The Value of Emotions for Knowledge

Edited by
Laura Candiotto

The Value of Emotions for Knowledge

Laura Candiotto
Editor

The Value of Emotions for Knowledge

palgrave
macmillan

Editor
Laura Candiotto
Free University of Berlin
Berlin, Germany

Acknowledgements

Contents

Notes on Contributors

Anja Berninger is a Lecturer (akademische Rätin) at the University of Stuttgart. Her main field of research is Philosophy of Mind. She is especially interested in empathy, emotion and memory. Recent publications include the journal article "Thinking Sadly. In Favour of an Adverbial Theory of Emotions" (*Philosophical Psychology* 2016) as well as the German-language book-length study: "Gefühle und Gedanken. Entwurf einer adverbialen Emotionstheorie" (Mentis 2017).

Michael S. Brady is a Professor of Philosophy at the University of Glasgow. His research centres on the Philosophy of Emotion, and its links with Moral Philosophy and Epistemology. In 2013 his book *Emotional Insight* was published by Oxford University Press. He was recently Co-Investigator on a major interdisciplinary project on the Value of Suffering, hosted at Glasgow; his book on this topic, *Suffering and Virtue*, was published by OUP in 2018. He was Director of the British Philosophical Association, having previously served as Secretary of the Scots Philosophical Association. Outside of academia, he has acted as a philosophical advisor on a number of productions by the Manchester-based theatre company Quarantine.

Laura Candiotto is Alexander von Humboldt Foundation Senior Research Fellow at the Institute of Philosophy of the Free University of Berlin. She has also held positions at University of Edinburgh, University of Aix-Marseille and Ca' Foscari University of Venice. Her research on emotions merges her interest in Social Epistemology and Philosophy of Mind, Social Ontology, Theories of Happiness and Well-Being, and the History of Philosophy. Her publications include, "Purification through emotions" (*Educational Philosophy and Theory* 2018); "Boosting cooperation. The beneficial function of positive emotions in dialogical inquiry" (*Humana. Mente* 2017); "Socratic Dialogue faces the history. Dialogical inquiry as philosophical and politically engaged way of life" (*Culture and Dialogue* 2017); "Extended affectivity as the cognition of the primary intersubjectivity" (*Phenomenology and Mind* 2016). The volume *Emotions in Plato,* edited with Olivier Renaut, is forthcoming for Brill.

Andy Clark is a Professor of Cognitive Philosophy at the University of Sussex. He is the author of several books including *Surfing Uncertainty: Prediction, Action, and the Embodied Mind* (Oxford University Press 2016). He is currently PI on a European Research Council Advanced Grant 'Expecting Ourselves', looking at Consciousness and the Predictive Brain.

George Deane is currently a Ph.D. student on 'Expecting Ourselves', a 4-year ERC-funded interdisciplinary project led by Professor Andy Clark, applying philosophy and cognitive science to investigate the nature and mechanisms of conscious experience within the predictive brain. His work focuses on integrating predictive processing with existing theories of consciousness and embodied cognition.

Roberta Dreon is an Associate Professor of Aesthetics at Ca' Foscari University, Venice. Among her publications, the Italian volume "Out of the Ivory Tower: John Dewey's Inclusive Aesthetics" (Marietti 2012) has been translated in French and published by Questions Théoriques in 2017. She recently edited "The Venetian Lectures" held by Joseph Margolis for Mimesis International (2017). She published many articles on Pragmatism, Evolutionary Aesthetics, Merleau-Ponty's Phenomenology, Hermeneutics, as well as a book on Heidegger in 2003 and a volume on

the connections between sensitivity and language in 2007. She is currently working on a project on Pragmatist Anthropology, focusing on the themes of sensitivity, habits and human enlanguaged experience.

Owen Earnshaw completed his Ph.D. in 2011 in the Philosophy of Psychiatry, focussing on the Phenomenology of Delusions. Since then he has been involved in teaching and research at the University of Durham and the Open University. His main interests are in Phenomenology, the Philosophy of Mental Health and understanding emotions and moods. He is a member of the Applied Phenomenology Research Cluster and chairs the Phenomenology Reading Group at Durham Philosophy Department.

Pascal Engel is the director of studies at Ecole des hautes études en sciences sociales, Paris. His work is mostly in the Philosophy of Mind and Epistemology. He is the author of *Truth* (Cambridge University Press 2002), *Ramsey, Truth and Success* (with J. Dokic, Routledge 2002), *Va savoir* (Editions Hermann 2007) and *Les vices du savoir* (Agone 2018).

Anthony Hatzimoysis is an Associate Professor of Contemporary Philosophy at the University of Athens, and a Visiting Professor in Epistemology at Alliance Manchester Business School, and at Pepperdine Graziadio Business School, having held academic posts at the University College London (UCL), the University of Manchester, and Massachusetts Institute of Technology (MIT). He is the co-founder of the European Philosophical Society for the Study of Emotions. His main works include the volumes *Philosophy and the Emotions* (Cambridge University Press 2003), and *Self-Knowledge* (Oxford University Press 2011), and the monograph *The Philosophy of Sartre* (Routledge 2013).

Raamy Majeed is a Lecturer in Philosophy at The University of Auckland, New Zealand. He has also held positions at University of Cambridge, Lehigh University and The University of Otago. He currently works in Philosophy of Mind and Cognitive Science, and is the author of "Do Gestalts Effects Show That We Perceive High-Level Aesthetic Properties?" (*Analysis* 2018), "Why the Canberra Plan Won't

Help You Do Serious Metaphysics" (*Synthese* 2017) and "Conceptual Instability and the New Epistemic Possibility" (*Erkenntnis* 2016).

Dina Mendonça is a research member of the Institute of Philosophy (IFILNOVA) at Faculty of Social Sciences and Humanities of NOVA University of Lisbon (FCSH/NOVA). Holding a Ph.D. in Pragmatism by the University of South Carolina, USA (2003), she researches on Philosophy of Emotions and Philosophy for Children. Her current research is focused on the application of the Situated Approach to Emotions to Ethics and Decision Making. She is the author of several papers on emotion theory and, in addition, promotes and creates original material for application of philosophy to all schooling stages, and as an aid in creative processes.

Cecilea Mun specializes in the areas of Philosophy of Emotion and Mind, Epistemology, and Feminist Philosophy. She is the founding president of the Society for Philosophy of Emotion, and the founding editor of the *Journal of Philosophy of Emotion*. She is also the author of *Interdisciplinary Foundations for the Science of Emotion: Unification Without Consilience* (Lexington Books *forthcoming*), and the editor of *Interdisciplinary Perspectives on Shame: Methods, Theories, Norms, Cultures, and Politics* (Lexington Books *forthcoming*).

Kathryn Nave is a Ph.D. student on the 'Expecting Ourselves' project, led by Professor Andy Clark. Her work focuses on how insights from both Predictive Processing and work in the Phenomenological tradition can be combined to help us better understand the structure and function of visual experience.

João Sàágua is the rector of NOVA University of Lisbon and a Full Professor of the Faculty of Social Sciences and Humanities of NOVA (FCSH/NOVA). Holding a Ph.D. in Contemporary Philosophy he has been teaching at NOVA since 1990 accumulating a long experience as a professor at all levels of university teaching in Portugal and abroad. He was a Visiting Scholar at the University of Oxford, and combined several management positions with an intensive teaching and research activity being the author of several books and articles on Logic, Philosophy of Language and Argumentation Theory.

Dave Ward is a Lecturer in Philosophy in the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh. He works on the relationships between perception, agency and understanding. Key recent publications include: "What's Lacking in Online Learning? Dreyfus, Merleau-Ponty and Bodily Affective Understanding", (*Journal of Philosophy of Education* forthcoming); (*with* G. Ongaro) "An enactive account of placebo effects" (*Biology and Philosophy* 2017); "Achieving transparency: An argument for enactivism" (*Philosophy and Phenomenological Research* 2016).

Sam Wilkinson is a Lecturer in Philosophy at the University of Exeter. Before that, he was a postdoctoral researcher at the University of Edinburgh, working on the European Research Council Advanced Grant 'Expecting Ourselves', led by Professor Andy Clark. Sam has published several papers in Philosophy of Mind and Cognitive Science in journals like *Mind and Language, Consciousness and Cognition* and *Review of Philosophy and Psychology*.

Part I

Introduction

1

From Philosophy of Emotion to Epistemology: Some Questions About the Epistemic Relevance of Emotions

Laura Candiotto

This volume consists of twelve full-length articles by leading and up-and-coming academics on the exciting topic of *philosophy of emotions*. This very multidisciplinary field of research focuses on the different roles that emotions play in our life. Significant work has been done in the area of moral philosophy, for example highlighting the weight of emotions and other affective states in nurturing our moral concerns, leading our decision-making, and also disclosing what it is that we care about most.[1] The aim of this volume is to analyse the role of emotions in knowledge acquisition, in its many and different processes and functions, especially focusing at the intersection between epistemology and the philosophy of mind and cognitive science. In fact, if we easily recognise the value of emotions in our moral life, it is difficult to not think that emotions impair knowledge, intrude on reasoning, and express our

---

[1]For a comprehensive overview of the most relevant work that has been done in this field in the last decades, see the second and the third volume of Ben-Ze'ev and Krebs (2018).

L. Candiotto (✉)
Free University of Berlin, Berlin, Germany
e-mail: laura.candiotto@fu-berlin.de

epistemic faults and self-deception.[2] And we should thus admit that these seemingly bad emotional behaviours have strong implications for moral judgements too. Emotions disrupt very often indeed, just think about the emotional biases and implicit prejudices that hinder our correct understanding of a thing, or how fast we judge a person poisoned by ill-founded jealousy and rivalry. But, we should concede, emotions can also make the process of inquiry appealing, for example nurturing our motivation towards knowledge acquisition, or letting us perceive the salience of an experience for our self-understanding, as the trend of emotional intelligence in pop-culture has very well emphasised—although with commercial exploitations, in certain cases.[3]

Neuroscientists have discovered the integrated functionality of emotions and reasoning in our mental life. Emotions are now understood as a constitutive element of human rationality, grounding concept creation and deliberative thinking, and partaking in the various cognitive processes, rather than being framed in opposition to rationality (Gray et al. 2002; Pessoa 2008, 2013).[4] Although the value of emotions in our mental life has been recognised by cognitive science, epistemology has remained a bit suspicious about it. It is true that some important pieces of work in the epistemology of emotions have been already developed especially by those philosophers who look at the science of mind for illuminating questions about the nature of emotions.[5] But as

---

[2]For the conceptualisation of emotions as misguided judgements at the beginning of our philosophical history, see Sorabji (2000). A prominent topic, widely discussed since the antiquity, is the possible contribution of emotions to *akrasia* (weakness of the will) and self-deception. Many philosophers, and then psychologists, have been dealing with it until now. Exemplary cases for the contemporary philosophical debate on emotions are Rorty (1987) and Mele (2000).

[3]For an academic study on emotional intelligence, see Barrett and Salovey (2002).

[4]It is important to mention a significant shift in cognitive science from the interactionist models of the late Nineties to the more recent integrationist models for which emotion and cognition are deeply entangled in our mental life. The integrationist model, first developed in neuroscience, is now assumed by many specific research fields, from developmental psychology (Labouvie-Vief 2015) to theories of learning and skilfull behaviour (Gardiner 2015).

[5]But this does not mean that there is a general consensus about what our best science says regarding the emotions. In fact, different research programs have been developed throughout the years and the different results are also reflected in the philosophical conceptualisations about them. Some good exemplars are de Sousa (1987) on the two systems theory (also called "Two-Track

Christopher Hookway (2003) has clearly stated, everyone would agree in ascribing to jealousy or anger our intellectual mistakes, however very few theorists would accept ascribing a positive role to our affective responses in the formation of reliable beliefs. In fact, granting a significant role to emotions in cognitive performances does not mean to also admit their truth-conduciveness—and only the latter should count for epistemology.

One of the innovative traits of this volume is to discuss the conditions that rule the beneficial role that affective states as emotions, meta-emotions, and moods can play in our epistemic practice, avoiding their popular shortcomings, but also the extreme rationalism which refuses to ascribe any positive epistemic function to them. In epistemology as a normative discipline, it is fundamental to ask what counts as acquiring or having the knowledge, what contributes to epistemic success and at which level, what possesses epistemic relevance and salience. This volume shows that *emotions do count for our epistemic enterprise*, and against the scepticism about their possible positive role in knowledge, it highlights *the how* and *the why* of this potential, also exploring aspects of their functionality in relation to specific kinds of knowledge.

But what does it mean that *emotions contribute to knowledge*? In the history of philosophy many answers have been offered to reply to this question, focusing on specific emotions—for example love as one of the best driving forces to truth in Plato and Scheler—beliefs' a priori grounding, as in the Scottish Sentimentalism and German Romanticism, metacognitive feelings as emotional attunements to reasons[6] or, as in the case of doubt, affective evaluations of the epistemic status of a belief in the Pragmatist philosophy, and affective dispositions

Mind"), Griffiths (1997) and DeLancey (2001) on the evolutionary approach and Ekman's basic emotions, Prinz (2004) in relation to Damasio's somatic marker hypothesis, and Thagard (2008) on the interactionist model. For a new account on emotions grounded in the new research program of the predictive mind, see Chapter 5.

[6]Contemporary philosophy of emotions differentiates between emotions and metacognitive feelings. See Proust (2015) and Carruthers (2017).

to world experience in the Phenomenological and Existentialist traditions.[7] But instead of specifically looking to our rich philosophical history,[8] this volume addresses the fundamental questions that underlie the deep entanglement between emotions and knowledge, asking for the criteria that can grant it, and assessing them. In addition to questioning the epistemological relevance of emotions, much of the work undertaken in this volume is directed to specific types of knowledge, such as self-understanding, group-knowledge, and wisdom, and to specific functions played by certain emotions in these cases, such as disorientation in enquiry and contempt in practical reason. Therefore, this volume draws special attention to how the function of emotions in knowledge is dependent on the types of knowledge—and why emotions' efficacy increase in relation to *specific epistemic practices*. This means that the volume highlights the necessity of studying the epistemic relevance of specific types of emotions in precise epistemic contexts. Focusing on epistemic practices, as the revision of beliefs or collective enquiry, the volume also considers the role played by *epistemic subjects and communities* in epistemic agency, thus discussing the epistemic significance of their affective states in belief-forming practices.[9]

In addition to highlighting the innovative character of the volume, the aim of this introduction is also to discuss the relevance that emotions can play in our epistemic life considering the state of the art of the philosophical debate on emotions. The debate that lies at the interface between epistemology, theory of emotions, and cognitive science has been taken into consideration by some other important publications in

---

[7]These are not only historical references to our glorious past since many contemporary approaches have renewed the traditions with novel accounts. For example, consider the prominent Neo-Aristotelian tradition in virtue epistemology (Zagzebski 1996) and the sentimentalist philosophy of mind (Slote 2014), or those approaches that are strictly anchored to the phenomenological tradition, as the ones of Matthew Rattcliffe (2008), Thomas Fuchs (2017), and Shaun Gallagher and Dan Zahavi (2008).

[8]For an excellent overview on the history of philosophy of emotions see the new edited volume by Cohen and Stern (2017).

[9]Recently Joëlle Proust (2018) has highlighted the importance of looking at epistemic activities for grasping the social dimensions of knowledge and thus for developing what it has been called *socially extended epistemology*. In this volume, especially in Chapters 11–13, we show how much the understanding of shared/collective emotions matter for this enterprise.

the last years, as the edited volumes *Epistemology and Emotions* by Georg Brun et al. (2008) and *Emotions and Value* by Sabine Roeser and Cain Todd (2014), the very popular handbook by Julien Deonna and Fabrice Teroni, *The Emotions* (Deonna and Teroni 2012), the introduction to the philosophy of emotion by Carolyn Price (2015), and the new one by Michael Brady (2018), and some of the chapters of the *Oxford Handbook of Philosophy of Emotion* edited by Peter Goldie (2010), as for example the ones by Adam Morton (2010) and Kevin Mulligan (2010). This new collection of essays brings new light to the debate, introducing new questions, for example about the role of emotions in participatory sense-making (see Chapter 11), the value of suffering in wisdom (see Chapter 9), and the rationality of moods (see Chapter 13), or new answers to old questions, as the intentional character of emotions (see Chapters 2, 3, and 10), and their epistemic value in moral knowledge (Chapters 7–10). It also introduces new conceptual frameworks for understanding the role of emotions in knowledge, as predictive processing (see Chapter 5), and brings to the contemporary debate the conceptual frameworks of important traditions of thought, such as Pragmatism (see Chapters 4 and 6), Phenomenology and Existentialism (see Chapter 8), explicitly emphasising the necessity of studying the role of emotions in epistemic contexts and within the embodied and situated experience of an epistemic agent (see Chapters 1, 4, 5, 6 and 11).

As it has been effectively summarised by Georg Brun and Dominique Kuenzle (2008), five epistemic functions 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. These different functions can be articulated within different conceptual accounts. My strategy is thus the one of focusing on the three main models on emotions as evaluative judgements, bodily feelings, and perceptions, following the *fil rouge* of emotional intentionality for rising questions about their epistemic functions.

Emotions are conceptualised as equipped of intentionality (de Sousa 1987; his account is extensively discussed here in Chapters 2 and 3), that means that they can be said to be about something. For cognitive theories, emotions have evaluative judgements for their content (Nussbaum 2001, 2004; Solomon 1976). This means that the

intentional object of an emotion is *the value of* something or, said in another way, emotional intentionality discloses what is of value to the subject, those features of the intentional object that are significant to us.[10] Emphasising the cognitive valence of emotions in this way allows us to investigate their epistemological status. For example, we could ask about the conditions of the correctness of their intentional character (D'Arms and Jacobson 2000), their appropriateness to the situation (Brady 2013) and their requiredness (De Monticelli 2015), or the apparent objective dimension of their fittingness (Todd 2014).[11] There are alternatives to this view that look at the intentional object for assessing if emotions are justified, such as the attitudinal theory for which the evaluative judgement does not belong to the intentional object, but to the disposition towards it (Deonna and Teroni 2012), models that take the evaluative judgements as non-doxastic states (Roeser 2011), "perspectival", meaning that emotional response varies with the evaluative perspective (Greenspan 2003), and experiential representations (Montague 2009, 2014). There are also accounts that ask to look at emotion's justification as the coherence between the evaluative content of emotion and the subject's internally justified values, therefore building an important bridge between the objective and the subjective dimensions of emotion, their truth and authenticity (Salmela 2014). Finally, there are theories that ascribe to emotions a very peculiar kind of intentionality, the one of the "feeling-towards" (Goldie 2000), recognised as fundamental for practical reasoning and moral judgements (Goldie 2007). This line of investigation discloses important overlaps

---

[10]It should be noted that this thesis could be assumed in many different ways, from one more apt to relate it to desires and hedonic levels (Helm 2001; Oddie 2005) to the ones that look at affects as the basic components of constructed emotions (Barrett 2017). These differences partially depend on the very much debated topic of emotional valence in affective science (for an overview of the issue, see Colombetti 2005). The thesis for which emotions disclose values (Johnston 2001) can also be used for explaining the analogy between perception and emotions— as perceptions are directed to objects, emotions are directed to values (Sauer 2012)—and can be articulated within different versions of intentionalism, as it has been recently argued by Vanello (2018). This thesis can be taken in functional terms too, and thus arguing that emotions are not a type of perception, but they function as perceptions (Price 2015).

[11]It is important to notice that the topic of the intentionality of emotions related to their fittingness has been one of the most debated since the pioneer book by Kenny (1963).

between epistemology and ethics, but also with aesthetics (Skellekens and Goldie 2011; Roeser 2018), for example investigating the epistemic significance of feeling moved by and feeling attracted to a piece of art and assessing the epistemic valence of the appreciation of a fictional intentional object, thus nurturing an interdisciplinary approach to emotion theories (on the relevance of fiction and literature for grasping the meaning of emotions in our life, see here Chapters 7 and 10).

