are not part of the neo-cortex, where most articulate cognitive activity appears to

4 A minimal consensus on what counts as a 'basic' emotion is that it should be like Paul Ekman's 'affect programs', grounded in specific brain circuits and processes; and cross-

be generated. Against this, an accumulating number of considerations support the view that emotions lie at the heart of the cognitive itself. Most easily available to common sense is the dependence of many individual emotions on specific beliefs. That suggests that emotions, like beliefs, are typically answerable to the analytic system. Second, emotions are (again, typically though not invariably) conscious, and consciousness is generally associated with the analytic system. Furthermore, some emotions, such as *interest* or the more elusive 'feeling of knowing', appear to be involved in knowledge and inference.

Both Christopher Hookway (1998) and Asher Koriat (2000) have suggested that there are a number of specifically epistemic feelings that have not until recently attracted sufficient attention from epistemologists. These epistemic feelings bridge the mind's two tracks, as we find certain deliverances of consciousness popping up ready-made, as it were, to take up positions as unargued yet unquestioned premises in explicit arguments. Hookway rightly comments that 'the fact that something serves as a first premise for conscious reasoning is compatible with its resulting from a process of unconscious reasoning (or other processes)' (Hookway, this vol., p. 53).

Some Varieties of Epistemic Feelings

Before proceeding further, it will be useful to sketch some varieties of epistemic feelings. They can be characterized according to whether or not they bear on a specific propositional object and by the phase at which they impinge on the pursuit of knowledge. Rather than an exhaustive classification, this is rather intended as a suggestion about what to attend to when looking at the way an epistemic feeling affects cognition.

1. Feelings such as *wonder* or *curiosity* motivate inquiry, but they need not presuppose specific suppositions or existing beliefs. Though I can wonder about whether a certain specific proposition is true, I can also be curious about a range of questions or topics without having formulated any clear yes-or-no questions.
2. Feelings such as *doubt* also motivate inquiry but they bear on hypotheses already entertained, propositions that have already made some claim on our assent.
3. In direct antagonism to feelings of doubt, there are feelings of *certainty* or *rightness*, which can also bear on specific beliefs or on the validity of *inferences*. The feeling of certainty freezes inquiry, to the extent that we may now feel we have the answer we were looking for. But it frees us for action based on the factual or normative propositions to which we now assent. Certainty about one proposition obviously does not preclude further inquiry into other questions. Close to the feeling of certainty which focuses on propositions, we shall encounter a distinct feeling of 'trust' targeting *persons*

culturally identifiable. For discussion see (Ekman and Davidson 1994), and particularly the contributions in that volume by Ekman, LeDoux and Panksepp.

directly and only indirectly bearing on propositions. Strictly speaking, trust is not an epistemic feeling, but because of the importance of testimony in the formation of most of our beliefs, it seems worth including in my purview, and I will shortly illustrate how trust, like many full-fledged emotions, can be curiously susceptible to direct chemical influence.

4. The *feeling of knowing* seems to bear on specific propositions, but without being able to specify what these are. It is a metacognitive feeling that induces us to believe that we know something before we are able to retrieve what we know. Familiar variants are the 'tip-of-the-tongue' phenomenon or the conviction that one has learnt something.

In what follows, I say a little more about some items in each of these four classes, particularly the last two. Along the way, I shall exhibit more items in my cabinet of curiosities. These are offered to illustrate the role of feelings as bridging the intuitive and analytic systems, and also to convey the sense that in some cases we seem to see into the very point where physiological process and subjective feeling coincide.

Aristotle opened his book on Primary Philosophy by remarking that 'all men by nature desire to know'. And Descartes ([1649]) listed the emotion of *wonder* at the top of his list of the passions. Its tamer relative, *interest* (Silvia 2006), also belongs to the first class. The feelings in this category get us started: they motivate inquiry, operating at a stage when we are in want of a hypothesis or explanation. But these emotions are not the only ones, among standard emotions, that can affect some of our deliberate or non-deliberate cognitive strategies.

Cognitive Foraging

While my next exhibit introduces epistemic feelings proper, it also extends the range of what one might have thought of under that heading. The reason is that it focuses on what we can think of as specifically epistemic species of two standard emotions, namely *greed* and *fear*, which we do not immediately associate with inference. It also illustrates the surprisingly direct influence of emotional chemistry on inference.