But it is not only the cognitive theories that can be a valid point of reference for an epistemological investigation of emotions. In fact, from the point of view of emotional phenomenology, we can assess *the intentionality of bodily feelings*, both inward and outward, also discussing the possible contribution of the pre-reflective dimension of the subjective experience of the living body in the generation of meaning (Petitmengin 2007), and specifically of the physiological dynamics of health and breath as constitutive of the subjective point of view (Depraz 2008). The feeling-centred theories of emotions, such as the neo-Jamesian model developed by Prinz (2004) through to the important influence of Damasio (1994, 1999), and—with different emphasis—the affective intentionality model by Slaby (2008) and the enactive approach by Colombetti (2014), ask to focus on embodied cognition and first-person experience for looking at emotional phenomenology, while also analysing their social embeddedness (Fuchs and De Jaegher 2009; Fuchs and Koch 2014), social sharing (Salmela 2012; De Jaegher 2015; Zahavi 2015), and social functioning and cultural scaffoldings (Griffiths and Scarantino 2008; Slaby 2014; Colombetti 2015; Colombetti and Krueger 2015; Candiotto 2016; Krueger and Szanto 2016).[12] From an epistemological point of view, these models invite, for example, to analyse the function played by affective bodily feelings in epistemic agency, also assessing their relevance, and to investigate affective bodily feelings from the perspective of first-person knowledge, in certain cases also challenging the very notion of *the subject of knowledge*. For Lorraine Code, in fact, mainstream epistemology has always focused on

---

[12]An important contribution to this debate is forthcoming, the *Routledge Handbook of phenomenology of emotions* (Szanto and Landweer, forthcoming).

the value of propositions, and not on the properties of the epistemic subjects, thus developing a "view from nowhere" (Code 1992). But, as Roberts has claimed, a philosophical study of emotions requires taking into account the point of view of the human participants in the creation of meanings and values (Roberts 2003, p. 37), since if we do not do so the risk of losing the epistemic significance of emotions is very high.[13]

Emotions are in fact what makes knowledge significant *to us*, what disclose *our* personal concerns, what nurture *our* epistemic responsibility and for which *we* need to be responsible—as Robert Solomon effectively said, reversing the Humean motto, about our choice to not be the passion's slave (Solomon 2003, p. 40).[14] But this does not mean avoiding epistemological investigation and instead moving to psychological or educational ones. As a matter of fact, in the last decades many new and comprehensive approaches to knowledge have been developed for taking the subject into account, such as virtue epistemology, and models of cognition that are more sensitive to the subject experience and her environment like the 4Ecognition approach, that deserves to be further investigated from the integrated point of view of epistemology and philosophy of emotions (see Chapters 1, 4–6, and 11). The alleged subjectivity of emotions could be very useful too, especially for deepening the analysis of certain kinds of knowledge, such as critical reflection and introspective knowledge (Goldie 2004, pp. 92–95; Solomon 2007, pp. 150–158). But—I want to stress it again clearly—if one could be convinced by the necessity of looking at the epistemic subject for better understanding our cognitive processes, it is not said that one would grant it for epistemology too. This challenge is crucial because

---

[13]It is important to notice that one of the main reasons of scepticism regarding a positive role of emotions in knowledge in our philosophical tradition has been exactly their subjective and private dimension that seems to be against the objective standards of knowledge. For an overview of the conceptualisation of emotions as subjective, see Calhoun (2004).

[14]This claim, that is strictly related to the cognitive account on the rationality of emotions, has been discussed well beyond the boundaries of the philosophical circles, notably in the Neo-Aristotelian affective turn in education (Nussbaum 1995; Kristjánsson 2018), but also in decision theory and economics (Kirman et al. 2010). But altought this widespread interest, the capacity to be responsible over our emotions is still controversial, especially if emotions are not conceptualized as cognitive, but as unbiden automatic responses that make us lose reason (Elster 1999, 2010).

it highlights the relevance of the investigation of this topic from the point of view of the epistemology of affective bodily feelings, in this case: the readers will find some argument for assuring a positive function for emotions in *embodied and situated epistemologies* in this volume, but we are far from a conclusive answer. And this is good news because we have many other paths to investigate in this exciting field of research. For instance, I see an interesting bridge between those accounts that stress the relevance of affective bodily feelings in our epistemic practices and the feminist tradition in epistemology (see for example Jaggar 1989; Alcoff and Potter 1992),[15] for which we could examine whether the epistemic agent's emotional involvement is a crucial aspect that underlies our epistemic practices, or if it is accidental, especially in relation to gender, race, and social status differences. Another chief path that deserves to be further explored is the one that assesses the necessity of overcoming the personal/social divide that has been quite strong in the feeling-centred approaches to emotions (an excellent guide to this route is Protevi 2009). And this path may bring one to travel through the enactivist approaches to emotions in knowledge, focusing on how organisms affectively perceive the environment through their action-tendencies (Ellis 2005; Slaby et al. 2013) and pragmatic and epistemic actions (Wilutzky 2015), also disclosing relevant bridges to dynamical approaches to knowledge (Livet 2016) and the pragmatist tradition that asks to move from the study of knowledge as true belief to epistemic activities (Hookway 1990). Therefore, emotions can be the trojan horse for bringing the issue of the subjectivity of knowledge to mainstream epistemology, asking for a non-reductionist approach to affective bodily-feelings that, without forgetting the complexity of the subjects' emotional experience, can benefit from important recent results in the intertwined field of embodied neuroscience and phenomenology.

---

[15]It should be noted that the feminist reflection on emotions has been mostly pursued within the field of political philosophy and cultural studies (see, for example, Mendus 2000; Ahmed 2014). However, I think that the prominent research activity pursued by feminist epistemology and philosophy of science can be a valid and challenging reference for addressing the topic of emotions in knowledge too.

Coming back to our main topic, a middle ground between the feeling-centred theories of emotion and the evaluative judgement theories have been offered by the perceptual model for which emotions are *perceptions of evaluative properties* (Elgin 1996, 2008; Tappolet 2000, 2012; Döring 2007), rather than inferences. For the perceptual account, emotions are cognitive—like the judgement theories, but their evaluation has a rich phenomenology that should be taken into account—like the feeling-centred theories, as directly revealing what is meaningful to the subject (Hatzimoysis 2003), or even what it is inaccessible to the subject in other ways (de Sousa 1998).[16] The point here is to look at alternatives to propositional knowledge, such as the epistemic immediacy of direct perception and disclosing response-dependent properties of an object (Elgin 1996, 2008). Just as when we say that the party last night was amusing or that the family dinner was depressing, thus getting access to the properties of the party and the dinner from the direct experience of the perceiver. For this model, propositional knowledge may rest upon non-propositional elements, such as these quasi-perceptions that are emotions. If emotions are perceptions, then they generate beliefs and, therefore, epistemology should ask if these beliefs are justified or reliable and on what grounds. The perceptual model has also found an immediate epistemological valence thanks to the research leaded by Linda Zagzebski (2003) in virtue epistemology. In this case, the fittingness of an emotion does not depend on beliefs only, but also by some character-traits of the epistemically responsible subject, such as her trustworthiness (Zagzebski 2012), her concern (Candiotto 2017a, b) and existential commitments (Slaby and Wüschner 2014), or intellectual humility (Tanesini 2008). The perceptual model is widely discussed in the volume, both in arguing for and against it, see for example Chapters 10 and 12.

From this brief overview of the implications for epistemology from the debate in philosophy of emotion, we can derive that the investigation regarding the role of emotions in knowledge does not only ask

---

[16]For an overview of the psychological literature about the role played by the affects in self-disclosure, see Forgas and Moylan (2002).

to assess the functions of emotions in knowledge, but also and more fundamentally requires to put into question *the very notion of knowledge*. Accordingly, some of the chapters not only explore the evaluative knowledge ascribed to emotions as its object, but also explore the role of emotions in those kinds of knowledge that put the subjects, the epistemic communities, and their practices at the centre, like self-awareness, reflexivity, revision of beliefs, social understanding, and group knowledge-building. The volume also tests the capacity of other epistemologies, like the one of enactivism and predictive processing, in giving a good account of the role of emotions in knowledge. This step towards *epistemic practices* and the *how* of knowing, fundamental for the pragmatist epistemology (Hookway 1990, 2000), has been argued for many years for example by Goodman and Elgin (1988), and in general by the externalist approaches to knowledge. The externalist turn in epistemology seems in fact to better fit the enquiry on the epistemic relevance of emotions because it works on those processes of acquisition and generation of beliefs, instead of a more traditional investigation of the internal justification, where emotions seem to perform significant functions, as markers of salience or motivations for enquiry, for example. Moreover, an externalist approach to emotions can also contribute to developing the understanding of epistemic agency in the different and many practices of epistemic subjects and communities.

However, if emotions are taken in the standard way to be private states of the mind, it seems implausible to not consider what the traits of a mature epistemic agent are which count most for knowledge, and ask if emotions play a part in these. Saying this, I want to highlight that in discussing the function of emotions in knowledge from the point of view of epistemology we also need to assess the conditions that allow the subject and the epistemic communities to be competent knowers employing emotional rationality, and so an internalist account of knowledge as virtue epistemology, for example, seems to be needed too. This volume does not take a partisan position and, beyond the divide between internalism and externalism in epistemology, wishes to nurture a multi-focal perspective for focusing on the different aspects related to the potential relevance of emotions in epistemic practices. This also means undertaking an interdisciplinary

approach that, converging on the shared need of furthering our understanding of emotions in knowledge, explores their role from different perspectives. Therefore, in this volume we both look at emotions from the perspective of cognitive science and one of literature and fiction, for example, highlighting the importance of the meeting between science and humanities.

This volume focuses on particular ways of carving out new territory at the intersection of epistemology and philosophy of mind and cognitive science, by exploring the epistemic role of emotions to the theory of knowledge in particular. The questions that have animated the debate around this volume are many, from the most fundamental, such as *why* and *how* emotions have a special significance in the acquisition of knowledge, to more specific ones, related to emotional rationality, epistemic practices, and communities of inquiry. The following theoretical questions highlight some of the challenges around which the volume is centred and are a representative sample of the many research questions which philosophy of emotions raises for epistemology:

- What does it mean for emotions to be rational? How can emotions be vehicles of knowledge? What is the relation between emotions and beliefs? What is emotions' biological function? What is the impact of meta-emotions in reasoning? Could moods be rational?
- How do our reflections on emotions shape our self-awareness and self-understanding? What makes these self-reflections more, or less, accurate? What is the role played by narrativity in practical reasoning, and how emotions contribute to it? Why does suffering matter to wisdom?
- What relevance do group level emotions have for acts of group cognition, participatory sense-making, and decision-making? What are the character traits that are beneficial or detrimental to group knowledge? What is the epistemic value of emotive bodily gestures and affective bodily feelings?

Contributors to *The Value of Emotions for Knowledge* engage with these questions as well as issues related to specific emotions. The contributions that follow all explore various important connections among emotion,

rationality, and knowledge, including the role that emotions play in different kinds of knowledge. The book is organized in 6 parts which highlight the different epistemic values that can be ascribed to emotions. Part I consists of the introduction to the volume and it addresses the main challenges that arise in bringing emotions under epistemic scrutiny, also presenting the many and different features of the new paths of investigation in the field. Part II addresses the epistemic status of emotion in rationality for answering the traditional questions about the intentionality of emotions, especially looking at the pioneering work by Ronald de Sousa in the field. Part III examines the epistemic function of emotions in the intersections among brain, body, and environment from the point of view of 4E Cognition and Pragmatism. Part IV explores the specific role performed by emotions in the kind of self-understanding that is involved in autonomous agency and in critical thinking. Part V travels through the possible epistemic value of negative emotions and painful feelings, also discussing their relationship to moral values. Part VI inquiries into group level emotions and moods for detecting their epistemic value both in individual and group knowledge. Therefore, the *leitmotiv* of emotional intentionality flows from the more speculative and conceptual explorations of Parts I, II and III to the application of the value of emotions to specific types of knowledge and epistemic practices, for providing innovative and original snapshots at the crisscross of philosophy of emotions and epistemology.

In what follows, I offer a brief overview of each specific chapter.

Chapters 2 and 3 analyse the fundamental question of the rationality of emotions discussing de Sousa's account. In Chapter 2, *How Emotions Know: Naturalizing Epistemology via Emotions*, Cecilea Mun highlights the import of the intentionality of emotions to the knowledge of the world that we can gain in virtue of our emotional responses. In Chapter 3, *What Can Information Encapsulation Tell Us About Emotional Rationality?,* Raamy Majeed assesses de Sousa's hypothesis about information encapsulation and, contrary to de Sousa, argues that it is not essential to emotion-driven reasoning, as emotions can determine the relevance of response-options even without being encapsulated.

Chapters 4–6 explore new conceptual frameworks for understanding the epistemic value of emotions, especially looking at those accounts that provide prominence to what the environment offers to knowledge-acquisition. Chapter 4, *A Pragmatist View of Emotions: Tracing its*

Significance for the Current Debate, by Roberta Dreon, drives us to the pragmatist account of emotions highlighting its paramount importance for understanding the role of emotions in the different processes of knowledge-building, also underlying significant connections with the contemporary debate, especially with the extended and enactive accounts to cognition. In Chapter 5, *Getting Warmer: Predictive Processing and the Nature of Emotion*, Sam Wilkinson, George Deane, Kathryn Nave, and Andy Clark offer predictive processing as a new perspective on emotion. The upshot is a picture of emotion as inseparable from perception and cognition, and a key feature of the embodied mind. Chapter 6, *Emotional Reflexivity in Reasoning: The Function of Describing the Environment in Emotion Regulation,* Dina Mendonça and João Sàágua explore the intentionality of meta-emotions arguing that the recognition of reflexivity of emotions is crucial for understanding the role of emotions in reasoning, also showing how adopting a Situated Approach to Emotions is well suited for explaining the refinement and complexity of emotion regulation.

Chapters 7 and 8 inaugurate the analysis on the epistemic functions played by emotions in specific kinds of epistemic practices. In Chapter 7, *Moving Stories: Agency, Emotion, and Practical Rationality*, Dave Ward discusses J. David Velleman's conception of being an agent as having the capacity to be motivated by a drive to act for reasons, arguing that our capacities to render ourselves intelligible are built upon a bedrock of emotionally suffused narrative understanding. In this chapter the epistemic value of emotions is understood as what mediate the kind of self-understanding that is involved in autonomous agency. In Chapter 8, *Disorientation and Cognitive Enquiry,* Owen Earnshaw suggests that the experience of disorientation is a background affect in intellectual enquiry, both motivating the enquiry and being necessary to instill certain epistemic virtues in the inquirer. Discussing *Bewilderment* through the lenses of the phenomenological and existentialist tradition, Earnshaw argues that it is an emotion that is evoked through the encounter with the "mystery" and that it has a role in cognitive enquiry as an indicator of where the boundary of sense has been overstepped by nonsense.

Chapters 9 and 10 offer two intriguing and different explanations about the epistemic value of negative emotions. In Chapter 9, *Learning*

from Adversity: Suffering and Wisdom, Michael Brady extends the positive perspective of the value of suffering, by examining the idea that suffering is necessary for *wisdom*. Chapter 10, *The Grapes of Wrath and Scorn*, by Pascal Engel, argues for a positive answer to the question about emotional rationality through an examination of two negative emotions, anger and contempt. The chapter suggests that these emotions are apt to deliver, albeit in an indirect way, a form of moral knowledge, and examine their expression in the writings of Jonathan Swift.

The last three chapters are dedicated to the role of emotions and affective atmospheres in the intersubjective and social dimensions of knowledge. Chapter 11, *Emotions In-Between: The Affective Dimension of Participatory Sense-Making*, by Laura Candiotto, discusses and evaluates the epistemic role of emotions in participatory sense-making, integrating 4Ecognition and virtue responsibilism in the understanding of emotions as socially extended motivations for a shared meaning. The chapter argues for a fundamental role played by emotions in boosting epistemic cooperation and determining the quality of social bonds. In a similar vein, but employing a different conceptual framework, the adverbial theory to emotion, Anja Berninger, in Chapter 12, *Group Emotions and Group Epistemology*, argues for understanding emotions within group contexts as a way of thinking that can facilitate cooperation and create a joint epistemic outlook. In Chapter 13, *In Search for the Rationality of Moods,* Anthony Hatzimoyisis assesses the three main options we have for grasping the meaning of a quite unexplored area of research, the intentionality of moods. In fact, this understanding is hindered by the complicated question about the rationality of a state that seems to be without object. And given that moods do not seem to bear an intentional relation to an object, it is hard to see how they can be in the offing for rational assessment. This chapter ends the volume returning to the fundamental question about the rationality of emotions which has been explored in Part I and II, but bringing new light to it thanks to the analysis of the embeddedness of moods.

The contributions to this volume testify to the rich overlaps between philosophy of emotions and epistemology, and the manifold links with philosophy of mind and cognitive science. They provide different paths for understanding why emotions matter for knowledge, also challenging

the narrow boundaries between disciplines and asking to enlarge the scope of the investigation to the different epistemic practices of our epistemic communities. Throughout the discussion of the three main approaches available nowadays about the intentionality of emotions, I recalled the idea that emotions are markers of values. I hope that this volume will disclose to the readers, thanks to our focus on the manifolds functions of emotions in the different epistemic practices, the value of knowledge too. Especially in its ethical dimension as epistemic responsibility (Fricker 2007), reliable processes of knowledge building and acquisition are necessary nowadays more than ever, in the time of fake news and epistemic injustice for the preservation of ignorance. May our affective concerns towards knowledge support our best epistemic practices.

References

Ahmed, S. 2014. *The Cultural Politics of Emotion*. Edinburgh: Edinburgh University Press.

Alcoff, L., and E. Potter (eds.). 1992. *Feminist Epistemologies*. London: Routledge.

Barrett, L.F. 2017. *How Emotions Are Made: The Secret Life of the Brain*. Boston, MA: Houghton Mifflin Harcourt.

Barrett, L.F., and P. Salovey (eds.). 2002. *The Wisdom in Feeling: Psychological Processes in Emotional Intelligence*. New York and London: The Guilford Press.

Ben-Ze'ev, A., and A. Krebs (eds.). 2018. *Philosophy of Emotion*, vol. 5. London and New York: Routledge.

Brady, M. 2013. *Emotional Insight: The Epistemic Role of Emotional Experience*. Oxford: Oxford University Press.

———. 2018. *Emotion: The Basics*. Abingdon: Routledge.

Brun, G., and D. Kuenzle. 2008. A New Role for Emotions in Epistemology? In *Epistemology and Emotions*, ed. G. Brun, U. Doğuoğlu, and D. Kuenzle, 1–32. Aldershot: Ashgate.

Brun, G., U. Doğuoğlu, and D. Kuenzle (eds.). 2008. *Epistemology and Emotions*. Aldershot: Ashgate.

Calhoun, C. 2004. Subjectivity and Emotion. In *Thinking About Feeling: Contemporary Philosophers on Emotions*, ed. R.C. Solomon, 107–121. New York: Oxford University Press.

Candiotto, L. 2016. Extended Affectivity as the Cognition of the Primary Intersubjectivity. *Phenomenology and Mind* 11: 232–241.

———. 2017a. Epistemic Emotions: The Building Blocks of Intellectual Virtues. *Studi di Estetica, XLV, IV SERIE* 7: 7–25.

———. 2017b. The Route of Goodness: Epistemic Emotions, Self-Realization, and Perfection. *Thaumàzein* 4: 243–258.

Carruthers, P. 2017. Are Epistemic Emotions Metacognitive? *Philosophical Psychology* 30 (1–2): 58–78.

Code, L. 1992. Taking Subjectivity into Account. In *Feminist Epistemologies*, ed. L. Alcoff and E. Potter, 15–48. London and New York: Routledge.

Cohen, A., and R. Stern (eds.). 2017. *Thinking About the Emotions*. Oxford: Oxford University Press.

Colombetti, G. 2005. Appraising Valence. *Journal of Consciousness Studies* 12 (8:10): 103–126.

———. 2014. *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge, MA and London: MIT Press.

———. 2015. Extending the Extended Mind: The Case for Extended Affectivity. *Philosophical Studies* 172 (5): 1243–1263.

Colombetti, G., and J. Krueger. 2015. Scaffoldings of the Affective Mind. *Philosophical Psychology* 28 (8): 1157–1176.

D'Arms, J., and D. Jacobson. 2000. The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophy and Phenomenological Research* 61 (1): 65–90.

Damasio, A. 1994. *Descartes' Error: Emotion, Reason, and the Human brain*. New York: Avon Books.

Damasio, A. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.

De Jaegher, H. 2015. How We Affect Each Other: Michel Henry's 'Pathos-With' and the Enactive Approach to Intersubjectivity. *Journal of Consciousness Studies* 22 (1–2): 112–132.

DeLancey, C. 2001. *Passionate Engines: What Emotions Reveal About Mind and Artificial intelliGence*. New York, NY: Oxford University Press.

De Monticelli, R. 2015. The Sensibility of Reason: Outline of a Phenomenology of Feeling. *Thaumàzein* 3: 139–159.

Deonna, J.A., and F. Teroni. 2012. *The Emotions: A Philosophical Introduction*. London and New York: Routledge.

Depraz, N. 2008. The Rainbow of Emotions: At the Crossroads of Neurobiology and Phenomenology. *Continental Philosophy Review* 41 (2): 237–259.

de Sousa, R. 1987. *The Rationality of Emotion*. Cambridge, MA: MIT Press.

———. 1998. Modeling Rationality: A Normative or Descriptive Task? In *Modeling Rationality, Morality and Evolution,* ed. P. A. Danielson, 119–134. New York: Oxford University Press.

Döring, S. 2007. Seeing What to Do. Affective Perception and Rational Motivation. *Dialectica* 61 (3): 363–394.

Elgin, C. 1996. *Considered Judgment*. Princeton, NJ: Princeton University Press.

—— 2008. Emotion and Understanding. In *Epistemology and Emotions*, ed. G. Brun, U. Doğuoğlu, and D. Kuenzle, 33–49. Aldershot: Ashgate.

Ellis, R.D. 2005. *Curious Emotions: Roots of Consciousness and Personality in Motivated Action*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Elster, J. 1999. *Strong Feelings: Emotion, Addiction and Human Behaviour*. Cambridge, MA: The MIT Press.

Elster, J. 2010. Self-Poisoning of the Mind. *Philosophical Transactions B* 365: 221–226.

Forgas, J.P., and S.J. Moylan. 2002. Affective Influences on Self-Perception and Self-Disclosure. In *The Social Self: Cognitive, Interpersonal and Intergroup Perspectives*, ed. J.P. Forgas and K.D. Williams, 73–96. New York: Taylor and Francis.

Fricker, M. 2007. *Epistemic Injustice: Power & the Ethics of Knowing*. Oxford: Oxford University Press.

Fuchs, T. 2017. *Ecology of the Brain*. Oxford: Oxford University Press.

Fuchs, T., and H. De Jaegher. 2009. Enactive Intersubjectivity: Participatory Sense-Making and Mutual Incorporation. *Phenomenology and the Cognitive Sciences* 8 (4): 465–486.

Fuchs, T., and S. Koch. 2014. Embodied Affectivity: On Moving and Being Moved. *Frontiers in Psychology* 5 (508): 1–12.

Gallagher, S., and D. Zahavi. 2008. *The Phenomenological Mind*. Abingdon and New York: Routledge.

Gardiner, M.F. 2015. Integration of Cognition and Emotion in Physical and Mental Actions in Musical and Other Behaviours. *Behavioral and Brain Sciences* 38: e76.

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

———. 2004. Emotion, Feeling, and Knowledge of the World. In *Thinking About Feeling: Contemporary Philosophers on Emotions*, ed. R. Solomon, 91–106. Oxford: Oxford University Press.

———. 2007. Seeing What Is the Kind of Thing to Do: Perception and Emotion in Morality. *Dialectica* 61(3): 347–361.

———. (ed.). 2010. *The Oxford Handbook of Philosophy of Emotion*. Oxford: Oxford University Press.

Goodman, N., and C.Z. Elgin. 1988. *Reconceptions in Philosophy and Other Arts and Sciences*. London: Routledge.

Gray, J.R., T.S. Braver, and M.E. Raichle. 2002. Integration of Emotion and Cognition in the Lateral Prefrontal Cortex. *Proceedings of the National Academy of Sciences USA* 99: 4115–4120.

Greenspan, P. 2003. Emotions, Rationality, and Mind-Body. In *Philosophy and the Emotions*, ed. A. Hatzimoysis, 113–125. Cambridge: Cambridge University Press.

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

Griffiths, P., and A. Scarantino. 2008. Emotions in the Wild: The Situated Perspective on Emotions. In *The Cambridge Handbook of Situated Cognition*, ed. P. Robbins and M. Aydede, 437–453. New York: Cambridge University Press.

Hatzimoysis, A. 2003. Emotional Feelings and Intentionalism. In *Philosophy and the Emotions*, ed. A. Hatzimoysis, 105–112. Cambridge: Cambridge University Press.

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

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

———. 2000. Regulating Inquiry: Virtue, Doubt, and Sentiment. In *Knowledge, Belief, and Character*, ed. G. Axtell, 149–160. Lanham: Rowman & Littlefield.

———. 2003. Affective States and Epistemic Immediacy. In *Moral & Epistemic Virtues*, ed. M. Brady and D. Pritchard, 75–92. Oxford: Blackwell.

Jaggar, A. 1989. Love and Knowledge: Emotion in Feminist Epistemology. *Inquiry* 32 (2): 151–176.

Johnston, M. 2001. The Authority of Affect. *Philosophy and Phenomenological Research* 63 (1): 181–214.

Kenny, A. 1963. *Action, Emotion and Will*. London, UK: Routledge & Kegan Paul.

Kirman, A., P. Livet, and M. Teschl (eds.). 2010. Rationality and Emotions. *Philosophical Transactions of the Royal Society B* 365 (1538).

Kristjánsson, K. 2018. *Virtuous Emotions*. Oxford: Oxford University Press.

Krueger, J., and T. Szanto. 2016. Extended Emotions. *Philosophical Compass* 11 (12): 863–878.

Labouvie-Vief, G. 2015. *Integrating Cognition and Emotion Throughout the LifeSpan*. Cham, Heidelberg, Dordrecht, London, and New York: Springer.

Livet, P. 2016. Emotions, Beliefs, and Revisions. *Emotion Review* 8 (3): 240–249.

Mele, A. 2000. Self-Deception and Emotion. *Consciousness & Emotion* 1 (1): 115–137.

Mendus, S. 2000. *Feminism and Emotion: Readings in Moral and Political Philosophy*. Basingstoke: Palgrave Macmillan.

Montague, M. 2009. The Logic, Intentionality, and Phenomenology of Emotion. *Philosophical Studies* 145 (2): 171–192.

———. 2014. Evaluative Phenomenology. In *Emotions and Value*, ed. S. Roeser and C. Todd, 32–51. Oxford: Oxford University Press.

Morton, A. 2010. Epistemic Emotions. In *The Oxford Handbook of Philosophy of Emotion*, ed. P. Goldie, 385–399. Oxford: Oxford University Press.

Mulligan, K. 2010. Emotions and Values. In *The Oxford Handbook of Philosophy of Emotion*, ed. P. Goldie, 475–500. Oxford: Oxford University Press.

Nussbaum, M.C. 1995. Emotions and Women's Capabilities. In *Women, Culture, and Development: A Study of Human Capabilities*, ed. M. Nussbaum and J. Glover, 360–395. Oxford: Oxford University Press.

———. 2001. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.

———. 2004. Emotions as Judgments of Value and Importance. In *Thinking About Feeling: Contemporary Philosophers on Emotions*, ed. R.C. Solomon, 183–99. New York: Oxford University Press.

Oddie, G. 2005. *Value, Reality and Desire*. New York: Oxford University Press.

Pessoa, L. 2008. On the Relationship Between Emotion and Cognition. *Nature Reviews Neuroscience* 9:148–158.

———. 2013. *The Cognitive-Emotional Brain: From Interactions to Integration*. Cambridge, MA: MIT Press.

Petitmengin, C. 2007. Towards the Source of Thoughts: The Gestural and Transmodal Dimension of Lived Experience. *Journal of Consciousness Studies* 14 (3): 54–82.

Price, C. 2015. *Emotion*. Cambridge: Polity Press.

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

Protevi, J. 2009. *Political Affect: Connecting the Social and the Somatic*. Minneapolis and London: University of Minnesota Press.

Proust, J. 2015. The Representational Structure of Feelings. In *Open Mind*, ed. T. Metzinger and J.M. Windt, Chapter 31. Frankfurt am Main: Mind Group.

Proust, J. 2018. Consensus as an Epistemic Norm of Group Acceptance. In *Socially Extended Epistemology*, ed. J.A. Carter, A. Clark, J. Kallestrup, S.O. Palermos, and D. Pritchard, 132–154. Oxford: Oxford University Press.

Ratcliffe, M. 2008. *Feelings of Being: Phenomenology, Psychiatry and the Sense of Reality*. Oxford: Oxford University Press.

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

Roeser, S. 2011. *Moral Emotions and Intuitions.* Basingstoke: Palgrave Macmillan.

———. 2018. Socially Extended Deliberations About Risks: A Role for Emotions and Arts. In *Socially Extended Epistemology*, ed. J.A. Carter, A. Clark, J. Kallestrup, S.O. Palermos, and D. Pritchard, 158–172. Oxford: Oxford University Press.

Roeser, S., and C. Todd. 2014. *Emotions and Value*. Oxford: Oxford University Press.

Rorty, A. 1978. Explaining Emotions. *The Journal of Philosophy* 75 (3): 139–161.

Salmela, M. 2012. Shared Emotions. *Philosophical Explorations* 15(1): 33–46.

———. 2014. *True Emotions*. Amsterdam: John Benjamin Publishing.

Sauer, H. 2012. Psychopaths and Filthy Desks: Are Emotions Necessary and Sufficient for Moral Judgment? *Ethical Theory and Moral Practice* 15 (1): 95–115.

Skellekens, E., and P. Goldie (eds.). 2011. *The Aesthetic Mind: Philosophy and Psychology*. Oxford and New York: Oxford University Press.

Slaby, I. 2008. Affective Intentionality and the Feeling Body. *Phenomenology and the Cognitive Sciences* 7(4): 429–444.

———. 2014. Emotions and the Extended Mind. In *Collective Emotions*, ed. M. Salmela and C. von Scheve, 32–46. Oxford: Oxford University Press.

Slaby, I., and P. Wüschner. 2014. Emotions and Agency. In *Emotions and Value*, ed. S. Roeser and C. Todd, 212–228. Oxford: Oxford University Press.

Slaby, I., A. Paskaleva, and A. Stephan. 2013. Enactive Emotion and Impaired Agency in Depression. *Journal of Consciousness Studies* 20 (7–8): 33–55.

Solomon, R. 2007. *True to Our Feelings: What Our Emotions Are Really Telling Us*. Oxford: Oxford University Press.

Solomon, R. 1976. *The Passions.* New York, NY: Doubleday.

———. 2003. *Not Passion's Slave: Emotions and Choice*. Oxford: Oxford University Press.

———. 2014. *A Sentimentalist Theory of the Mind*. Oxford: Oxford University Press.

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

Szanto, T., and H. Landweer (eds.). Forthcoming. *The Routledge Handbook of Phenomenology of Emotions*. Abingdon: Routledge.

Tanesini, A. 2008. Virtues, Emotions, and Fallibilism. In *Epistemology and Emotions*, ed. G. Brun, U. Doğuoğlu, and D. Kuenzle, 67–82. Aldershot: Ashgate.

Tappolet, C. 2000. *Emotions et valeurs*. Paris: Presses Universitaires de France.

———. 2012. Emotions, Perceptions, and Emotional Illusions. In *Perceptual Illusions: Philosophical and Psychological Essays*, ed. C. Calabi, 207–224. Basingstoke: Palgrave-Macmillan.

Thagard, P. 2008. How Cognition Meets Emotion: Beliefs, Desires and Feelings as Neural Activity. In *Epistemology and Emotions*, ed. G. Brun, U. Doğuoğlu, and D. Kuenzle, 163–184. Aldershot: Ashgate.

Todd, C. 2014. Relatively Fitting Emotions and Apparently Objective Values. In *Emotions and Values*, ed. S. Roeser and C. Todd, 90–104. Oxford: Oxford University Press.

Vanello, D. 2018. Affect, Perceptual Experience, and Disclosure. *Philosophical Studies* 175: 2125–2144.

Wilutzky, W. 2015. Emotions as Pragmatic and Epistemic Actions. *Frontiers in Psychology* 6: Article 1593.

Zagzebski, L. 1996. *Virtues of the Mind.* Cambridge: Cambridge University Press.

———. 2003. Emotion and Moral Judgement. *Philosophy and Phenomenological Research* 66 (1): 104–24.

———. 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford: Oxford University Press.

Zahavi, D. 2015. You, Me, and We: The Sharing of Emotional Experience. *Journal of Consciousness Studies* 22 (1–2): 84–101.

Part II

Emotional Rationality

2

How Emotions Know: Naturalizing Epistemology via Emotions

Cecilea Mun

The topic of intentionality has drawn considerable contemporary attention within the area of philosophy of mind, especially since Daniel Dennett's (1987, 1990) challenge to identify what original intentionality is and to explain how it fits into a naturalistic picture of the world. I refer to this challenge as *the problem of intentionality* and I discerned that it entailed the problem of adequately answering the following four questions:

1. What is the explanandum of original intentionality?
2. What is the genus of original intentionality?
3. What is the differentia of original intentionality?
4. How does original intentionality fit into a natural and hopefully scientific understanding of the world?[1]

---

[1]I say "hopefully" here because providing an adequate scientific theory of original intentionality—a genuine mental life—may be more dependent on the politics of scholarly

---

C. Mun (✉)
Las Vegas, NV, USA
e-mail: cecileamun@icloud.com

27

I will refer to the above interpretation of the problem of intentionality as *Mun's problem of intentionality*, and I will refer to each of the above questions as *an aspect of Mun's problem of intentionality*. In this chapter, I respond to Dennett's challenge by providing an answer (to a certain extent) to the fourth question presented above—the fourth aspect of Mun's problem of intentionality: I argue that we can understand how original intentionality fits into a natural and scientific understanding of the world through an understanding of the import of the intentionality of emotions to our knowledge of the world in which we live.

To do so, I first rely on Anthony Kenny's ([1963] 1966) and Ronald de Sousa's (1987) accounts of the intentionality of emotions in order to highlight the significance of the logical relation between an emotion and its target. I then introduce what I will refer to as *de Sousa's problem of intentionality*, and I relate it to the fourth aspect of Mun's problem of intentionality in order to argue that original intentionality—having a genuine mental life—is implicated in the intentionality of emotions. In doing so, I argue that, given the intentionality of emotions, as evidenced by the fact that we judge the rationality of the ordinary language use of emotion terms, emotions constitute a category of experiences that demonstrate the genuine mental life of emotional beings. In short, emotions demonstrate our original intentionality (i.e., a genuine mental life).[2]

I then move on to explain how, given a framework that regards emotions as being rational,[3] the intentionality of emotions is necessary for us to have knowledge of the world in virtue of our emotional responses: the intentionality of emotions makes knowledge of the world, in virtue of our emotions, possible by endowing

---

theorizing, which, unfortunately, has not always been deferential to truth above everything else. For historical examples, see Feyerabend ([1975] 1988).

[2]Note that this demonstration does not necessarily entail that the subject of an emotional experience has any theoretical knowledge about their emotional experience.

[3]For example, de Sousa (1987), Nussbaum (2001), Solomon (2007), Mun (2016a), and Furtak (2018). Cf. with Prinz (2004). I note this because intentionality alone is insufficient for knowledge or even the possibility of knowledge.

our emotions with knowable content, i.e., by making it possible for our emotions to be vehicles of knowledge. I do so by first relating the two aspects of de Sousa's problem of intentionality (the problem of composition and the problem of complex objects) to two aspects of what I will refer to as *Millikan's problem of knowledge*, showing how a solution to de Sousa's problem of intentionality is also a solution to a restricted version of Millikan's (2017) problem of knowledge. I then focus on one aspect of de Sousa's problem of intentionality (the problem of composition) and one aspect of Millikan's solution to her problem of knowledge (the problem of developing a naturalistic theory of cognition), and in defending de Sousa's solution to his problem of composition, I argue for the claim that emotions are vehicles of knowledge.[4]

In my final section, I connect de Sousa's solution to his problem of complex objects with Millikan's solution to solving the problem of developing a naturalistic theory of what Millikan refers to as "natural information" (her second course to solving her problem of knowledge), and thereby complete my explanation (with the help of de Sousa and Millikan) of how the intentionality of emotions is significant to our knowledge of the world. And, I briefly discuss how, by relating de Sousa's problem of intentionality with both the fourth aspect of Mun's problem of intentionality and Millikan's problem of knowledge, the foregoing explanation provides an answer (or at least a detailed sketch of one) to the fourth aspect of Mun's problem of intentionality. Specifically, it does so by allowing, at the least, the neuroscience of emotion to provide an explanation of how we can know in virtue of our emotional experiences—how epistemology can be naturalized via emotions.

---

[4]Cf. my arguments here, which in some sense can be taken as extending de Sousa's (1987) arguments, with others who have argued for the claim that emotions are vehicles of knowledge (for example, Nussbaum 2001; Solomon 2007; Furtak 2018). I say "in some sense" because there are some significant differences between my view and de Sousa's view, although there are also some considerable overlaps. For example, I do not take desires to be emotions (see 34n).

1   The Intentionality of Emotions

The intentionality of emotions—that emotions are about some aspect of the world (real or imagined[5])—is now a widely accepted fact within the area of philosophy of emotion (see Mun, forthcoming). Both Kenny ([1963] 1966) and de Sousa (1987) have discussed the intentionality of emotions in terms of an emotion's "formal object." For Kenny, the intentionality of emotions, or more so the "intensionality" of emotions,[6] in terms of their formal objects, entailed that there is a logically necessary relation between an emotion type (e.g., anger, fear, sorrow, and joy) and what an emotional experience is about (i.e., the intentional content of an emotion) rather than a merely contingent, causal relation. Thus, given the logical relations between emotions and their intentional contents, Kenny concluded that the intentional contents of emotions were essential to what emotions are (191).

Furthermore, Kenny supported his premise that there is a logical relation between emotions and their intentional contents with analyses of the rationality of the ordinary language use of emotion words. He

---

[5]All my uses of "or," without the use of "either," ought to be taken as mutually inclusive disjunctions. Every mutually exclusive disjunction is indicated by the use of "either, or."