Angela Yu and Peter Dayan looked at the application to science of the trade-off between exploration and exploitation (Yu and Dayan 2005; Cohen and Aston-Jones 2005). Exploitation of known resources is safe but likely to yield diminishing returns: one can see it as expressing a form of fear of risk or preference for a sure thing. On the other hand, giving up well-trodden paths for the sake of exploration may yield a jackpot of discovery, but it is inherently risky: it can be seen as a variant of greed, a willingness to take risks for the sake of large gain.

That trade-off is well-known to students of ant foraging (Bonabeau, Dorigo and Théraulaz 1999). An ant faced with an established path may either follow it, in the expectation of finding food where many others have already found it, or else strike out in an unexplored direction. The latter option is risky but will pay off, if not for the individual at least for the colony, when the original sources of food are exhausted. This is a good example, then, of a mechanism first instantiated at the most basic level of animal foraging. What is surprising is that it seems to apply equally well to sophisticated scientific cognitive strategies, where it appears still to be controlled

validity of inference. That, as we saw, was something like the feeling aroused in Descartes by the contemplation of the *cogito*.

Doubt and certainty are both equally indispensable to rational inquiry. Once seized of any random belief, I would have no motive to undertake any further inquiry if I were incapable of experiencing doubt. Freedom from doubt is lauded as Faith, one of the three Theological Virtues, by those who have an interest in promoting mental docility in their followers; but it is notoriously unfriendly to scientific progress. On the other hand, if I did not have the feeling of certainty about a valid inference, I would not rely on it; and if I lack the feeling of conviction about a conclusion, I would remain, like Hamlet, unable to take any action based on it (Hookway 1998). Insofar as any valid argument is intended to persuade, it relies for its impact on the listener's feeling that its conclusion is not more implausible than its premises. If that is not the case, the listener is justified in turning the argument on its head to constitute a *reductio* of that among its premises that least arouses a feeling of certainty.

It took a couple of millennia before anyone had a convincing refutation of Zeno's argument from the possibility of infinite dichotomy against the reality of space and motion. Nevertheless, though it might not have been clear just which of its premises was false and why, it was plain enough to most people that the conjunction of its premises was no more certain than the denial of the argument's conclusion. So it was not because everyone was irrational that scarcely anyone doubted that space and motion were possible. Unless we spontaneously recognized the validity of some basic pattern of inference, such as *modus ponens* or *modus tollens*,[5] no instruction manual could save our inferences from sinking into logical quicksand. That is the lesson of Lewis Carroll's story of Achilles and the tortoise: if we required every applicable rule of inference to be written down as a premise, the simplest inference would require us to endorse an infinite number of finite steps, as we would need another rule to tell us that the present inference was an instance of the rule last specified, and so on for ever (Carroll 1895). Whatever our level of logical sophistication, there are inferences of which we feel the validity, others which we immediately feel are invalid. At the level of conscious policy, therefore, the dialectic of doubt and certainty is a fruitful one. But that dialectic is manifested below the level of conscious deliberation: epistemic feelings seem to serve precisely the function of providing premises elaborated at the subpersonal or intuitive level for use in explicit inferences.

The Chemistry of Trust

A near relative of feelings of rightness and certainty is the *feeling of trust*. It differs from rightness in that its object is neither a specific proposition nor the validity of inference, but a *person*. In a recent article in *Nature* which got almost as widely reported in the press as the research that indicated a genetic cause for

5 These two patterns are not necessarily on a par as a matter of psychological fact. The point made here, like the point made in the *Meno*, holds a priori: it is that unless some patterns of transition are built into the architecture of the brain, no process of reasoning can get going. What those patterns are is an empirical question: in theory – and sometimes in practice – they might include a disposition to *believe whatever you are told.*

women's differential capacity to experience orgasm, researchers at the University of Zürich have shown that 'intranasal administration of oxytocin, a neuropeptide that plays a key role in social attachment and affiliation in non-human mammals, causes a substantial increase in trust among humans' (Kosfeld et al. 2005, 673). Their results also support the conclusion that 'the effect of oxytocin on trust is not due to a general increase in the readiness to bear risks. On the contrary, oxytocin specifically affects an individual's willingness to accept social risks arising through interpersonal interactions' (Kosfeld et al. 2005, 673). The experimental set-up in each run of the experiment involved two subjects, an 'investor' and a 'trustee'. Both received 12 monetary units and each unit invested by the investor was tripled by the experimenter. Thus if the investor handed over all of his 12 units, the trustee now had 48, comprising his original 12 plus the tripled investment. He could then return any amount to the investor.