[6]I am simply highlighting here the distinction between "intentionality," spelled with a "t," and "intensionality," spelled with an "s." Kenny's use of the word "intensionality" indicates the underlining connection between his notion of intensionality and the meaning of a linguistic item, which Kenny takes to be essentially mental. The "meaning" of a word, phrase, or sentence is often referred to in the area of philosophy of language as the "intension" of a word, phrase, or sentence (see Putnam 1973), and it is related to the Fregean notion of "sense," which is contrasted with the notion of reference (see Kripke 2011). Comparing Kenny's notion of intensionality with more recent works on intentionality, such as de Sousa's (1987), suggests the idea that talk of intentionality/intensionality need not be associated with or directly associated with the meaning of linguistic items. de Sousa's use of the word "intentionality" stands at the intersection of uses that necessarily entail a relation to the meaning of linguistic items and those that do not (such as an understanding of intentionality in terms of information processing that has no linguistic component; for example, the intentionality of some perceptual experiences). This is primarily because de Sousa, while acknowledging the significance of the meaning of a linguistic item to concerns about intentionality, also, unlike Kenny, considers the significance of the properties of the targets of emotion, and the causal relations between an emotion and its target, in his discussions about the intentionality of emotions. As I suggest at the end of the following section, what explains the shift from Kenny's use of the word "intensionality" to de Sousa's use of the word "intentionality" is de Sousa's interest in naturalizing emotions.

argued that the various ways in which we understood the rationality of our ordinary language use of emotion words placed restrictions on our emotion concepts, and these concepts reflect an essential aspect of what emotions are.[7] Thus, given that we cannot understand the rationality of one's expression of fear or remorse if one's use of the word "fear" or "remorse" violates a rational restriction on what the concept of fear or remorse allows,[8] the intentional content of fear or remorse demonstrates a logically necessary relationship with what fear or remorse is (192).

To further illustrate this point, consider a sincerely uttered sentence like "I am overjoyed that I lost all my money in the recent stock market crash."[9] When taken as a sincere utterance, such a sentence would cause one to question the speaker's rationality, but only if one was not also given a *sensible* reason for the speaker wanting to lose all of their money during a recent stock market crash. Given that one typically regards the use of the English word "overjoyed" as being rational on occasions in which one's desire or appropriate desire is fulfilled, the necessity of (at the least) a logically consistent reason rather than just any reason suggests that the concept of being overjoyed logically restricts the intentional content of the experience of being overjoyed to (at the least) the fulfillment of one's desire or appropriate desire. So, to be overjoyed logically entails (in virtue of what it means, perhaps analytically, to be "overjoyed") that one is (at the least) in a state that is about or directed at the fulfillment of one's desire or appropriate desire.

---

[7]Cf. this discussion about the relationship between emotion concepts and the ontology of emotions, especially the intentionality of emotions, with my discussion of "the lack of metaphysical dependence between the cultural diversity of emotion words, or concepts, and the objective kind status of Emotion" (Mun 2016b, p. 265). In that paper, I was specifically addressing concerns regarding the metaphysical status of emotion (and emotion types/species) as an objective kind whereas what I am speaking of here concerns the import of the rationality of emotions to the intentionality of emotions. One can, however, regard what is stated in this chapter as a clarification or extension of some of my points in my previous paper on the rationalities of emotion (Mun 2016a).

[8]All references to words when used as words are placed in double-quotation marks, although some words placed in double-quotation marks are not intended to indicate references to words but instead indicate a quoted passage. The context should help the reader differentiate between these cases.

[9]This example was inspired by Kenny's example of a man saying that he is afraid of winning £10,000 (192).

In contrast with Kenny, de Sousa (1987) spoke of an emotion's formal object, and therefore its intentionality, in terms of a criterion of success (20; 158–159).[10] According to (R1), de Sousa's first principle of rationality, which he named "Success,"[11] the normative condition under which any representational state, such as an emotional experience, can be judged as being rational is expressed in terms of that representational state's formal object (i.e., its criterion of success) (159). In other words, the fulfillment of a particular emotional experience's condition of success that is expressed in terms of that emotion's formal object, entails, as with Kenny's notion of a formal object, restrictions on at what a particular emotional experience can be directed or be about in order for that emotional experience to be regarded as a rational experience. A criterion of success can therefore be understood as the normative condition that must be fulfilled by the target of an emotional experience not only for that emotional experience to be regarded as the type of emotion that it is (126), but also in order for it to be regarded as a rational experience (158–159). When an emotion's formal object fits the actual object or "target" of an emotional experience, one can conclude, in accordance with de Sousa's principle of Success (R1), that the emotional experience was successful, and therefore rational (20).

Also unlike Kenny, de Sousa conceived an emotion's formal object as a second-order property rather than a logical relation (122). For example, the formal object of fear would be the second-order property of being frightening, which would fit any target with the first-order property of being dangerous (122). Such targets, which are targets of a successful emotional experience, can also be understood, in de Sousa's

---

[10]de Sousa only explicitly noted that he was following C. D. Broad (1954) in his understanding of an emotion's formal object (de Sousa 1987, p. 121), yet a careful reading of de Sousa's (1987) and Kenny's ([1963] 1966) works suggests that de Sousa was also inspired by Kenny. Although the comparisons between Kenny and de Sousa's thoughts presented in this paper suggest this, compare de Sousa's notion of a criterion of success with Kenny's discussion of a criteria for success for the emotion of enjoyment ([1963] 1966, p. 150) for further support for this claim. Also see de Sousa (2018). Searle (1984) and Chalmers (2010) also spoke of something very similar in regard to the intentionality of mental states as "conditions of satisfaction."

[11]de Sousa states this principle, the principle of Success, in the following way: "The formal object of a representational state defines that state's criterion of success, in terms of which the rationality of that state is assessed" (de Sousa 1987, p. 159).

terms, as a "proper target" compared to a *mere* target. A mere target, according to de Sousa, is an actual particular to which an emotion is related or directed whereas a proper target of an emotion is the target of an emotion that would remain that emotion's target even when the subject had full, relevant knowledge of the circumstances of their emotion (116). For example, consider the following case.

A young man in a small town visited the local store with his father, and the young man noticed a homeless man, whom many members of the community considered the "town drunk," going about his routine of purchasing alcohol and asking for change every now and then near the storefront. The young man turned to his father, and with contempt he questioned his father, "What's wrong with that guy? He's such a waste of a person! The world would be a better place without people like that." The father simply observed his son's response, stated that he didn't know, and they both went on their way. Later on, the father and son visited the local store again, and they saw the same homeless man standing near the storefront. The young man again showed his contempt for the homeless man, "What a waste!" But, unbeknownst to the young man, the father had met with the homeless man earlier that week and spoke with him about his life. So, this time, the father responded by telling his son about the homeless man's life: Well you see son, that guy has a disability that makes it difficult for him to succeed in this town, especially because we don't really have many resources here to help him. His parents used to take care of him, but they both passed away a while ago. Since then, no one has taken the time to help him and to teach him what he needs to know in order for him to flourish in this town, or to be his friend.[12] So, that's what's "wrong" with that guy.

---

[12]According to the McKinney-Vento Homeless Assistance Act, as amended by the HEARTH Act (42 U.S.C. §103), "chronically homeless" is defined in terms of the following criteria, "(i) is homeless and lives or resides in a place not meant for human habitation, a safe haven, or in an emergency shelter; (ii) has been homeless and living or residing in a place not meant for human habitation, a safe haven, or in an emergency shelter continuously for at least 1 year or on at least 4 separate occasions in the last 3 years; and (iii) has an adult head of household (or a minor head of household if no adult is present in the household) with a diagnosable substance use disorder, serious mental illness, developmental disability (as defined in section 102 of the Developmental Disabilities Assistance and Bill of Rights Act of 2000 (42 U.S.C. 15002)), post traumatic stress disorder, cognitive impairments resulting from a brain injury, or chronic physical illness or

Now what do you think? The son's response turned from contempt to pity after learning that what was ultimately "wrong" with the homeless man was that he was born the way he was in a community that did not care enough to befriend and help people like him flourish.

In the above example the young man's contempt was directed at the homeless man because the young man believed or perceived that the homeless man had properties that were contemptible: e.g., being a person who is wasting their life away. de Sousa referred to such properties as "focal properties" (de Sousa 1987, p. 116). The young man thought that the homeless man, who was the target of his contempt, had characteristics that made the formal object of contempt fit the homeless man. The homeless man was therefore the target of the young man's contempt. Furthermore, that the homeless man was wasting his life away was, at the least, a significant aspect of what the young man's contempt was about.[13] The young man's contempt, however, changed to pity once he learned some truths about the homeless man. So, the young man no longer had contempt for the homeless man, although he still pitied him, and perhaps also himself and the community he shared with the homeless man. This rational shift in the young man's emotion,[14] from

---

disability, including the co-occurrence of 2 or more of those conditions." Also, according to the April 2014 policy briefing paper, "Discrimination and Economic Profiling among the Homeless of Washington, DC," by the National Coalition for Homelessness, "approximately 93% (132) of the respondents from a sample of 142 respondents experienced discrimination; 70.4% from private businesses, 66.6% from law enforcement, 49.7% from medical services, and 43.7% from social services" (5). For ideas about alternative possibilities, see the Camphill Association of North America website (www.camphill.org).

[13]That the homeless man was wasting his life away was a significant aspect of what the young man's contempt was about rather than being all of what the young man's contempt was about since the young man's contempt also included information about how the homeless man affected the young man's well-being if the homeless man was actually wasting his life away. There may be various ways in which the homeless man can affect the young man's well-being if the homeless man were in fact wasting his life away. For example, the homeless man, in wasting his life away, can be taken by the young man as a contrasting case for the young man's self-understanding. This may be why contempt is often associated with pride, in that pride often motivates contempt, and contempt often reinforces pride. Cf. this account of contempt with the accounts discussed by de Sousa (2019).

[14]The emotional shift from pity to contempt is a rational shift since the shift occurred as a result of the young man maintaining the logical consistency between the information he acquired from

contempt to pity, indicates that the target of his contempt (the homeless man) was not the proper target of his contempt since the young man no longer had contempt for the homeless man once he learned some truths about the homeless man's life. In this case, the homeless man did not remain the target of the young man's contempt. The young man realized that the homeless man actually had characteristics that made the formal object of pity rather than contempt more appropriately fit the homeless man, even if the young man may have also simply experienced this realization as an emotional shift from contempt to pity.

de Sousa's point in differentiating between a mere target and a proper target was to highlight the normative condition under which a target of an emotional experience can be judged with respect to the rationality of that emotional experience, which, according to de Sousa's principle of Success (R1), is determined by the emotion's formal object. Unlike proper targets, which fulfill the normative conditions given by an emotion's formal object, an emotional experience can have a multiplicity of contingent targets (i.e., mere targets) to which it is related or directed. For example, it is possible for one to fear any number of things, including things that one might regard under normal conditions to be very benign. Yet only a restricted set of targets is regarded to be appropriate for a rational experience of fear, and so only a restricted set of targets can be fear's proper target. In other words, an emotion's formal object is thought of as appropriately fitting only a restricted set of targets. Consider again de Sousa's example of fear. The formal object of fear, which is that of being frightening, would fit targets with the focal property of being dangerous, but not targets with the focal property of being disarming. Thus, for de Sousa, an emotion's formal object restricts the set of targets, at which an emotion can be appropriately or rationally directed or be about, to that emotion's proper targets.

We can therefore understand how both Kenny and de Sousa conceived the intentionality of an emotional experience, in terms of a

---

his father and the constraints on rationality placed on him by his emotion's formal object. For a more detailed discussion of the constraints on rationality that can be placed by an emotion's formal object, see Mun (2019a).

formal object, as the formal object placing logical restrictions on the kinds of objects at which a particular emotional experience can be directed or be about, in order to regard that emotional experience as a rational emotional experience. For in doing so, both Kenny and de Sousa emphasized the significance of the logical relation between an emotion and its target. This fact about the logical relation between an emotion and its target, which is expressed in terms of that emotion's formal object, is also a fact about how emotions demonstrate the genuine mental life of emotional beings. It is also a fact about how emotions demonstrate our original intentionality. de Sousa's account of an emotion's formal object, however, differed from Kenny's in at least one important respect: whereas Kenny formulated his notion of intentionality within the ontological framework of mental attitudes as his primary unit of explanation,[15] thereby focusing on the logical relation between an emotion and its formal object, de Sousa explicated his notion of an emotion's formal object in terms of the logical property and causal relations between an emotion and its target, focusing primarily on the appropriate fulfillment, by an emotion's target, of the normative condition that is given by its formal object. The most notable explanation for this difference is that de Sousa, unlike Kenny, was especially interested in the project of naturalizing the intentionality of emotions.[16]

2   The Problems of Intentionality

de Sousa also referred to the problem he was addressing as "the problem of intentionality" (74–75), and he referred to his two formulations of this problem as "the problem of *composition*" and "the problem of complex objects" (75). I will refer to this problem of intentionality,

---

[15]A "unit of explanation" can be generally understood as that through which an explanation is given and understood. Also see Mun (2019b).

[16]As noted in the previous section, Kenny's criticisms of Descartes' and Hume's theories of emotion, as well as Ryle's theory of emotion, was a rejection of a purely causal, and therefore contingent relation, between an emotion and its proper target, and such causal explanations often indicate naturalistic approaches.

which are constituted by his two formulations, as *de Sousa's problem of intentionality*. The problem of composition can be understood as a challenge to provide a materialistic account that can make sense of the observation that our emotional experiences have formal objects; it is therefore a challenge to provide a materialistic account that can make sense of the intentionality of emotions. More specifically, the problem of composition is a challenge for materialistic, functionalist, neuroscientific accounts of emotion, such as Pansepp's (1982) theory of emotion,[17] to provide an account of how a materialistic framework (which appeals to mechanistic, physiological, or neural functions or processes as the primary unit of explanation) can make sense of the logical relation between an emotion type and its proper target, which is expressed in terms of that emotion's formal object and is instantiated by a particular emotional experience of a certain type. Such explanations would, according to de Sousa, "*break the simulation barrier*" (71). de Sousa's point about the existence of a simulation barrier between such materialistic accounts of emotions and mentalistic accounts of emotions (that can readily make sense of the logical relation between an emotion type and its proper target) can be likened to what others in the area of philosophy of mind refer to as the "explanatory gap,"[18] which is the general problem of providing a materialistic explanation that maintains the mentalistic features of a mental phenomenon.[19]

---

[17]See de Sousa (1987, pp. 60–63, 75–76).

[18]Princess Elisabeth of Bohemia is usually identified in the discipline of philosophy as the first person to have conceived the "problem of the explanatory gap," in response to René Descartes' new physics, especially in regard to concerns about the interaction between physical bodies and mental minds. Also see Chalmers (2010) for a related discussion about the problem of the explanatory gap within the contemporary discourse in the philosophy of consciousness.

[19]A similar point is also made by Jackson's (1982) thought experiments. Furthermore, the problem of complex objects may also be understood as the problem of specifying bridge principles that link materialistic, functional accounts to mentalistic, intentional accounts, although this would be the case only under the assumption of an approach to the problem of complex objects that assumes the necessity of something like a unified language for the science of emotion, broadly construed. Note that my proposal for an interdisciplinary science of emotion is not for a unified language of emotion or emotions, but rather for a unified meta-language (a language about theories of emotion) and for the necessity of translators or speakers of many languages (see Mun, forthcoming).

The problem of complex objects is the problem of explaining how, given a materialistic framework, it is possible for emotions to be about anything at all—"how they can have any sort of meaning" (de Sousa 1987, pp. 75–76). It is therefore a challenge to go beyond a solution to the problem of composition by providing a particular account of how a particular solution to the problem of composition can explain the meaningfulness (i.e., the intentionality) of our emotional responses. It is the challenge to provide an explanation of how a materialistic, functional account, which would be posited as a solution to the problem of composition, can in fact be a solution to the problem of composition by requiring such a materialistic, functional account to also include an explanation of how the materialistic, functional aspects of our emotional experiences can account for the meanings of our emotional responses, which is evidenced by the logical relation between an emotion and its formal object, as observed by both Kenny and de Sousa.[20] An answer to this problem of complex objects amounts to what de Sousa referred to as breaking the "*contextual barrier*" (71).

Given de Sousa's two formulations of his problem of intentionality, which are in fact two interdependent aspects of de Sousa's problem of intentionality,[21] de Sousa's problem of intentionality can also be understood as a restricted version of the fourth aspect of Mun's problem of intentionality, which was introduced at the beginning of this chapter: the problem of proving how the notion of a genuine mental life (original intentionality) fits into a natural and (hopefully) scientific explanation of the world. The problem of composition and the problem of complex objects are both variants of the fourth aspect of Mun's problem of intentionality restricted to the domain (i.e., the context) of emotion. Therefore, a solution to de Sousa's problem of composition would constitute one part of an answer to the fourth aspect of Mun's problem of intentionality, and a solution to de Sousa's problem of complex objects

---

[20]If the problem of composition is understood in terms of a problem of providing bridge principles for a unified science of emotion, then the problem of complex objects can be understood as the problem of justifying these bridge principles.

[21]See de Sousa's (1987) discussion of the contextual barrier and its relationship to the simulation barrier (74).

would constitute the second part of an answer to the fourth aspect of Mun's problem of intentionality.

In the next section, I will focus on de Sousa's problem of composition with regard to the fourth aspect of Mun's problem of intentionality, although I will begin by relating de Sousa's problem of intentionality to Millikan's problem of knowledge. I will conclude the section with an argument for the claim that emotions are vehicles of knowledge.

3   Emotions as Vehicles for Knowledge

de Sousa's problem of intentionality, as well as the fourth aspect of Mun's problem of intentionality, is also associated with Kant's question of how knowledge is possible but from a contemporary, naturalistic perspective, as considered by Millikan in her book *Beyond Concepts* (Millikan 2017, p. 3). I will refer to this problem of the possibility of knowledge as *Millikan's problem of knowledge*. This problem can be more precisely understood as the problem of explaining how knowledge is possible given the fact that we are the kind of creatures that we are—creatures that have evolved the capacity to know the world and to express this knowledge through language. Millikan also observed two courses for addressing her problem of knowledge: (1) developing a naturalistic theory of cognition and (2) developing a naturalistic theory of what she referred to as "natural information" (Millikan 2017, p. 109).[22] Both of these courses can be understood as responding to a more generalized version of de Sousa's problem of intentionality, from within the area of philosophy of language.

The first course that Millikan took can be understood as addressing de Sousa's problem of composition, and the second course can be understood as addressing de Sousa's problem of complex objects. Thus, de Sousa's problem of intentionality can also be understood as a version of Millikan's problem of knowledge, which has been restricted to the domain of emotion. A solution to de Sousa's problem of intentionality,

---

[22]What Dretske (1981) referred to as the "flow of information."

which includes the problem of composition and the problem of complex objects, should therefore also provide, at the least, a solution to a restricted version of Millikan's problem of knowledge (restricted to the context of emotion). And, given the conclusion of the preceding section, a solution to de Sousa's problem of intentionality should also provide an answer to the fourth aspect of Mun's problem of intentionality. In the remaining passages of this section, I will focus on de Sousa's solution to the problem of composition and how it can fulfill what Millikan suggested as the first course to a solution to her problem of knowledge: the course of developing a naturalistic theory of cognition.

Toward the end of his chapter, "The Rationality of Emotion," de Sousa argued that "emotions tell us things about the world," and the specific way in which they do so is by being a solution to the philosophers' frame problem (203). This is essentially de Sousa's solution to his problem of composition.[23] The philosophers' frame problem, as discussed by de Sousa (1987, pp. 192–195), can be summarized as the problem of discerning the relevance of available information to solving a particular problem: given that we are supplied with and have access to a profusion of information about the world via our sensory organs, the question is how would beings like us determine the relevance of the available information to solving any particular problem?[24] Furthermore, the intentionality of emotions—specifically, for de Sousa, the fact that emotions have formal objects—is what allows emotions to have the function of directing our attention toward information that is relevant to solving the problems that we face as the kind of beings that we are (i.e., to solving the philosophers' frame problem). As de Sousa's *New Biological Hypothesis 2* (BH2) states, "Emotions are species of determinate patterns of salience among objects of attention, lines of inquiry,

---

[23]Note that this solution to the problem of composition should be understood as being given within the context of a more general explanation of how our emotions, given our individual natures, which are consequences of both our evolutionary and biographical history (phylogenesis and ontogenesis), allow us to know about the world, and that at least one way in which emotions do so is in virtue of their intentionality (de Sousa 1987, p. 203; also see de Sousa 2011, Chapter 3).