Is trust an emotion? Like love, trust is often manifested dispositionally rather than by occurrent qualities of awareness. When it is a full-fledged emotion, however, the subject is expected to be able to report that she feels trusting. In this case, the ascription of trust is based on functional criteria rather than on any first-person report of experience, of which there is no mention in the authors' account of the case. They speak of 'modulating trusting behaviour' rather than generating a feeling of trust. It seems safe to say, therefore, that we are dealing with a feeling operating at the subpersonal level.

However that may be, trust is a nice bridge between the strictly epistemic and the strategic; for although it inclines the subjects to believe identifiable propositions – in this case, the proposition that investment in this particular trustee would prove profitable for himself – it does so but only in the context of a transaction envisaged with a person. For the effects of oxytocin on the investor was not matched by any effect on the trustee. That showed that the causal factor responsible for the effect was not a general increase in benevolence or good feeling. For a vague benevolence would have presumably led to larger returns from the trustees as well as a larger amount risked by the investor. Another significant control contrasted the original set-up with a situation in which the investor was told that a computer, not a person, would determine the return to the investor. Again, in the alternative situation the oxytocin had no significant effect, despite the fact that the sheer probability of getting or not getting a return was not affected. This seems to show that oxytocin did not simply shift the estimate of risk down a notch; rather, it worked specifically on the feeling component of trust.

The authors note that there is substantial independent evidence 'that oxytocin promotes prosocial approach behaviour by inhibiting defensive behaviours' (Kosfeld et al. 2005, 675). In the light of this known effect of oxytocin on social approach in other mammals, they tend to minimize the specific effect on belief: 'the investors given oxytocin show more trusting behaviour but do not hold significantly different beliefs about the trustworthiness of others' (Kosfeld et al. 2005, 675). That is paradoxical, if we assume that in either case the behaviour of the investor follows a roughly Bayesian strategy. It can be partly though not wholly explained, according to the authors, by appealing to an evaluative rather than a strictly cognitive appraisal: what the chemical has done is help the investors 'overcome their betrayal aversion

in social interactions' (Kosfeld et al. 2005, 675). Still, the *consequence* of the diminished 'betrayal aversion' is equivalent to a change in the probability measure of the expectation of return. So we have here a kind of primitive, purely causal case of direct biological influence over a process that is functionally equivalent to making an inference, even though no explicit inference is made.

In any case, as might be expected, the effect of the oxytocin is not determining. It may contribute to a Svengali effect, but cannot guarantee its success and could hardly be credited with one all by itself. (The median amount entrusted by investors who had absorbed oxytocin was 25 per cent higher than those sprayed with a placebo.) The precise nature of the chemical influence on feelings, and thereby on cognitive processes, invites further research, both in the sort of case just mentioned and in cases of bona fide valid inference.

The Feeling of Knowing

Feelings of the fourth class have only recently come under scrutiny. These are metacognitive in their import, bearing not on beliefs already acquired or on hypotheses currently being entertained, but rather on the existence of hidden knowledge: on whether *we already know* something, *without knowing what it is we know*. *Feelings of knowing* include the familiar tip-of-the-tongue feeling, as well as the judgement that one has learnt something.