[24]For de Sousa's discussion of how the philosophers' frame problem relates to the discussion of frame problems in artificial intelligence, see de Sousa (2011, p. 154).

and inferential strategies" (196).[25] As such, not only are emotions responses to the philosophers' frame problem, they are also vehicles of knowledge—emotions carry information that can be known.[26]

Although de Sousa's solution to his problem of composition seems like an intuitively accurate response, one might argue that certain aspects of de Sousa's view inhibit us from drawing the conclusion that emotions can be understood as vehicles of knowledge, given that they tell us about the world. For example, one might argue that, given the following three claims about emotions, which are all aspects of de Sousa's particular account of emotions (and therefore aspects of his solution to his problem of composition), one cannot infer that emotions are vehicles of knowledge:

1. Emotions as patterns of salience solve the philosophers' frame problem by directing our attention toward relevant information (203).
2. Emotions are not justified (197–198).
3. Emotions need not be conscious (149–150).

First, one might argue that given de Sousa's view, emotions cannot be "vehicles" of knowledge (such as, perceptions and beliefs), although they can be a necessary condition for knowledge since, as patterns of salience, they merely facilitate the acquisition of knowledge by solving the philosophers' frame problem. Thus, one might conclude that, for de Sousa, emotions cannot be *vehicles* for knowledge since patterns of salience cannot be vehicles of information although they can organize vehicles of information, such as vehicles of knowledge. The significance of this criticism is to bring out a point about how one ought to

---

[25]Also see de Sousa's (1987) BH1, *New Biological Hypothesis 1* (195).

[26]Note that being a vehicle of knowledge does not entail that what is represented in virtue of the vehicle of knowledge (e.g., the contents of an emotion) is necessarily known. A vehicle of knowledge simply makes it possible for the content of the vehicle to be known. The vehicle of knowledge is not what makes the content of the vehicle known. It simply allows the content of the vehicle to be known. Furthermore, not all vehicles of representation are vehicles of knowledge. For example, street signs are vehicles of representation, but they are not vehicles of knowledge since they are not states of epistemic beings.

understand what it means for something to be a "pattern of salience" and a "vehicle of knowledge."

Within the discourse on intentionality, especially within the areas of philosophy of mind and emotion, *vehicles* of knowledge are generally understood to be some kind of *vehicle of representation*, which ought to be distinguished from what has been referred to as a "mode of presentation."[27] A vehicle of representation can be understood as that which carries information. For example, thoughts, beliefs, perceptions, physiological states (of all kinds), as well as physical states, such as rocks, clouds, surface reflections, and light can all be understood as vehicles of representation. Millikan refers to such vehicles of representation as "infosign vehicles" (Millikan 2017, p. 110). Thus, the main point behind this criticism is that emotions, as patterns of salience, cannot be vehicles of knowledge since they cannot be vehicles of representation. One might therefore conclude that de Sousa's solution to his problem of composition cannot also work as a solution to Millikan's problem of knowledge since there is a problem of understanding how patterns of some kind (such as patterns of salience) can be vehicles of some kind (such as vehicles of knowledge).

The problem with this criticism is that it fails to recognize the fact that emotions, or any other phenomenon, can be understood from a multitude of perspectives. Therefore, given the fundamental assumption that *any object of inquiry can be understood from various perspectives, and that each perspective may yield a distinct view, theory, or language*, it follows that any truths of any particular perspective necessarily entails not only that these truths can be true, but also that they are true, from all other perspectives, insofar as all the relevant perspectives are perspectives about the same phenomenon.[28]

---

[27]See Chalmers (2010, p. 362). Generally speaking, modes of presentation are ways in which one can understand, conceptualize, or conceive a referent of a term. As such, it would be more accurate to liken modes of presentation to an emotion's formal object or intentional content rather than liken modes of presentation to vehicles of representation.

[28]See de Sousa (2011), Chapter 2, "What Emotions Have to Say," for a list and discussion of possible truths from various perspectives about what emotions are (26–44). Also, due to some ambiguities in his text and some of his comments, which were personally conveyed to me, it's not clear to me whether de Sousa would agree with this defense of his solution to his problem

Given that emotions have a logical relation with their formal objects, the challenge then for materialistic, functional perspectives, and any other relevant perspective, in regard to de Sousa's problem of composition, is to provide an account of how such established truths are true from their own perspective. Given this, bridge principles are unnecessary for a solution to de Sousa's problem of composition since it is the object of inquiry that ultimately unites the truths of every perspective. What is necessary, however, is the recognition that these perspectives are all about the same phenomenon, which is the shared object of inquiry.[29] Accordingly, emotions, as patterns of salience, not only can be, but also are vehicles of representation, as long as one is willing to accept that an understanding of emotions as patterns of salience and an understanding of emotions as vehicles of representation are both understandings of what emotions are.

I refer to the fundamental principle with which the foregoing conclusion was drawn as the (αω) *alpha-omega principle of interdisciplinary enquiry*:

> Alpha-Omega Principle of Interdisciplinary Enquiry (αω): Any object of enquiry can be understood from various perspectives, which may each yield a distinct view, theory, or language, and the truths of each perspective are ultimately united by the shared object of enquiry.

It is the alpha principle in that it is the preliminary assumption with which any adequate interdisciplinary research ought to begin, and it is the omega principle in that it is the final conclusion that ought to be inferable from the complete, totality of the results of any adequate interdisciplinary research. Given the alpha-omega principle of

---

of composition. But if he does not, one can consider that this is at least one point at which de Sousa's view and mine diverge.

[29]The validity of this argument can be understood with the use of a Venn diagram. If we grant, from one perspective, that it is true that all emotions are patterns of salience, and we grant, from another perspective, that it is true that emotions are vehicles of knowledge, then it will necessarily be the case that all emotions as patterns of salience are emotions as vehicles of knowledge, and vice versa, as long as both perspectives share the same object of inquiry—emotion or an emotion.

interdisciplinary inquiry, one ought to be able to understand that emotions as patterns of salience can also be vehicles of representation, and therefore can be vehicles of knowledge, especially if one considers the fact that the object of inquiry, which unifies these two perspectives, *is experienced*,[30] regardless of whether or not the subject is *aware* of having such an experience.[31]

Second, even if one grants that emotions can be vehicles of knowledge, one might argue against the application of de Sousa's solution to the problem of composition to Millikan's problem of knowledge by arguing that although emotions can "tell" us about the world, what they tell us is not something that is known since emotions are not justified and since emotions are not conscious. The problem with this convergent criticism is that they fail to consider the possibility of an externalistic, reliablist notion of justification, such as the notion of *warrant*,[32] which need not require a subject to know that they know (and therefore does not require subjects to be aware of their knowing) in order to know. Thus, if we accept that "knowledge" can be something that can be warranted rather than justified, and that a subject of knowledge need not be *aware of knowing that they know in virtue of their emotions* in order to *know in virtue of their emotions*, then emotions (as experiences *versus* experiences of which one has a categorical or discrete awareness), even as patterns of salience, can be vehicles of knowledge.

For example, experiences of fear, anger, sorrow, joy, love, etc., can provide us with knowledge about aspects of the world that bear on our well-being,[33] and the way in which they can do so, in virtue of their

---

[30]By "experienced" I mean that there is something it is like for a subject of an emotional experience to have such an experience, although the subject need not be aware that they are having such an experience. For example, a person may have an experience of fear without being aware that they are having an experience of *fear*. They may simply have an experience of freezing without being aware that this experience is an experience of fear. Cf. Chalmers' (1995) discussion of experience, consciousness, and awareness. I avoid using the word "consciousness" in my explanations in order to stay away from introducing any unnecessary complications.

[31]See de Sousa (1987, pp. 149–150), on phenomenology. I would also like to thank James A. Russell for helping me understand the distinction between an experience and one's awareness of that experience.

[32]See Prinz (2004) for a discussion of warrant with respect to emotions.

[33]See Nussbaum (2001), Prinz (2004), and Mun (forthcoming) regarding the intentionality of emotion and how it is constituted by a relation between aspects of the world and our well-being. de Sousa, however, denies that emotion, as a class, has a formal object (1987, p. 20), although he

intentionality, is by directing our attention to those aspects of the world that are significant to our well-being. This directing of our attention is the instantiation of patterns of salience—it is the instantiation of an emotion's formal object, in our experience of that emotion, in virtue of the recognition that a target has a relevant focal property. When one has compassion toward another, the experience of their compassion tells them about an aspect of their world and how it relates to their well-being. It tells them that the world can be unjust by drawing their attention to the relevant information that helps them understand how the world can be unjust. Through my compassion, as patterns of salience, I attend to your unjust pain and suffering, and I can learn and know from this experience—from my experience of attending to your unjust pain and suffering—that, among other things, the world can be unjust (although it may not need to be).

Given the above, one can understand how de Sousa's solution to his problem of composition also fulfills Millikan's course of providing a naturalistic theory of cognition. In other words, one can understand how de Sousa's solution to one of the aspects of his problem of intentionality also works to fulfill one of the courses suggested by Millikan to solve her problem of knowledge (i.e., her problem of explaining how knowledge is possible for creatures like us).

I now turn to connecting the second aspect of de Sousa's problem of intentionality (the problem of complex objects) with Millikan's second course to solving her problem of knowledge (the course of providing a naturalistic theory of natural information) in order to complete my solution (or a sketch of one, with the help of de Sousa and Millikan) to the fourth aspect of my problem of intentionality (the problem of explaining how original intentionality—a genuine mental life—fits into a natural and scientific understanding of the world). And, I conclude with a brief discussion of how emotions as vehicles of knowledge, which is made possible by the intentionality of emotions, allows the

---

does not necessarily deny that every emotion involves a relation between aspects of the world and our well-being as indicated by my discussion in this passage.

neuroscience of emotion to provide an explanation of how we can know in virtue of our emotional experiences—how epistemology can be naturalized via emotions.

4   Naturalizing Epistemology via Emotions

According to de Sousa, the rationality of human emotions is *sui generis* (de Sousa 1987, p. 203).[34] Emotions, as patterns of salience, run outside the standards of rationality against which the rationality of judgments, perceptions, and functional desires are or ought to be judged,[35] yet they do not work independently of judgments, perceptions, and functional desires.[36] Although *their rationality is independent of* the rationality of judgments, perceptions, and functional desires, *emotions work in conjunction with* judgments, perceptions, and functional desires in order to solve the philosophers' frame problem.[37] de Sousa's (1987) explanation of this *sui generis* nature of the rationality of our emotions, which is inclusive of evolutionary, biological, ecological, cultural, and individually idiosyncratic explanations (see Chapter 4), constitutes his solution to his problem of complex objects. de Sousa expresses this explanation through principles relating types of objects to emotions and the causes of emotions (see Chapter 5), and these principles can also be understood as constituting a solution to Millikan's problem of providing a naturalistic explanation of what Millikan referred to as "natural information." de Sousa's solution to his problem of complex objects is

---

[34]Also note that for de Sousa (2011), "mere desires" are "degenerate or zero-level cases of emotions" (101). de Sousa's and my account can be said to be distinct in at least this one respect. I do not regard desires to be emotions, although I regard them to be significant components of at least some emotions.

[35]de Sousa (1987, p. 203). Note, however, that according to more recent comments personally given to me by de Sousa (December 2018), he would no longer include perceptions in this statement. According to de Sousa, when he spoke of emotions as being perceptions he was using perception as an analogy and not an identification, and part of his reason for not intending to make an identity claim was his belief that perceptions cannot be irrational although emotions can be.

[36]Cf. with Mun's (2016a) observation that "emotion or an emotion as a superordinate inference rule runs outside considerations of traditional or standard logical systems that dictate how assessments of warrant, rational thoughts, and rational judgments are to be evaluated" (54).

[37]See de Sousa (1987, pp. 194–202).

therefore also a solution to Millikan's second aspect of her problem of knowledge. Thus, we can understand how the intentionality of emotions allows us to know the world through our emotional responses.

Previously, I argued that if human emotions are at times rational and at times irrational, then there is a normative standard against which our emotional experiences can be judged as being rational or irrational (Mun 2016a, p. 50). This conditional expresses what I referred to as the criterion for the ontological rationality of emotion (COR^e), which states that there exists some normative standard that is given by what emotion or an emotion is (the type of emotion that it is) against which our emotional responses can be judged or evaluated in virtue of the fact that our emotions manifest our rationality, i.e., the capacity for being both rational and irrational (Mun 2016a, p. 51). Furthermore, I explained that what fulfills the criterion for the ontological rationality of emotions is what emotions are as *superordinate inference rules*, and it is this aspect of emotions that is ultimately under investigation by researchers in the field of emotion (Mun 2016a, p. 54). Thus, there is, as of yet, no definitive answer as to *what* emotions *are* as superordinate inference rules. Regardless of this, I also observed that the essential nature of an emotion as a superordinate inference rule is "that which is intended by speakers of some language, especially ordinary or natural languages, to be tracked by or reflected by the intentional content of their use of relevant emotion term(s), i.e., the intentionality of emotion or an emotion type" (54). de Sousa's principles for solving his problem of complex objects can also be understood as expressing what I referred to as "emotions as superordinate inference rules."[38] We can therefore conclude that emotions have their own *sui generis* "canons" of rationality, which express what emotions are as superordinate inference rules.[39]

---

[38]Note that I am not suggesting that I am adopting de Sousa's expressions of superordinate inference rules as my own. For an example of my proposal for the superordinate inference rule of shame, see Mun's (2019b).

[39]It is important to note here that the foregoing ought not to be understood as an explication of de Sousa's (1987) view about the rationality of emotion, but instead as an explication of my view regarding the rationality, intentionality, and epistemic import of emotions, which builds on some of the foundations laid by de Sousa and Millikan. To be sure, there are some considerable overlaps between my view and de Sousa's view, however, there are also some significant differences between our two views, some of which I have noted throughout this chapter.

These canons of rationality (at the least) in some way express the ways in which human minds operate, and emotions as superordinate inference rules are the primary shared objects of inquiry for interdisciplinary neuroscience of emotion.[40] These canons of rationality express the natural laws that govern the operation of human minds. As such, insofar as the neuroscience of emotion aims at discovering the natural laws that govern human minds, one can conclude that the enterprise of the neuroscience of emotion is also concerned with providing an account of *how* emotions can know. As such, adequate theories from the neuroscience of emotion should therefore be able to adequately explain, to a reasonable extent, how emotions can know (or, more precisely put, how it is possible for human beings to know in virtue of our emotional experiences).[41] And, the same may be said *mutatis mutandis* for other areas of study that aim to identify the natural laws of human minds.

---

[40]This claim does not entail that emotions as superordinate inference rules only constitute aspects of human emotions (see Mun 2016a), although my conclusion here is restricted to human emotions.

[41]For an example of a proposal of a neuroscientific approach that may be able to sufficiently do so, see Adolphs and Andler (2018).

References

Adolphs, Ralph, and Daniel Andler. 2018. Investigating Emotions as Functional States Distinct from Feelings. *Emotion Review* 10 (3): 191–201.

Broad, C. D. 1954. Emotion and Sentiment. *Journal of Aesthetics and Art Criticism* 13 (2): 203–214.

Chalmers, David J. 1995. Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies* 2 (3): 200–219.

———. 2010. *The Character of Consciousness*. Oxford: Oxford University Press.

Dennett, Daniel C. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press/Bradford Books.

———. 1990. Evolution, Error and Intentionality. In *The Foundations of Artificial Intelligence: A Sourcebook*, ed. D. Partridge and Y. Wilks, 190–211. Cambridge, UK: Cambridge University Press.

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

———. 2011. *Emotional Truth*. New York: Oxford University Press.

———. 2018. The Rationality of Emotion: Biology, Ideology and Emotional Truth. Interview by Caroline Price, *Emotion Researcher*. http://emotionresearcher.com/the-rationality-of-emotion-biology-ideology-and-emotional-truth/. Accessed November 26, 2018.

———. 2019. Is Contempt Redeemable? *Journal of Philosophy of Emotion*. Forthcoming.

Dretske, Fred I. 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Feyerabend, Paul. [1975] 1988. *Against Method: Outline of an Anarchist Theory of Knowledge*. New York: Verso.

Furtak, Rick A. 2018. *Knowing Emotions: Truthfulness and Recognition in Affective Experiences*. New York: Oxford University Press.

Jackson, Frank. 1982. Epiphenomenal Qualia. *The Philosophical Quarterly* 32 (127): 127–136.

Kenny, Anthony. [1963] 1966. *Action, Emotion and Will*. London: Routledge and Kegan Paul.

Kripke, Saul A. 2011. Frege's Theory of Sense and Reference: Some Exegetical Notes 1. In *Philosophical Troubles: Collected Papers*, vol. 1. Oxford: Oxford University Press.

Millikan, Ruth G. 2017. *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford: Oxford University Press.

Mun, Cecilea. 2016a. The Rationalities of Emotion. *Phenomenology and Mind* 11: 48–57. https://doi.org/10.13128/phe_mi-20105.

———. 2016b. Natural Kinds, Social Construction, and Ordinary Language: Clarifying the Crisis in the Science of Emotion. *Journal of Social Ontology* 2 (2): 247–269.

———. 2019a. Rationality Through the Eyes of Shame: Oppression and Liberation via Emotion. *Hypatia*. Forthcoming.

———. 2019b. Unification Through the Rationalities and Intentionalities of Shame. In *Interdisciplinary Perspectives on Shame: Methods, Theories, Norms, Cultures, and Politics*, ed. C. Mun. Lanham, MD: Lexington Books. Forthcoming.

———. Forthcoming. *Interdisciplinary Foundations for the Science of Emotion: Unification Without Consilience*. Lanham, MD: Lexington Books.

National Coalition for the Homelessness. 2014. *Discrimination and Economic Profiling Among the Homeless in Washington, DC.* http://nationalhomeless.org/wp-content/uploads/2014/04/DiscriminationReport20141.pdf. Accessed December 1, 2018.

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

Panksepp, Jaak. 1982. Toward a General Psychobiological Theory of Emotion. *The Behavioral and Brain Sciences* 5 (3): 407–422.

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

Putnam, Hilary. 1973. Meaning and Reference. *The Journal of Philosophy* 70 (19): 699–711.

Searle, John R. 1984. Intentionality and Its Place in Nature. *Synthese* 61: 3–16.

Solomon, Robert. 2007. *True to Our Feelings: What Our Emotions are Really Telling Us*. New York, NY: Oxford University Press.

3

What Can Information Encapsulation Tell Us About Emotional Rationality?

Raamy Majeed

1 Introduction

The last few decades of emotion research in philosophy, and to an extent in cognitive psychology, has involved a full-blown assault on what Solomon dubs the "The Myth of the Passions": the treatment of emotions "as irrational forces beyond our control, disruptive and stupid, unthinking and counterproductive, against our 'better interests', and often ridiculous" (1977, p. 106). Emotions, it has been argued, and with some qualifications empirically demonstrated, play a role in reasoning, aid us in action, are to a certain extent under our control, and exhibit intentionality. These factors, especially the role emotions play in reasoning, are also said to make them, *contra* orthodoxy, rational. Or more carefully, emotions are things that can contribute, in a positive way, to whether their bearers are rational.

R. Majeed (✉)
Department of Philosophy, University of Auckland,
Auckland, New Zealand
e-mail: r.majeed@auckland.ac.nz

One way of arguing for this conclusion draws on the cognitive architecture of the mind. That is to say they concern ways of carving up cognitive processing theorised by cognitive science, and have their roots in computational theories of the mind. Such accounts are interesting in two respects. First, ways of modelling cognitive architecture in cognitive science tend to be both empirically informed and empirically tractable; they make claims about ways of carving up cognitive processing that, at least in theory, can be empirically tested. What is to be gained by such accounts, therefore, are empirically-driven descriptions of the precise way emotions assist reasoning, which can in turn confirm or disconfirm our initial a priori speculations about emotional rationality. Second, and relatedly, if cognitive architecture has a bearing on emotional rationality, we can't expect to have a complete account of emotional rationality simply by armchair theorising because how cognitive processing is actually structured will turn out to be an empirical issue.