Asher Koriat has elaborated the hypothesis that feelings of knowing 'serve to interface between implicit-unconscious-automatic processes on the one hand, and explicit-conscious-controlled processes on the other' (Koriat 2000, 152). Under favourable conditions, such feelings appear to afford a moderately reliable indication that an item of information is retrievable from memory. This provides guidance on whether or not further attempts at retrieval are likely to be worthwhile.[6]

Apart from the way in which it emerges into consciousness from unavailable sources, what is most remarkable about the epistemic feelings is the subtlety and complexity of their mechanisms which remain, precisely because they normally escape awareness, exceedingly hard to analyse. There is some evidence that in the course of skilled physical activity humans' animal sense of what to expect follows Bayesian lines more accurately than we could ever consciously compute. In this respect, humans behave no differently than other animals whose inferential resources do not include explicit reasoning in language. Thus when a tennis player returns a ball, anticipation of the ball's position at the moment of impact depends on a Bayesian computation guided on the one hand by prior expectations and on the other by current sensory input. Körding and Wolpert (2004) report that the respective contributions of these two parameters are finely tuned to take account of the amount

6 Just as other emotions can be 'misleading' (Goldie, this vol.), however, it is relatively easy to trip up the feeling of knowing. Koriat notes that under certain experimental conditions 'impressive results documenting systematic discrepancies between metacognitive feelings and actual performance have been obtained not only with regard to [feeling of knowing] judgments … but also with regard to judgments of learning … and confidence judgments' (Koriat 2000, 161).

of uncertainty about the accuracy of the sensory output that may arise from poor light conditions, implementing the Bayesian calculus with remarkable accuracy. Aided by computers, we may come dispassionately to do as well in aiming rockets at Mars; but so long as our conscious deliberations are left to their own unaided and explicit devices, we can only rely on a kind of arbitrage among epistemic feelings, which itself can only proceed on the basis of more of those same feelings. Those who urge us to 'trust our intuitions' have this much on their side: that it will take a good deal of hard work to understand exactly how our minds do what appears to take no work at all. The findings of Körding and his colleagues serve to show that while the complete analytic understanding of these processes may be hard to achieve, it is not doomed to remain for ever out of reach. For the moment, though, while we may be able to provide mechanical criteria for the assessment of validity in deductive arguments, we are very far from knowing how to do this for inductive arguments.

4. The Normativity of Correct Inference

The correctness of inference remains a normative matter. How does this fact relate to our propensity to experience feelings of knowing or feelings of certainty? At the most basic level, the equipment at our disposal for assessing the goodness of a cognitive claim or information consists in learnt and innate dispositions to epistemic feelings, as well as social responsiveness. We innately possess, and refine through learning, certain dispositions to respond with feelings of rightness to the deliverances of our senses and to the emergence in consciousness of certain propositions. In addition, our feelings of rightness are subject to psycho-social facts about the power of social sanctions over individuals.

In a moment, I shall adduce a striking example of the depth of that power, not merely over individual emotions, but even over the content of perception. But it must be acknowledged at the outset that social pressure is sometimes – indeed, perhaps *often* – deplorably wrong-headed. The key to doing better lies in acknowledging something deeper than social pressure: an original evolutionary basis for some of the 'intuitive' judgements that are codified in social consensus. The point is clear in its application to our capacity to apprehend logical validity: while that ability is far from infallible, we have seen that it cannot dispense with the feeling of rightness. But this conclusion is not confined to the domain of logical inference. Let me illustrate this in terms of one more exhibit in my gallery.

Obsessive Compulsive Disorder (OCD)

In OCD, or obsessive compulsive disorder, patients experience a recurring need to check on whether they have closed the door, washed their hands or accomplished some other – sometimes necessary but often utterly trivial – task. I lay no claim to a diagnosis or explanation of this syndrome; but at least some aspects of it invite a description in terms of a failure of the feeling of certainty or rightness to be triggered by the memory of the required action. I closed the door. I *remember* closing the door. I may even be sure that I closed the door. But the memory fails to trigger the right sort

of feeling. There is some sort of disconnection of the normal feeling of certainty from the recent memory of having taken necessary precautions. The patient suffering from OCD does not lack the memory of having done what was needed, but lacks those normal feelings of rightness or certainty that would normally be triggered by that memory.

OCD used to be thought of as a neurotic syndrome calling for psychoanalytic diagnosis and therapy. Remarkably, however, some of these cases are capable of clearing up under the influence of a drug such as Prozac (Kramer 1993). A crucial epistemic feeling can be triggered or at least facilitated by a simple chemical agent, which apparently determines the presence or absence of conviction in a particular proposition. This suggests that this apparent complexity is sometimes an illusion. The feeling of doubt – or the failure of expected feelings of certainty – manifests itself on the cusp of consciousness in such a way as to play a crucial epistemic role. In its susceptibility to a drug targeting specific receptors for certain neurotransmitters, it also leaves traces of its underlying physiological or chemical nature.