These features make an investigation into an account of emotional rationality grounded in cognitive science philosophically pressing. But they do so only on the proviso that cognitive architecture actually has a bearing on emotional rationality. A possible way it does is hypothesised by de Sousa (1987). This hypothesis draws directly on a key concept of cognitive science, viz. information encapsulation. Very roughly, a system is informationally encapsulated if the function it computes is insensitive to information stored in other systems, especially systems which encode high-level representations like beliefs and desires. Fodor (1983) takes information encapsulation to be the "essence" of modularity, and argues most mental activity besides central cognition, e.g. perceptual processing, language processing and motor-control, are informationally encapsulated. For instance, he argues visual processing is modular given that visual illusions, like the muller-lyre illusion, persist despite our explicit knowledge to the contrary. In the muller-lyre illusion, we continue to see the lines as being of different lengths even when we come to discover that they are actually of the same length.

de Sousa's hypothesis is this: "the role of emotions is to supply the insufficiency of reason by imitating the encapsulation of perceptual modes" (1987, p. 195). The idea, briefly, is when we consider

a response-option to a given situation, reason alone can't determine which information is relevant from the vast store of information which the agent knows, as assessing whether each piece of information is relevant would be too cognitively demanding. Emotions assist reasoning by pre-highlighting certain pieces of information as relevant or "salient". This is achieved by emotions "mimicking" the information encapsulation of perception, i.e. by emotion processing being temporarily insensitive to information stored outside certain channels, e.g. channels to do with belief.[1]

This hypothesis can be challenged on two fronts. First, on whether emotion processing can be informationally encapsulated, even temporally. Second, assuming it can be, on whether it can still, nevertheless, play the role de Sousa ascribes to it in emotion-driven reasoning. As it stands, there is empirical evidence to suggest that emotion processing can, sometimes, be informationally encapsulated. That is, multipathway models of emotion processing are confirmed at a neurobiological level, where some emotion processing is found to occur "off-line" without considering what the agent already knows.[2] A careful examination of the second challenge, however, has not been taken up. This paper aims to make up for this neglect by addressing the question, What can information encapsulation tell us about emotion-driven reasoning? In other words, What can information encapsulation tell us about emotional rationality?

In what follows, I provide an exposition of de Sousa's hypothesis, and then explain why it is misleading. In particular, I argue information encapsulation is not essential for emotion-driven reasoning, as emotions can fulfil their role of highlighting certain pieces of information as salient without being informationally encapsulated (§2). However, I argue

---

[1]The role attributed to emotions here is similar to that posed later by Damasio (1994). Though, crucially, Damasio doesn't rely on information encapsulation, nor any other notion from cognitive science. See Brady (2013: §1.2) for a brief overview of accounts that postulate a relation between emotion and salience more generally.

[2]Multi-level theories of emotion processing have been proposed by Leventhal (1979), Barnard (1985), LeDoux (1996), amongst others. See Teasdale (1999) for a review.

information encapsulation still proves relevant for emotional-driven reasoning in that it ensures (i) emotions can fulfil this role in an efficient manner, and (ii) the fulfilment of this role won't, insofar as emotions are encapsulated, be overridden by what the agent knows (§3).

2   How Emotions Assist Reasoning

What can information encapsulation tell us about emotional rationality? de Sousa, as a matter of fact, provides six principles of rationality, each of which lends itself to an independent account of how emotions can be rational. The account of rationality that has a bearing on our question, however, only concerns his principle of *strategic rationality*. Rationality, it is assumed here, is a teleological concept: anything can be rational to the extent that it fulfils its function. The function of emotion, broadly speaking, is to guide us in reasoning. But more specifically, their biological function is to do something reason can't, viz. to determine the salience of features of perception and reasoning. The information encapsulation of emotions on this picture bears on emotional rationality in that it plays a role in how emotions fulfil this function.

   An exposition of this picture is provided in the context of how emotions can help solve what de Sousa calls "the philosophers' frame problem":

> [W]e need to know when not to retrieve some irrelevant information from the vast store of which we are possessed. But how do we know it is irrelevant unless we have already retrieved it? I proffer a very general biological hypothesis: Emotions spare us the paralysis potentially induced by this predicament by controlling the salience of features of perception and reasoning; they temporarily mimic the information encapsulation of perception and so circumscribe our practical and cognitive options. (de Sousa 1987, p. 172)

Before unpacking this picture, it is worth stressing that the original frame problem has its roots in artificial intelligence. This problem has

been taken up and formulated in several distinct ways,[3] all of which, though related, arguably only bear a loose resemblance to the problem identified above. Evans (2002), for instance, goes so far as to say that de Sousa's discussion of the above problem *qua* frame problem is a red herring. In order to avoid muddying the waters, the scope of this paper will be restricted to *the philosophers' frame problem*, i.e. the one outlined above, as opposed to the plethora of related problems discussed under the umbrella of "the frame problem".

Even with our scope restricted in this way, the problem, as well as the proposed solution to it, still requires interpretation along several distinct lines; some of which will turn out to be relevant to an exposition of the role emotions play in assisting reasoning broadly construed, and others more specific to how their encapsulation enables them to play this role. I begin by discussing interpretations along the first of these lines.

The problem, in its broadest form, is a problem about how to restrict the amount of information to be computed for us to get an appropriate response-option. de Sousa himself takes it as a constraint of the problem that we not only restrict the amount of information, but that we are also able to restrict information to those that are relevant or salient: "No logic determines *salience*: what to notice, what to attend to, what to inquire about" (1987, p. 191).[4] It is here that emotions come in by fulfilling their biological function of determining salience. But why pure reason alone can't determine salience isn't entirely clear. Spelling this out is crucial for getting clear on the precise role information encapsulation is supposed to contribute to this picture.

There appears to be at least three factors relevant to why pure reason can't tell us which response-option to consider, and thereby giving rise to the philosophers' frame problem. de Sousa himself sometimes talks as if the information that needs to be restricted is that which the organism already knows, i.e. how we determine what's relevant "from

---

[3]Chow (2013), for instance, discusses six different versions of the frame problem.

[4]de Sousa's (p. 194) exposition of why the frame problem isn't the problem of induction also makes explicit that the frame problem he has in mind is one concerning which information is relevant.

the vast store of which we are possessed" (p. 172). He elaborates, "The frame problem arises only when we consider what to do with information interpreted and stored in an intentional system" (p. 195). If the information we need to determine as salient is information the agent already knows, there are two possibilities as to why reason alone can't determine which information is relevant. First, it would simply be too cognitively demanding to individually figure out whether each piece of information is relevant; and that remains so even if we only consider those available to the agent. Second, even if such a task were cognitively possible, it remains unclear whether reason alone can determine which pieces of information are relevant under the time-constraints required to respond to certain situations.

To elaborate, some discussions of the frame problem build in a time-constraint to determining which pieces of information are relevant. For example, in Dennett (1984) we find the tale of a robot who when faced with a ticking bomb whiles away precious time considering everything it knows. This constraint is not explicit in de Sousa's version of the problem, but it stands to reason that some real-life cases will come with significant time-constraints, e.g. figuring out the best response-option when faced with a mugger. Fodor (1987, p. 26) calls this "Hamlet's problem: How to tell when to stop thinking". If reason can only determine which information is relevant by considering each piece of information individually, reason alone can't determine which pieces are relevant under such time-constraints, i.e. except sometimes by random. So here we have another factor that explains why reasoning alone can't solve the philosophers' frame problem, at least when we consider real-life instances of where the problems arise.

Now consider a different way of interpreting the information that needs to be restricted. This stems from a related, though, distinct formulation of the philosophers' frame problem: "Assume all the powers already listed — logic, induction, and more-than-encyclopaedic knowledge: the *philosophers' frame problem*, roughly, is how we make use of just what we need from this vast store, how not to retrieve what we don't need" (de Sousa, p. 193). On this way of understanding what is at issue, the problem is not just how to restrict propositional knowledge

the agent already knows, but also all the inferences they can draw, and all the response-options they can consider. (So more accurately, the situation is such that we need to not just restrict the information over which to compute, but the computations themselves as well.) Reason alone can't restrict information *thus construed* because of the two afore-mentioned factors. It would simply be too cognitively demanding to draw each inference before assessing whether it is relevant, not to mention this would take an unrealistically long amount of time. de Sousa himself considers the inferences drawn to be "from a potential infinity" (p. 195). If this turns out to be the case, even setting aside the previous two factors, it would be straight out impossible for reason to determine which inferences, from a set of infinity, are relevant to a given response-option. This is the third factor that accounts for why reason alone can't determine salience.

The above discussion is telling not only because it helps us get clear on exactly what gives rise to the philosophers' frame problem, but because this in turn lets us see more clearly what is required to solve it. The role of emotions, recall, is to supply insufficiency of reason. For de Sousa, this is achieved by emotions mimicking the information encapsulation of perception. What the above ways of understanding the factors which give rise to the philosophers' frame problem bring out, however, is if information encapsulation is relevant to solving the problem, it is relevant only to the extent that it helps determine the salience of information. This is because on this picture, the role of emotions, *in essence*, is to supply the insufficiency of reason by determining the salience of information.

So how exactly do emotions determine which pieces of information are relevant? On certain ways of understanding emotions, it is in their very nature that they determine the salience of patterns of perception and reasoning—and crucially, they do so regardless of whether emotion processing is sometimes informationally encapsulated. There are two features, ones we uncontroversially ascribe to emotions, which work in conjunction to explain how they determine salience. First, emotions have an affective component, and the types of affective components they do have marks them out as having a valence. The idea, very

roughly, is emotions feel good or bad; they appear positive or negative.[5] Second, emotions are not only representational, but they have an evaluative component: they assign a positive or negative value to their intentional objects, i.e. what they are about or in response to.[6] The fact that emotions have an affective component with a valence explains why their representational component is evaluative. In brief, emotions feel good or bad, and in virtue of that, their intentional objects are represented as also being good or bad; as being positive or negative. This helps explain why emotions bias certain pieces of information over others. We are drawn to pieces of information represented as positive or negative whilst neglecting ones that aren't represented in either of these ways.

Crucially, emotions can determine the salience of information whether or not they are informationally encapsulated. To be informationally encapsulated is for our emotion generation systems to be insensitive to top-down influence, e.g. from propositional knowledge already possessed by the agent. Emotions can bias certain pieces of information whilst being insensitive in this way. This is confirmed by multi-pathway models of emotion generation at the neurobiological level. Emotional responses, especially fear responses, can be triggered by stimuli "off-line" without activating the cortex, and thereby without the agent being consciously aware of the stimuli. Such emotion processing is said to occur without any top-down influence from the agent's beliefs. To the extent that such emotions have an evaluative component, they can bias certain response-options over others. However, emotional responses, like those involved in fear, can also be triggered "on-line", and in a way that is susceptible to top-down influence. Such emotional responses result from neural circuits that activate the cortex, consciously register the stimuli, and are influenced by background information possessed by the agent.[7] Emotions triggered in this way, to the extent that

---

[5]See Colombetti (2005), Prinz (2010), and Carruthers (2018) for overviews of emotional valence.

[6]The evaluative nature of emotions is discussed by Solomon (1976), de Sousa (1987), Greenspan (1996), amongst others.

[7]This is proposed by several multi-level theories of emotion generation, and confirmed for fear generation at the neurobiological level by LeDoux (1996).

they have an evaluative component, can also successfully bias certain response-options over others.

In summary, a careful examination of the possible factors which give rise to the philosophers' frame problem, as well as de Sousa's proposed response to it, puts emotional salience at the heart of how emotions solve this problem. Emotions assist reasoning, and *ergo* make their bearers rational, insofar as they determine patterns of salience. But what's more, we now see that we can provide an explanation of how emotions determine salience in virtue their very nature *sans* any further claims about emotion processing being informationally encapsulated. This makes the hypothesis under investigation—"the role of emotions is to supply the insufficiency of reason by imitating the encapsulation of perceptual modes" (de Sousa 1987, p. 195)—mysterious. Precisely what is it that the temporary encapsulation of emotions contributes to how they determine the salience of features of perception and reasoning, and thereby assist reasoning?

3   The Role of Information Encapsulation in Emotion-Driven Reasoning

3.1   The Efficiency of Response-Options Selections

As I see it, there are two major ways the information encapsulation of emotion processing contributes to emotion-driven reasoning. In neither way does information encapsulation *determine* the salience of features of perception and reasoning. Instead, encapsulation makes a contribution to how such salience is determined, and in a way that is conducive to how emotions help solve the philosophers' frame problem. In the first instance, encapsulation helps emotions determine the salience of information in an efficient manner. There are two facets to how it does so, one of which concerns speed, and the other cognitive cost.

The need for speed speaks directly to the philosophers' frame problem construed as one having a time-constraint, i.e. we assume the problem comes with Hamlet's problem built in. One benefit of a system

being informationally encapsulated, and thereby ignoring background information, is we can focus and process in-coming information in a quick and efficient manner. As Fodor notes, "speed is purchased for input systems by permitting them to ignore lots of facts" (1987, p. 70). It might, therefore, be assumed that emotional processing being informationally encapsulated is required to solve this particular version of the frame problem.

It is plausible that encapsulation is required to solve the frame problem under certain time-constraints. Seeing a snake on a hike-trail or wearing off the path of an oncoming car, for example, are cases where an emotional response trigged bottom-up would be beneficial, and precisely for guaranteeing the speed of an appropriate response. But top-down effects can also be quick, and it isn't clear whether all instances of Hamlet's problem, or even most, require emotion processing to be encapsulated. Take Dennett's example. It is hard to see why it would be necessary for the robot's system to be encapsulated for it to escape the ticking bomb. The processing that generates its response-options might be susceptible to top-down influence, but it may still select a response-option within the required time. (In fact, it is highly likely that the selection of an optimal response-option in this situation requires top-down influence, as the robot will need to recall its knowledge of possible escape routes, its knowledge that bombs are hard to diffuse etc.) Not all instances of the frame problem, even when they arise with significant time-constraints, then, seem to require the super quick reflex-like response-options that can only be triggered by emotional responses being encapsulated.

The take-home message shouldn't be that the speed at which response-options are selected is unimportant, nor that encapsulation doesn't matter for these purposes. Rather, the point is encapsulation does make us select response-options quickly, but the specific speeds at which only it can help us do so is only necessary in a limited range of instances—even within the subset of frame problems seen through the lenses of Hamlet's problem.

While the facet of speed only matters in a limited range of instances, the contribution encapsulation makes to minimising cognitive costs

plausibly applies to all instances of the philosophers' frame problem. For any given instance where there is a range of response-options to choose from, background information may bear on selecting the best possible options, and in a timely fashion required by the specific instances. Nevertheless, were it possible to select the relevant options without considering such information, we can resolve frame problems without the cognitive costs acquired when we consider such information. A potential benefit of this is agents are better able to allocate their cognitive resources to other tasks, e.g. how to best follow through with the response-options they have selected.

It is important, however, not to overestimate how much cognitive cost is actually minimised when emotional processing is encapsulated. Encapsulated systems are insensitive to information stored in other channels. When perceptual processing is encapsulated, for example, this is taken to mean background information, e.g. what the agent believes, does not influence what we perceive. By contrast, when this process is claimed to be "cognitively penetrated", the claim is not that what the agent actively considers affects what they perceive; though it very well might. Instead, the claim is the agent's beliefs affect their perception regardless of whether they actively consider these beliefs. In this way, any cognitive costs to the agent that are spared on account of emotion processing being encapsulated needn't involve anything resembling deliberation, or even any conscious activity for that matter. Subsequently, emotion-driven reasoning, when emotion processing is encapsulated, no way ensures any significant gains when it comes to the minimisation of cognitive costs.

Both facets discussed above play a role in how emotions determine the salience of features of information, though the role they play is minimal and in most cases inessential. The next way encapsulation of emotion processing contributes to emotion-driven reasoning will prove far more crucial. This can be brought out in the context of why information encapsulation was thought relevant for the original frame problem(s) in artificial intelligence, and why it fails to play the same role when it comes to the philosophers' frame problem.

3.2   The Maintenance of Response-Option Selections

The first discussion of how information encapsulation bears on the frame problem was discussed by Fodor (1983, 1987, 2000). In his discussion, Fodor argues that frame problems don't arise for informationally encapsulated systems because there is only a small amount of information over which to compute. Chow provides the following exposition:

> To be more precise, encapsulated systems avoid relevance problems in two subtly distinct ways: Not only does the small amount of information contained in the system's database constitute all the information that the system can consider, thus considerably reducing the number of computations needed for information search, but that small amount of information constitutes the one and only set of background information against which relevance is determined. The more encapsulated a system is, the more tractable its computations will be, and the less relevance problems will be problems. (2013, pp. 315–316)

Chow here is discussing one version of the frame problem, which he calls the 'Generalised Relevance Problem': "how a cognitive system can make determinations of what is relevant to a given task with reasonable levels of success" (p. 313). Since informationally encapsulated systems are insensitive to propositional knowledge the agent already knows, such systems significantly restrict the amount of information that can bear on a response-option. Fodor argues that systems like perceptual processing are encapsulated, but he also argues that central cognition isn't given its holistic nature. The upshot of this is that frame problems still arise for computations that are to be carried out by central cognition.

In this context, it is not far-fetched to suppose de Sousa's proposal is a way of explaining how we can solve the frame problem faced by processing at the level of central cognition. Emotions "mimic" the encapsulation of perception, and *ergo* restrict the amount of information over which to make computations at this level. We find an exposition compatible with this picture when we closely examine what de Sousa supposes is achieved by the encapsulation of emotion processing:

[A]n 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 among which it will choose. (de Sousa 1987, p. 195)

If we take the inspiration for this to be Fodor's suggested proposal as to how we solve the frame problem, a plausible interpretation has it that the temporary information encapsulation of emotions don't really solve the problem for central cognition as much as prevents the problem from arising despite its usual holistic nature. There are, however, two major worries with this proposal. First, it remains unclear how the Fodor-inspired move actually addresses the frame problem understood as a problem of not just restricting information, but restricting it to that which is relevant to achieving reasonable levels of success when it comes to choosing response-options.

This worry is accounted for given the overall framework de Sousa proposes. Emotions bias certain pieces of information, and they acquire the biases they do have for a variety of cultural and biological reasons. de Sousa's exposition employs the notion of "paradigm scenarios" where we first acquire our knowledge of which emotional responses are appropriate for a given situation. The appropriateness of future emotional responses is measured against the backdrop of the responses acquired during these scenarios. The details don't really matter for current purposes. What does is the availability of plausible developmental stories which explain why we acquire our emotional biases. These suffice to explain why our biases confer salience to information relevant to achieving reasonable levels of success. In brief, both cultural and biological evolution selects for emotional biases that, on the whole, achieve reasonable levels of success. Once again how the philosophers' frame problem is resolved, then, is explained without the need to make any commitments to emotion processing being informationally encapsulated.

The second worry, like the first, concerns a lack of success when we apply the Fodorian move to the philosophers' frame problem. The crucial feature of an encapsulated system, de Sousa notes, is it "limits the range of information that the organism will take into account" (p. 195).

This is ambiguous as to whether the information limited is background information or in-coming information. Informationally encapsulated systems successfully limit background information. To be an encapsulated information processing system is simply to be insensitive to background information—except for any information stored within that very system. Consequently, there is no question of whether they limit background information. Informationally encapsulated systems, however, do not limit in-coming information. This is significant because insofar as solving the philosophers' frame problem also requires limiting in-coming information, the information encapsulation of emotion processing will not fully resolve this problem.