A close neighbour of the feeling of certainty in the space of epistemic feelings is the *feeling of familiarity*. Certainty and familiarity are in part related as knowledge by description is to knowledge by acquaintance, since certainty relates to items of putative knowledge, while familiarity bears on particulars, as objects of perception or as apprehended scenes and situations. Most of us have experienced 'déjà vu' feelings, the misplaced feeling of familiarity that gives us the erroneous impression of having been here before. In pathological cases, following brain damage in certain fronto-lateral circuits of the brain, these can become insistent and deeply disruptive (Moulin et al. 2005). We can surmise that there might be an explanation for the existence of a mechanism for generating familiarity. There may have been, in the long run of evolution, a correlation between familiarity and beneficence. But from the epistemological point of view the tendency of the former to cause inference to the latter can hardly count as a justification. Yet the feeling of familiarity plays an important part in our lives even in cases where it is not conscious as such. In some cases where it breaks down, we get a pathological condition known as the Capgras Syndrome. This is the next exhibit on my tour.

The Capgras Syndrome

Patients afflicted with the Capgras syndrome persist in believing that a person close to them – wife or father – is an impostor. They never accuse persons of being impostors if they are merely acquainted with them, but not emotionally significant (Ramachandran and Blakeslee 1998; Pacherie 2005; Mangan 2001).

The best explanation for this strange disorder is that a direct link normally exists between the facial recognition mechanism and the areas controlling the appropriate emotional responses (particularly the amygdala). The sight of a close relative – a parent, in the case of Ramachandran's patient Arthur – normally triggers an affective response, which is itself subject to a 'familiarity' evaluation. In Arthur's case, the direct link to the area in charge of generating the affective response is missing. As a result, the affective response to his father is not produced. This sets up an incongruity between the strictly cognitive familiarity check that applies to the face

and the missing familiarity check effected by the expected affective response. The Capgras delusion is then no more than a perfectly reasonable inference (though of course one that is neither conscious nor explicit): *I get a characteristic thrill when my father appears; I'm not getting that father-thrill now; therefore the person before me is not my father. Yet he looks exactly like my father. Therefore he is an impostor, a stranger who looks just like my father.*

The present hypothesis explains why the 'impostor' delusion occurs only with persons to whom the person is emotionally bound. It does not occur with mere acquaintances, because in most cases of recognition a more or less indifferent emotional reaction is normal, not aberrant. That suggests that the emotional aspect of recognition is subject to an independent familiarity marker. Where the person recognized is both familiar and affectively significant, both markers, the heart and the head, as it were, must pass the required ID check.

This interpretation of the Capgras syndrome raises suggestive questions: could an animal be subject to the Capgras syndrome? If we think not, is it because their brains are organized differently, or because they only need the emotive marker – 'the heart'? If so, is this evidence that what constitutes the 'head' is causally tied to the presence of language? Or could one make a quasi-analytic, *theoretical* case for the impossibility of Capgras syndrome affecting a being without language?

Leaving the reader to pursue these idle speculations if she pleases, I will confine myself to an observation about the object of the feeling of familiarity. The passage from what I rendered as 'I didn't get the father-thrill' to what I rendered as 'he's an impostor' looks like an inference. But it is not experienced subjectively as an inference, but as intuitive conviction about the supposed father. Second, while the feeling of familiarity acts as a marker, it does not present itself as a marker of correct *inference* as such. In the OCD case, by contrast, it seems to be doing just that. For the absence of the feeling of rightness in OCD signals unreliability in the inference that goes from 'I remember locking it' to 'I locked it'. What is conveyed here by the feeling is a form of metacognition, though it is not recognized as such; for the presence or absence of the marker normally indicates that I am recognizing someone who has a significant role in my life. Here again, one might raise the question of whether that sort of duality requires a two-track mind, whether animals could have OCD, and what role is played by the explicit thought represented in the sentence 'I locked it'.