To elaborate, as we noted earlier, the frame problem in its broadest form is a problem about how to restrict the amount of information that is to be computed for us to get an appropriate response-option. This understanding of the problem gives way to three possible interpretations: we need to restrict (i) background information, (ii) in-coming information, or (iii) both. Fodor's discussion of the frame problem assumes (i), whereas prevailing discussions of the philosophers' frame problem, including de Sousa's own, centres on (iii). The inclusion of (ii) is perhaps clearest in Faucher and Tappolet's exposition of de Sousa's proposal where they write, it "is natural to think that emotions influence what we attend to and determine what information we take in" (2002, p. 108). Quite independent of de Sousa, the function of emotion, construed as a mechanism that biases in-coming perceptual information, is discussed within the context of both evolutionary psychology and neuroscience. For evolutionary psychologists Ketelaar and Todd, "specific emotions might help to solve the problem of what information to attend to in specific environmental circumstances" (2001, p. 194). Likewise, for neuroscientist Panksepp, emotions help "modulate attentional and sensory-motor sensitivities relevant for the evolved behavioural tendencies" (2000, p. 22). Faucher and Tappolet themselves argue the thesis that emotions bias perceptual data is empirically borne out, at least with regards to the emotion of fear and anxiety.

Assuming another way emotions assist reasoning is by limiting in-coming information, we face similar problems when we try to find a role for how information encapsulation of emotion processing helps

with this task as we do when we try to determine the precise role it plays in assisting reasoning when it limits background information. Informationally encapsulated systems don't restrict in-coming information. They are neither insensitive to certain kinds of in-coming information, nor do they mark certain pieces of such information as more salient than others. As before, how emotions determine the salience of information, in this case in-coming information, can be explained in terms of their affective component. Emotions bias certain in-coming information over others because they have an affect component, which helps evaluate certain pieces of information as being positive or negative. Moreover, emotions acquire the biases they do have for a variety of cultural and biological reasons. A story of how emotions determine the salience of features of perception can, therefore, be told *sans* any commitment to emotion processing being encapsulated.

To recap, a focus on information encapsulation within the context of the philosophers' frame problem, I suggest, stems from the role Fodor attributes to it with regards to the frame problem in artificial intelligence. The two problems discussed above, however, demonstrate how information encapsulation of emotion processing cannot resolve the philosophers' frame problem. The encapsulation of emotion processing fails to resolve this problem because it does not determine the salience of information—be it in-coming information or the relevant background information. We then reach a negative verdict. Nevertheless, putting aside the role Fodor attributes to encapsulated systems helps us see an alternative way encapsulation of emotion processing might be relevant for the philosophers' frame problem. Emotion processing, on occasions it is encapsulated, doesn't determine our emotion-driven biases, but rather ensures these biases are not overridden.

We have predetermined biases which emotions seek to highlight in virtue of their affective components. Provided our emotional responses, and their ensuing biases, are triggered without considering background information, this ensures that these biases are manifested in a way that isn't overridden by what the agent knows or believes. For instance, there might be various evolutionary or sociological reasons why our fear responses are more acute for unrecognised noises, especially when they are coming from the periphery, or other cases of visual disintegration,

say walking in a dark alleyway. These responses seek to highlight such noises over other perceptual data, e.g. the beautiful reflection of yellow light from a streetlamp on the wet pavement. Moreover, they do so regardless of what the agent believes. An aesthete might have certain beliefs about the good life being one where she appreciates beauty wherever possible. These beliefs may, in practice, result in tendencies to prioritise perceptual data which lead to aesthetic experiences. Nevertheless, her fear responses, where they are informationally encapsulated, ensure these responses, and the ensuing biasing of unrecognised auditory data, are not overridden by her aesthetic beliefs and belief-based tendencies. She may live her life religiously practising her appreciation of beauty, but her fear responses ensure she prioritises potential sources of danger when considering response-options, e.g. when walking in a dark alleyway.

Information encapsulation of emotion processing plays the same role when it comes to restricting non-perceptual information. Emotions bias certain inferential patterns over others. Moreover, emotion processing being encapsulated ensures these biases are not overridden. The aesthete, for instance, might be prone to contemplating the beautiful, even when walking in dark alleyways by herself. But the information encapsulation of the mechanisms that trigger her fear responses ensure these responses are triggered by unrecognised noises, regardless of her intentions. Moreover, these responses in turn prioritise inferences concerning the selection of response-options to threat over the sorts of inferences involved in contemplating the beautiful.

The claim here isn't that there can't be any top-down effects of cognition on emotion processing. Through training, various beliefs might come to penetrate such processing, especially diachronically. Cognitive behavioural therapy might, for example, be employed to dull our fear responses to walking in dimly lit streets.[8] However, where emotion processing "mimics" the encapsulation of perceptual processing, emotions can continue their role of acting as biasing mechanisms which seek to

---

[8]A subject might undergo such therapy if trauma has induced debilitating fear responses to walking unaccompanied in dark streets.

prioritise certain pieces of information over others. In other words, the information encapsulation of emotion processing guarantees that these biases are not overridden by what the agent believes. Encapsulation doesn't, then, determine the salience that emotions confer information, but rather turns out to be necessary for guaranteeing that emotions can fulfil this role of determining salience.

4   Conclusion

This paper concerned itself with what cognitive architecture could tell us about emotion-driven reasoning. To that end, we assessed a hypothesis about a particular feature of such architecture owing to de Sousa: "the role of emotions is to supply the insufficiency of reason by imitating the encapsulation of perceptual modes" (1987, p. 195). We now find this interpretation to be misleading. The central role of emotion in emotion-lead reasoning isn't to mimic the information encapsulation of perception but to determine the salience of information; something it can do without being informationally encapsulated. However, we also see there is a sense in which de Sousa was on the right track. Emotions do "control" the salience of features of perception and reasoning. The positive parts of this paper fleshed out precisely how this happens. The encapsulation of emotion processing ensures that emotions can fulfil their central role in an efficient manner, and in a way that isn't overridden by what the agent knows.

If this account is correct, analyses of emotional rationality can proceed without any presuppositions about cognitive architecture. That is, we can, at a level of abstraction, explain how emotions assist reasoning without making any commitment to emotion processing being informationally encapsulated. Nevertheless, if we are serious about providing a detailed story as to precisely how emotions assist reasoning, including an account of the mechanisms by which it does so, information encapsulation will be part of the parcel of such a story. How philosophically pressing knowing about cognitive architecture is to the endeavour of explaining emotional rationality will, then, depend on just how deep you want your theory to go.

References

Barnard, P. 1985. Interacting Cognitive Subsystems: A Psycholinguistic Approach to Short-Term Memory. In *Progress in the Psychology of Language*, vol. 2, ed. A. Ellis, 197–258. London: Erlbaum.

Brady, M. 2013. *Emotional Insight: The Epistemic Role of Emotional Experience*. Oxford: Oxford University Press.

Carruthers, P. 2018. Valence and Value. *Philosophy and Phenomenological Research* 97 (3): 658–680.

Chow, S.J. 2013. What's the Problem with the Frame Problem? *Review of Philosophy and Psychology* 4 (2): 309–331.

Colombetti, G. 2005. Appraising Valence. *Journal of Consciousness Studies* 12 (8–10): 8–10.

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

Dennett, D.C. 1984. Cognitive Wheels: The Frame Problem of AI. In *Minds, Machines, and Evolution*, ed. C. Hookway, 129–152. Cambridge, MA: Cambridge University Press.

de Sousa, R.B. 1987. *The Rationality of Emotion*. Cambridge and London: MIT Press.

Evans, D. 2002. The Search Hypothesis of Emotions. *British Journal for the Philosophy of Science* 53 (4): 497–509.

Faucher, L., and C. Tappolet. 2002. Fear and the Focus of Attention. *Consciousness and Emotion* 3 (2): 105–144.

Fodor, J.A. 1983. *The Modularity of Mind*. Cambridge and London: MIT Press.

———. 1987. Modules, Frames, Fridgeons, Sleeping Dogs and the Music of the Spheres. In *The Robot's Dilemma: The frame Problem in Artificial Intelligence*, ed. Z. Pylyshyn, 139–149. Norwood, NJ: Ablex.

———. 2000. *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

Greenspan, P.S. (1996). Emotions and Reasons: An Inquiry into Emotional Justification. *The Philosophical Quarterly* 46 (183): 281.

Ketelaar, T., and P.M. Todd. 2001. Framing Our Thoughts: Ecological Rationality as Evolutionary Psychology's Answer to the Frame Problem. In *Conceptual Challenges in Evolutionary Psychology: Innovative Research Strategies*, ed. H.R. Holcomb III, 179–211. Norwell, MA: Kluwer Academic Publishers.

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

Leventhal, H. 1979. A Perceptual-Motor Processing Model of Emotion. In *Perception of Emotion in Self and Others*, vol. 5, ed. P. Pilner, K. Blankstein, and I.M. Spigel, 1–46. New York: Plenum.

Panksepp, J. 2000. The Neuro-Evolutionary Cusp Between Emotions and Cognitions: Implications for Understanding Consciousness and the Emergence of a Unified Mind Science. *Consciousness and Emotion* 1 (1): 15–54.

Prinz, J. 2010. For Valence. *Emotion Review* 2 (1): 5–13.

Solomon, R.C. 1976. *The Passions*. Garden City, NY: Anchor Press and Doubleday.

———. 1977. The Rationality of Emotions. *The Southwestern Journal of Philosophy* 8 (2): 105–114.

Teasdale, J. 1999. Multi-level Theories of Cognition-Emotion Relations. In *Handbook of Cognition and Emotion*, ed. T. Dagleish and M.J. Power, 665–681. New York: Wiley.

Part III

Emotion Regulatory Affordances

4

A Pragmatist View of Emotions: Tracing Its Significance for the Current Debate

Roberta Dreon

1   Introduction

Among the many experts on human emotions, mention of William James is still due. For better or worse, James's position on the subject represents a milestone: supporters of a cognitivist approach to emotions, even of a soft cognitivist thesis, such as Martha Nussbaum, appeal to James in order to contest his shortcomings in focusing on cognitive and evaluative aspects of emotions (Nussbaum 2001); neuroscientists, even ones of a non-reductivist sort, such as Antonio Damasio, refer to James to emphasize his crucial insights into the primarily embodied dimension of emotions (Damasio 1994, 1999).

Things could not be otherwise because the core of James's theory is famously constituted by a basic overturning of "our natural way of thinking about [..] standard emotions", according to which "the mental perception of some fact excites the mental affection called the emotion",

R. Dreon (✉)
Ca' Foscari University of Venice, Venice, Italy
e-mail: robdre@unive.it

so that "this letter state of mind gives rise to the bodily expression. My thesis, on the contrary, is that *the bodily changes follow directly the* PERCEPTION *of the exciting fact, and that our feeling of the same changes as they occur* IS *the emotion*" (James 1884, pp. 189–190).[1] James rejected the standard picture of emotions as mental states causing their bodily expressions by stating that an emotion consists in the feeling of the changes happening in one's body, caused by the perception of something troubling, exciting or joyful in the environment.

However, my contention is that James's contribution is only one side of a more rounded pragmatist approach to emotions, which was further developed by another two champions of classical pragmatism, John Dewey and George Herbert Mead—even if it must be remembered that at the time of the formulation of his thesis (from 1884 to 1891) on this subject James had not yet pronounced himself in favor of the pragmatist movement.[2]

Consequently, my suggestion here is to consider James's proposal not as a complete theory that can be criticized in itself—even if this is a perfectly legitimate approach, of course. My attempt here is to deal with the various aspects of his conception as the beginning of an open inquiry, which was further developed over the two following decades. The present paper will consider Dewey's and Mead's points of view on James's reflections, by focusing the reader's attention on some convergences and elements of disagreement, but also on some mutual, and sometimes even radical, amendments. This kind of approach allows us to recover a challenging perspective on emotions which basically emphasizes the continuity between mental and bodily aspects, emotive behaviors and cognitive practices, while highlighting the already social characterization of the environment where individual emotive responses take place.

This kind of approach sounds refreshing if compared with the prevailing brain-centered, Cartesian and methodologically solipsistic trend

---

[1]Both italics and capitals are James's.

[2]However, it must be also recalled that Dewey's turn from Hegelianism to pragmatism was deeply connected with his reading of James's *Principles of Psychology* (James 1981) and Darwin's *On the Origin of Species* (Darwin 1996).

in neurosciences, the limits of which have already fruitfully been highlighted by Noë (2015, p. 94 and ff.). On the contrary, it allows us to draw attention to some interesting convergences with current embodied, embedded, enactive and socially extended interpretations of affectivity and emotions (4E).

While leaving aside Darwin's attempt to identify a series of basic emotions and their correspondent bodily expressions[3] (still a preeminent enterprise in mainstream studies on emotions),[4] James, Dewey and Mead took from Darwin the idea that emotions had to be considered primarily with reference to life, namely to the structural interactions between human organisms and their natural and naturally social environment. From this point of view, the pragmatists largely anticipated the enactivist approach to emotions brilliantly developed by Giovanna Colombetti (2014) as well as the socially extended form of enactivism supported by Hanne De Jaegher and Ezequiel Di Paolo, on the one hand, and by Shaun Gallagher, on the other (De Jaegher and Di Paolo 2007; Gallagher 2013, 2017; Krueger 2013). Moreover, Dewey developed the idea that we experience the things, persons and situations around us primarily in terms of what they can do for us, against us or with us—in his language, in terms of their qualitative or aesthetic meanings and values for our life (Dewey 1981; Dreon 2012). This position seems to be convergent with Colombetti's idea of "primordial affectivity" (Colombetti 2014, 2017), namely the idea that affectivity is a basic character of human life and that emotions are not isolated episodes in a basically an-affective cognitive mind. Dewey's insight into the basic aesthetic or qualitative connotation of our experience of the world around us was also contemporary to (although completely independent of) the

---

[3]James distanced himself from Darwin and Lange's attempt to identify a series of basic emotions and their correlated gestures in the re-elaboration of his essays on emotions, constituting Chapter XXV of his *Principles of Psychology*. Here he added an introductory section highlighting the limits of their "sort of descriptive work". James makes explicit reference to Darwin (1998) and Lange (2012).

[4]See Eckman (1999). In the second chapter of her book, Colombetti develops a detailed criticism of the theory of basic emotions (Colombetti 2014, p. 25 and ff.).

phenomenological account of the so-called *Befindlichkeit* provided by Scheler (1923) and Heidegger (1967). Mead's thesis of a social and communicative configuration of emotions in humans is probably more radical than Colombetti's position and could be considered (even though it was only roughly sketched out) as basically convergent with Fogel's conception of emotions as dynamic processes defining themselves only in already socially shared contexts (Fogel et al. 1992).[5] Some conceptual tools and arguments developed in the current analysis of emotions in the field of 4E cognition can help define in a more analytical way some insights already suggested in a very condensed form by the classical pragmatists and sometimes in need of being further articulated.[6]

Finally, the sort of inquiry I am suggesting here makes it possible to adopt a clear position with reference to the main subject of this volume, namely the relations between emotions and cognition. It is well known that all the classical pragmatists, from Charles Peirce onward (Alexander 2002; Calcaterra 2003), have supported the thesis that qualitative, affective or aesthetic aspects pervade cognition: they are seen to lie at the basis of current cognitive practices as well as to constitute a means of selecting, controlling and testing inquiries (Dewey 1984, 1988a, 1991). Furthermore, affective forms of attunement are seen to lie at the roots of verbal communication, both phylogenetically and ontogenetically (Mead 2011). This kind of background—in a formula, the idea of a structural intertwining between sensibility and rationality—should be considered as supporting Dewey's thesis of a proto-evaluative meaning of emotive behavior and his correlative rejection of the assumption of an allegedly merely descriptive level of perception giving rise to a subsequent judgment.

---

[5]See Candiotto (2016) for an interesting comparison between extended mind approaches to intersubjectivity and enactivistic treatments of sociality.

[6]I originally developed an inquiry into the pragmatist conception of emotions independently of any engagement with this kind of current debate (in Dreon 2015). However, I think that the comparative inquiry I tried to develop in this work (thanks to Laura Candiotto and Pierre Steiner), makes the whole reconstruction clearer in terms of its theoretical implications as well as, hopefully, more challenging for current discussions.

I will begin my inquiry by trying to read James's germinal essays on emotions through Dewey's and Mead's eyes. I will then focus on Dewey's adoption, correction and integration of both James's and Darwin's conception of emotions. Finally, I will consider four main aspects of emotions examined by Mead, which partially develop and somewhat redirect Dewey's interpretation of emotions.

2   Dewey's James, or What Dewey (and Mead) Found Crucial in James's Theory of Emotions

What outstanding elements in James's theory might have inspired Dewey first and then, immediately afterward, Mead?

(1) My suggestion is that one of the most stimulating aspects in James's position for both these younger pragmatists was a negative one, that is its explicitly avoiding any recourse to mental states understood as the cause of bodily changes. James did not generally refuse to make any reference to mental processes: the point is that he was giving up the idea that we should assume allegedly previous mental states as the causes of emotionally laden bodily changes. We could sum this up as the no "mind-stuff" argument, in James's words. The two pragmatists' perspective was by no means that of the behaviorists, who notoriously avoid any use of introspective analysis and limit themselves to considering only observable actions because of the epistemological problems determined by introspection.[7] On the contrary, both Dewey and Mead were seeking to understand mind and consciousness no longer as entities or substances of a distinct kind, but rather as particular forms of interaction between human organisms and their natural and naturally social environment or as peculiar phases of experience, basically continuous with the environment they belong to and which they contribute to transforming from the inside. In other words, long before the enactivists

---

[7]On the difference between Pragmatism and Behaviorism, see Mead (1934).

(see Colombetti 2014: XVI–XV), they considered organic life in an environment as the necessary background for interpreting mental processes.[8]

Of course, as is well known, James was still adopting the kind of dualistic parallelism that configured the traditional framework of his *Principles*, even though he radically overturned it, at least in the formulation of his theory of emotions.[9]

Nevertheless, James's conception does not simply emphasize the implication of the whole body—as opposed to just the brain—in a wide range of emotions. What I find interesting in his conception of emotions is the idea of the body that James was adopting here. It might be argued that this was a (2) second point of interest for Dewey and Mead, which was instead at least partially ignored by those who criticized James's theory as reductionist and as rejecting the idea of any intentional dimension to the emotions (Cunningham 1995). We might summarize this point by stating that the other pragmatists may have seen in James's seminal text his structurally interactive conception of the human body and of the nervous system in particular—where "interactive" is more than "intentional".

Indeed, in James's words, the feelings of the body are not understood as mere private and internal feelings, principally impeding or marginalizing the role of the alleged external world and of the objects "out there". "[..] my first point", James argued, "is to show that their [of the emotions] bodily accompaniments are *more far reaching* and complicated than we ordinarily suppose" (James 1884, p. 191).

---

[8]It must be said that Dewey and Mead reserved the terms "mental" and "cognitive" for interactions between human organisms and their environments. Dewey preferred to speak of "sensitivity" in the case of animals capable of locomotion and Mead focused on conversations based on gestures between animals as emotional, not yet verbal gestures. In both cases, the interactions of nonhuman life forms are understood as being meaningful but not yet capable of establishing forms of triadic signification. This point highlights a difference in comparison to the enactivist equation of every form of "sense-making" with cognition. Dewey and Mead strongly felt the need to consider the peculiarities of human behaviour against the background of a basic continuity with other forms of life, although they did not deny that nonhuman animals can display a kind of intelligent and sensible behaviour.

[9]On this point, see Dewey (1988b).