5. Social Influence: The Price of Non-Conformism

The last item in my cabinet of curiosities is perhaps the oddest. It stems from a new take on some famous experiments on the power of social conformism done in the 1950s. Solomon Asch had found that when asked to make a judgement of a visual quantity, some 40 per cent of subjects went along with the false judgement of their peers (Asch 1962). In the new variant, a group of researchers explored 'potency of social pressure in inducing conformity and how information that originates from humans, versus inanimate sources, alters either perception or decision making and the neural basis for such changes' (Berns et al. 2005, 250). The surprising aspect

of their findings is that they seem to show that social influence is exerted *directly* on perception: 'the effects of social conformity are exerted on the very same brain regions that perform the task' (Berns et al. 2005, 251). Using fMRI data, they found 'highly suggestive' the 'lack of concomitant activity changes in more frontal areas' (Berns et al. 2005, 251), where one might have expected activity if the subject's judgement had resulted from a decision to override their own judgement in favour of that of the majority. Berns et al. concluded that no special cognitive activity was detected in the cortex of those who conformed to others' false opinion. Rather, it was those who saw and stood up for an independent truth who suffered emotional disturbance. In other words, the distorting effect of conformity did not require any calculation of costs and benefits. Resisting conformity to stand by the truth, by contrast, did have an emotional cost. What is intriguing about this discouraging finding is that the emotion involved (though the authors make no attempt to pin-point it in our repertoire of normal emotions) does not seem to be among those listed above – in terms of simple common sense – as epistemic emotions.

6. Concluding remarks

In the various empirical observations I have cited, the feelings in question were sometimes objects of awareness, but at other times they were inferred from their effects on subsequent belief or behaviour, or from their neurophysiological underpinnings. This may seem like cheating: to be sure, the mere activation of neuromodulating chemicals, as instantiated in the cases I have just sketched, cannot be assimilated to the presence of a feeling, let alone an emotion. Emotions, as I have already observed, are typically ascribed at the *personal* level; the activation of neuromodulators is a *subpersonal* phenomenon. But some of what we are learning about the involvement of specific parts or functions of the brain in reasoning, illustrated by the exhibits in my gallery, implicates just such subpersonal factors, and it is not clear to what extent we are justified in assuming that if the same factors are involved in emotion, that shows that emotions, as commonly conceived, are involved in reasoning. I have used the word 'feeling' precisely in order not to be pinned down on where, in the personal/subpersonal dichotomy, they rightly belong. So long as we remain willing to countenance a range of feelings, akin to yet distinct from full-fledged emotions on standard lists, it seems we can hazard four concluding observations.

On the Connection between Biology and Rationality

Social pressure, however powerful, will not suffice to guarantee the normative correctness of our inclinations to draw inference. A sort of core of basic procedures – perhaps *modus ponens* and *modus tollens* – had to be installed by natural selection.[7] We have a native disposition to recognize some inferences as valid. This disposition

7 This should not be read as taking sides in favour of the existence of an abstract topic-neutral logical capability. That some procedures must be taken to trigger the feeling of certainty is quite compatible with the modularist view that different modules trigger this separately in different domains.

is separable from our capacity to provide an explicit and conclusive argument for their validity. When codified by rules of language, logic or mathematics, however, inference patterns become less reliable rather than more; the reason is that the fit between the 'native' disposition and its implementation in explicit language is not itself part of that system of mechanisms on which natural selection has put its certificate of warranty.

On the Normativity of Inference

There is a prima facie presumption of functionality to any heritable disposition the complexity of which makes it unlikely to be accidental. But what has been put in place by natural selection, however useful to our ancestors in the environment of evolutionary adaptation, may not be worthy now of any evaluative endorsement. Justification cannot be infallibly grounded. It has to run in circles, the capacity of which to inspire respect depends directly on their size. Ultimately we trust our epistemic feelings to tell us when the circle is big enough for comfort. The degree of comfort we feel in that big circle is the currency in which the normative warrant of inferences is ultimately cashed out.

Some Emotional Mechanisms Guide Cognitive Strategies at the Proximate Level

The surprising lesson of the sort of recent psychological and brain research of which I have exhibited a sampling is that in some cases a relatively abstract inference is triggered by what appears to be a fairly simple chemical agent. It would obviously be greatly exaggerated to conclude from the Zürich experiments that trust was simply triggered by oxytocin, or from those of Yu and Dayan that strategic research decisions were determined by noradrenaline. But these experiments are part of a accumulating body of evidence that suggests that emotional factors, more obviously linked to non-cortical brain and hormonal activities, are important to our judgements of what inferences are or are not acceptable.

The Link to the Two-track Mind

I have suggested that we must take seriously the hypothesis that there are two communicating but largely independent levels of processing: the 'two-track mind', and that emotions serve crucial functions in managing the traffic between them. Emotions are almost omnipresent in both, but much remains to be done in science and philosophy before the detail of that involvement is clear. Epistemic feelings cast a special light on this, because while they appear to hug the ground of neurophysiology more tightly than emotions in general, they are also manifestly important to different aspects of cognition which we generally assume are most prominently governed by 'reason'.

In the final analysis, the increasing precision of our understanding of the brain mechanisms underlying the actual and the normatively correct practices studied by psychology and epistemology may blur the very image I have sought to sketch. The idea that one should be able to distinguish specific contributions of feelings

to cognitive strategies, inference and knowledge, and even that there is always a clear answer to the question of what inferences have been drawn, presupposes that we can draw clear lines between the mind's two tracks, between emotional and merely evaluative determinants of decision making, and between the effects that brain chemicals have on reasoning and those they have on feelings. But the fine-grained picture, when it emerges, may overwrite the lines drawn in the sand by the presuppositions implicit in our concepts of inference, of normativity, of feeling and of emotion.

References

Asch, Solomon. 1962. *Social Psychology*. Englewood Cliffs, N.J.: Prentice-Hall.

Baars, Bernard J. 1997. 'In the Theatre of Consciousness. Global Workspace Theory, a Rigorous Scientific Theory of Consciousness'. *Journal of Consciousness Studies* 4 (4): 292–309.

Becker, Gary S.; Rubinstein, Yona. 2004. 'Fear and the Response to Terrorism. An Economic Analysis'. 'Preliminary and Incomplete' version online at http://www.ilr.cornell.edu/international/events/upload/BeckerrubinsteinPaper.pdf.

Berns, Gregory S.; Chappelow, Jonathan; Zink, Caroline F.; Pagnoni, Giuseppe; Martin-Skurski, Megan E.; Richards, Jim. 2005. 'Neurobiological Correlates of Social Conformity and Independence During Mental Rotation'. *Biological Psychiatry* 58: 245–53.

Bonabeau, Eric; Dorigo, Marco; Théraulaz, Guy. 1999. 'Ant Foraging Behavior, Combinatorial Optimization, and Routing in Communications Networks'. In: *Swarm Intelligence. From Natural to Artificial Systems*. Oxford: Oxford University Press.

Carroll, Lewis. 1895. 'What the Tortoise Said to Achilles'. *Mind* 4: 278–80. [Available online at http://www.ditext.com/carroll/tortoise.html.]

Cohen, Jonathan D.; Aston-Jones, Gary. 2005. 'Decision Amid Uncertainty'. *Nature* 436: 471–2.

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, Mass.: MIT Press.

de Sousa, Ronald. 2000. 'Deux Maximes de Rationalité Émotive'. In: Angehrn, Emil; Baertschi, Bernard (eds). *Emotion und Vernunft – Émotion et Rationalité. Studia Philosophica,* vol. 59. Bern: Paul Haupt.

Descartes, René. [1641] 1986. *Meditations on First Philosophy with Selections from the Objections and Replies*. Transl. John Cottingham, Robert Stoothoff, and Dugald Murdoch. Cambridge: Cambridge University Press.

Descartes, René. [1649] 1984. *The Passions of the Soul*. Vol. 1 of *The Philosophical Writings of Descartes*. Transl. John Cottingham, Robert Stoothoff, and Dugald Murdoch. Cambridge: Cambridge University Press.

Ekman, Paul; Davidson, Richard J. (eds). 1994. *The Nature of Emotions. Fundamental Questions*. New York, Oxford: Oxford University Press.

Wilson, Timothy D. 2002. *Strangers to Ourselves. Discovering the Adaptive Unconscious*. Cambridge, Mass.: Harvard University Press.

Yu, Angela P.; Dayan, Peter. 2005. 'Uncertainty, Neuromodulation, and Attention'. *Neuron* 46: 681–92.

Index