The body is not conceived as a kind of closed entity, with an alleged external stimulus having the mere function of giving rise to a mechanical process; on the contrary, by emphasizing its plasticity, James understands the nervous system as both an active and passive function of the environment, capable of reacting to some selected aspects of it and of ignoring others.[10]

To put it in the language of the extended mind theory, we could say that for James the nervous system in a body constitutes a system coupled with its surroundings (Clark and Chalmers 1998). Moreover, it should be noted that James's emphasis on the plasticity of the nervous system in *The Principle of Psychology* could be seen as basically foreshadowing more recent ideas that the brain is at least partially configured after birth in relation to the experiences of the individual in a social and already cultural context, as contended by Fogel and Greenough (Fogel et al. 1992; Greenough et al. 1987).[11]

James understood emotions as attitudes our nervous system displays to react to certain environmental stimuli or even as inclinations anticipating environmental factors to which a given organism is sensible. Our nervous system and our visceral apparatus are not seen as being completely absorbed in themselves, so to speak; instead, they are "a sort of sounding-board, which every change of our consciousness, however slight, may make reverberate" (James 1884, p. 191). Consciousness here is seen to be constituted by nothing more than the perception of a certain environmental aspect, which is crucial or at least important for that particular organism. In other words, during an intense emotional experience, the body does not simply feel itself as a sort of closed entity, but feels itself suffering or enjoying the surrounding world.[12]

---

[10]Jim Garrison suggests the idea that Dewey admired James's functionalistic account of the psychic and quotes Dewey by pointing out that James's idea of the organism was not a static one: on the contrary, he tended to "think life in terms of life in action" (Garrison 2003, p. 405).

[11]On this aspect, see in particular the chapter James devoted to habits in his book (James 1981). Here he speaks about the development of neural paths through use and environmental exposition.

[12]On this conception of the feeling body, see Goldie (2002, p. 236). This point is further confirmed by James's *The Physical Basis of Emotion* (James 1983), where he argues that an emotion is a kind of secondary feeling indirectly aroused by an object, that is a feeling of one's own body being affected by an external object. Besides, in the 1891 paper James speaks about the possibility

(3) This leads us to a third aspect which makes the cords of Dewey's soul resonate, that is his emphasis on the affective or aesthetic dimension of our experience of the world. James begins his 1884 paper by claiming that most physiologists have only focused on cognitive and volitional performances when exploring the functions of the brain while ignoring "the *aesthetic* sphere of the mind, its longings, its pleasures and pains, and its emotions" (James 1884, p. 188).

In my opinion, by adopting and radically extending James's idea of so-called "affectional facts" in experience (James 1976, Chapter V), Dewey came to understand our suffering or enjoying situations, other individuals and objects as our primary, pre-reflective way of experiencing the world (Dewey 1981, 1989). First of all, we feel things, persons and contexts as comfortable and welcoming or as dangerous, bitter or annoying: our experience of the world is already laden with these aesthetic qualities and meanings before becoming proper knowing, according to Dewey. It is a mode of experiencing things that is already revealing of the actions that the overwhelming environment triggers in on us and which therefore implies both a proto-evaluative character and a proto-cognitive import about the world around us prior to any analytic distinction between subject and object, or any feeling, evaluation and cognition (Dreon 2007, 2012; Quéré 2013). In his early essay on emotion, James went so far as to argue that a purely "intellectual feeling hardly ever does exist thus unaccompanied" (James 1984, p. 202) by bodily reverberations. Dewey instead explicitly supported the thesis that our cognitive experience is secondary or derived, reflecting as it does the need to respond differently to critical situations, where habitual behavioral responses do not work and the whole indeterminate situation must be reflectively analyzed (Dewey 1991, 2004). From this point of view, Dewey largely anticipated what Giovanna Colombetti calls "primordial affectivity" (Colombetti 2014, 2017), namely the idea that human life

---

of both afferent currents in the nervous system going from the objects to the body and of other kinds of affective currents, going from the center to periphery.

is always affectively oriented in one way or another and that emotions are not isolated episodes in a basically an-affective (or "an-esthetic", as he says in Dewey 1989) knowing mind.

In turn, Mead may have been inspired by James's focus on the aesthetic aspects of human experience for at least two reasons. The first reason is that by linking emotions to pain and pleasure, to longings and refusals, we can understand both the continuity between animal behavior and human experience and the peculiarity of the latter, consisting in the symbolic character of the objects causing painful or enjoyable interactions (Mead 2011).[13] Secondly, Mead may have found inspiration in James's acknowledgment of the fact that "the most important part of my environment is my fellow man. The consciousness of his attitude towards me is the perception that normally unlocks most of my shames and indignations and fears" (Mead 2011, p. 195). In *The Social Character of Instinct* Mead was to state that the "primitive consciousness even of the physical world is social", or—to put it more explicitly—that our consciousness is primarily affectively oriented toward other people's reactions to our own actions. Consequently, this affective social perception of the world "becomes physical consciousness with the growing power of reflection".[14] This idea of human emotions as basically dependent upon human sociality is probably more radical than the one supported by Colombetti's book. Nevertheless, developmental psychologists such as Stern (1985) and Trevarthen (1979) seem to go in Mead's direction, as does Fogel et al. (1992). Mead's intuition about a primarily social configuration of affectivity in humans could push the idea of participatory sense-making further (De Jaegher and Di Paolo 2007), making its affective characterization more explicit (see Gallagher, Fusaroli, Kruger and also Colombetti 2017).

---

[13]Mead focuses on this point specifically in an abstract entitled *A Theory of Emotions from the Physiological Standpoint*.

[14]For an interesting convergence, see also Merleau-Ponty (1942, p. 191): "[…] la perception commençante est, beaucoup plus qu'une opération cognitive et désintéressée, un contact émotionnel de l'enfant avec les centres d'intérêt de son milieu […]".

3   Dewey's Contribution

Dewey begins the first of his two papers on emotions published in 1894 and in 1895 by explicitly declaring that his suggestions on this subject must be understood "as a possible outline for future filling in, not as a proved and finished account" (Dewey 1971a, p. 152).[15] James's theory of emotions, together with Darwin's one, constitutes the central focus of his analysis—respectively in the second paper, "The Significance of Emotion", and in the first one, "Emotional Attitudes". Moreover, Mead's work on emotion is also expressly mentioned in a footnote, where Dewey expresses his hope that the whole theory, which was being formulated by his colleague and friend, might "soon appear in print" (Dewey 1971a, p. 167). These elements support my interpretation of the pragmatists' work on emotion (including that of James) as a kind of open laboratory for ideas and hypotheses.

Concerning Dewey, we must also remember that this subject will remain a key one for the development of his thought: important reflections are presented both in *Experience and Nature* and in *Art as Experience*, but there is not enough room to focus on this later production in the present paper.[16]

The common ground of the two essays lies in the attempt to establish an interaction between Darwin's theory and James's one, by correcting some of their crucial mistakes and bringing out some points of divergence, which James had not explicitly considered when formulating his ideas on emotions.

According to Dewey, the crucial problem faced by both the above-mentioned authors is constituted by the dichotomy between the psychical and the physical aspects of the emotions that unfortunately continued to be maintained even in James's conception. His statement that emotion is the feeling of the bodily changes following directly

---

[15]Dewey quotes not only from the *Principles*, but also from James's paper *The Physical Basis of Emotion*.

[16]On this subject, see Dreon (2012, Chapters 2 and 3). See also Quéré (2013), who points out that Dewey's early conception of emotions should be integrated with the idea he developed more explicitly in the 1920s and 1930s (in particular in Dewey 1988a, 1989).

upon the perception of an exciting fact causes a radical reversal of the elements at play: James ultimately assigns priority to physical factors over psychic ones. However, in doing so, he confirmed and eventually reinforced the mind–body dualism, even though he probably already had trouble with his own presupposition—for sure, he later openly sought to abandon it, as already noted.

Nevertheless, let us proceed with order, by following Dewey's train of thought.

Darwin's unquestioned premise is that the emotions are prior to the "organic peripheral action[s]" and that for this reason alone facial and superficial bodily changes can be understood as the outer expression or communication of emotions themselves. Incidentally, it should be pointed out that also the current theory of basic emotions (Eckman 1999) maintains that there must be something previous to changes in the body, causally eliciting them. Most current views of emotions take neural affect programs (namely, an already fixed set of them corresponding to the alleged basic emotions) as constituting the mental stuff preceding bodily changes—meaning that a mind-centered stance is replaced by a brain-centered approach.

Dewey clearly refuses to interpret visible bodily movements connected to an emotional experience as the outer expression of an alleged previous mental state and argues that this interpretation is based on a psychologist's fallacy, namely on the fact of conflating the observer's point of view with that of the facts observed. As noted by James in his *Principles*, "The psychologist […] stands outside of the mental state he speaks of", but he can unfortunately forget this and make fatal mistakes (James 1981, Chapter VII). So, if a gnashing of teeth can be interpreted as the communication of anger by an observer, the man gnashing his teeth is not doing that in order to communicate his rage to other people, at least not primarily—and differently from an actor performing on the stage. For him his bodily changes are simply an integral part of his being angry—Dewey will later observe that they are movements or acts inhering to his being angry, which is to say that they are emotional attitudes. In other words, bodily movements and actions tendencies to action, on the one hand, and feeling, on the other, are not separated, but integral parts of an emotion.

However, the psychologist's fallacy argument, in my opinion, does miss one important point, which will be later taken up by Mead: our actions are always social in principle, in the sense that they are subject to be seen by other people and we are almost always partially aware of this (as James noted in his seminal paper, although he did not further expound upon the idea elsewhere). In any case, we will return to this aspect later on, when discussing Mead's contributions to the pragmatist conception of the emotions. For the moment, let's focus on Dewey's texts.

The older pragmatist offers the reader a careful inquiry into Darwin's principles in order to explain emotions, starting from the first one, that is his principle of "serviceable associated habits". According to this principle, bodily changes, which are useful for communicating one's own emotions to others, are selected and acquired by following generations. In Dewey's opinion, Darwin's most important contribution to the understanding of emotions lies precisely in his teleological approach—the idea that we have to understand emotional attitudes, gestures or behaviors according to a functional perspective related to organic interactions in an environment. However, from Dewey's point of view, Darwin missed the target: emotional attitudes are not primarily significant with reference to an alleged preexisting psychic state, but with reference to useful movements; this means that they are functions of certain acts rather than of certain emotions understood as mental states. In other words, emotional attitudes must be interpreted as "acts originally useful not *qua* expressing emotion, but *qua* acts – as serving life" (Dewey 1971a, p. 154). Dewey suggests the example of laughing, which according to his interpretation is not primarily functional toward communicating a pleasurable state of mind, but rather constitutes the termination of a period of effort.

On the other hand, Dewey ascribes an important role to Darwin's third principle for explaining emotions, while denying any basic significance to his second principle—that of "antithesis". The former principle deals with so-called cases of "direct nervous discharge". According to Dewey's interpretation, those idiopathic cases in which no clear cause for a certain bodily movement can apparently be found must be understood as "cases of the failure of habitual teleological machinery"

(Dewey 1971a, p. 139), where previously successful habits can no longer supply good behavioral answers to a given situation. The functional approach is clearly always at work: the discharge of the nervous system is not meaningless and causeless, it is rather connected to a moment of crisis. Past movements and actions "formerly useful for a given end" (Dewey 1971a, p. 175) or previously adjusted to a certain environmental request prove to be inappropriate, so that new forms of coordination and interaction have to be found.

Summing up Dewey's adoption and correction of Darwin's theory, his position could be condensed as follows: Dewey recognizes and further emphasizes the importance of the functional interpretative key for understanding emotions as behavioral aspects. However, he connects emotive gestures and bodily changes to human (inter)actions embedded in and responding to a specific environmental context, rather than to an alleged interior state of mind of which they are mere external expressions.

So, what about James?

Dewey focuses his attention on James's approach to the emotions in his second paper, where what emerges is his stance based on the idea of continuity between mind and body, between the psychic and the physical. Furthermore, he offers a sharp critique of the traditional dualism between cognition and feeling, which appear to be integral parts of emotion, conceived as "a concrete whole of experience" (Dewey 1971a, p. 171), including the actions of the surrounding context upon us and our behaviors or acts as ways of interacting with it.

The problem with James is that he only focused on the feeling aspects of emotional experience while neglecting a series of other crucial elements, which are first of all experienced and perceived "as a whole carrying its distinctions of value within it" (Dewey 1971a, p. 173). From this point of view we might argue that Dewey's starting point is that of "our ordinary, everyday way of thinking of the emotions, and the phenomenology of emotional experience", whose unity, in Peter Goldie's opinion, is prior to any distinction between the mental and the material, but also, according to Dewey, between feeling, knowing, evaluating and acting (Goldie 2002, p. 247). This intertwining of feeling and cognition represents an important point of convergence with Colombetti's

conception of emotions, which also highlights recent neuroscientific accounts questioning the traditional assumption that "cognition and emotion are distinct psychological faculties implemented in separate neural areas" (Colombetti 2014, p. 98).

Dewey amends James's interpretation by pointing out that the feeling of bodily changes is a structural part of an emotion, but does not exhaust the concrete emotional experience. Emotions must be considered modes of actions and behaviors, because within them, first of all, we can detect a "readiness to act in certain ways" (Dewey 1971a, p. 172) in response to a certain situation or a certain object with which an organism is interacting; they are dispositions toward other men and women, modes of conduct or ways of behaving in a certain situation. This behavioral side of emotion is characterized as being "primarily ethical" by Dewey because, even if we can intellectually distinguish a feeling aspect from the action or reaction we are assuming, this same behavior carries a kind of proto-evaluation within it about what is better or worse, dangerous or enjoyable. It is a primitive form of evaluation, which is not the result of a judgment but depends on how I feel or perceive the situation I am embedded in. In this way, Dewey avoids the traditional objection to James's approach coming from the cognitivistic conception of emotions, by questioning at the same time the idea of a two-level experience of the world: firstly, an alleged merely perceptive and neutrally descriptive way of experiencing what there is and, secondly, a judgment about its congruity or dissonance with one's own life and interests (Dreon 2012, Chapter III). In his *Ethics* Dewey says that our reflective morality, which is to say our capacity to consciously reconsider our behaviors and norms, rests on a customary morality, based on habits as well as on our structural tendency to affectively accept or reject what is going on around us, because our lives are radically dependent on and exposed to an environment. In this sense, Dewey can be seen as largely agreeing with Colombetti's idea that we should distinguish between a primary appraisal and more reflective judgments or a secondary appraisal (Colombetti 2014, p. 85).

A further aspect of a full emotional experience is constituted by its being always about a certain object, person or situation—its being oriented toward something. This is its so-called intentional

structure,[17] which Dewey called "prepositional reference". Many years later, in *Experience and Nature*, he was to suggest that, by looking at the grammar of emotions, we should reject the typically modern philosophical *topos* that emotions are essentially confined within our private consciousness and can in some lucky cases be expressed to the outer world by means of almost magic works of art. In fact, emotions are always *for* something, *toward* a certain person or event, *against* a given situation, *because of* a terrible or wonderful fact. That is, rather than confining each one of us to his or her own subjectivity, they reveal our structural exposition to the environmental and social forces surrounding us.

We could characterize this intentional or prepositional side of a concrete emotion as its cognitive aspect. However, even from this point of view Dewey's approach is original because of the overlapping of the cognitive factors with the aesthetic ones, which makes the emotional perception of an object something that cannot be divided into two allegedly separate phases, as argued for example by Martha Nussbaum through her weak cognitivistic or neo-Stoic approach (Nussbaum 2001). In Dewey's perspective, we do not primarily perceive or register a matter of fact and then ascribe it some value. For example, I do not simply perceive a bear with my sense organs or as the mere descriptive content of an experience and then feel afraid because I am in the forest and not at the zoo looking at the bear in its cage. On the contrary, I perceive the bear as frightening if I am in the forest, or as amazing (or miserable) if I am at the zoo. Only later, can I intellectually abstract the qualitative or aesthetic elements, namely the significance of the bear in my life, from the bear in general or from the alleged mere perception of the bear. According to the later Dewey, this abstraction is not mistaken per se, because the reflective returning to an immediate experience in order to analyze its different phases can be a way to find a solution to a moment of crisis and can enrich the following

---

[17]"Intentional structure" here does not involve an alleged content in one's own mind; on the contrary, it involves the quality of being referred to something or about something, according to the phenomenological usage of the term.

experiences. Conversely, it is a philosophical fallacy to assume the result of the abstraction—the alleged mere descriptive sensitive or conceptual contents—as a first neutral element that can be evaluated and judged only later.

In my opinion, this is a point that will play an outstanding role in the construction of Dewey's philosophy, according to which our experience of the world around us is primarily aesthetic or qualitative and not purely cognitive or conceived as the object of an alleged mere sensory perception.

But, to stick to this text, we may note that Dewey is perhaps more radical here, because he goes so far as to suggest that in a concrete emotional experience we cannot primarily separate our feeling afraid from the frightening bear that is scaring us, while both the object and we as the subject emerge from a basically unitary experience (Dewey 1971a, p. 176).

Many commentators[18] have rightly pointed to the significance of Dewey's paper on the reflex arc concept (Dewey 1971b) for his interpretation of the emotions, by underlying the crucial role of the so-called "organic circuit" in human behavior—an idea largely preceding not just Alva Noë and Shaun Gallagher's insistence on the intimate intertwining between perception/cognition and action, but even Merleau-Ponty's phenomenological inquiries into the structures of behavior and perception (Merleau-Ponty 1942, 1945). However, Dewey's emotional circle does not only imply the interaction of acting and knowing, but also their being intertwined with the aesthetic or affective factor. Its qualitative meaning consists in a proto-evaluation about what I am suffering or enjoying and about what I have to do that can—and very often should—be further articulated into a reflective judgment, yet is not primarily a judgment in itself, as much as a matter of feeling. Dewey often uses the word "aesthetic", which implies a reference to sensibility as feeling (desire or rejection) rather than to sensory perception assumed as the basis of the cognitive building process; but he sometimes also uses the word "affective", which recalls James's preference

---

[18]Cunningham (1995), Garrison (2003), Quéré (2013) and Baggio (2015).

for this term—which is now key to Colombetti's interpretation of our emotive life. A passage from her 2013 book seems to perfectly fit with Dewey's account of emotions: "Emotion is not a distinct step in a perception-action sequence or a distinct representation added at some point to the sequence; emotion is rather an inescapable pervasive dimension of brain activity on which sensory information impinges and from which action progresses" (Colombetti 2014, p. 64).

4   Mead's Further Developments

Mead's contribution to this discussion about human emotions is especially notable in his papers dating from the period between 1894 and the first decade of the twentieth century. It is on these works, therefore, that I will be focusing. The first text to be considered is the abstract for *A Theory of Emotion from the Physiological Standpoint* (Mead 1895), a paper he presented to the American Psychological Association in 1894 (that is, the very year in which Dewey published his essay *Emotional Attitude*). The second text is *Emotion and Instinct* (Mead 2011, pp. 27–29), an unpublished essay, probably written sometime between the previous paper and another crucial text, *The Social Character of Instinct* (Mead 2011, pp. 3–8). This is another unpublished paper, which was certainly written after 1905 because it contains a reference to James Rowland Angell's book *Psychology*, which appeared in a first edition in 1904 (Angell 1904). There are also other essays to be taken into account in the background, that is *Social Psychology as Counterpart to Physiological Psychology* (Mead 2011, pp. 9–17), published in 1909, *What Social Objects Must Psychology Presuppose?* (Mead 2011, pp. 19–25), publicly read in the same year and published in 1910, and an unpublished essay entitled *A Psychological Study of the Use of Stimulants* (Mead 2011, pp. 73–82).

Mead's writing style is often very dense, so that it is not always easy to find a detailed articulation of the different stages of his arguments. In order to clarify Mead's contribution to the subject of the emotions, I am going to identify four main research trajectories within the complex web of his papers.

The first research strand, which emerges from the very beginning of Mead's interest in the emotions, is his urge to develop the physiological components implied in Dewey's teleological or functional conception of emotion. Mead stresses the need to articulate a physiological theory of pleasure and pain as the biological precedents of human emotions. One point of philosophical interest in this proposal is the fact that he adopts—as he usually does—a deeply continuist approach, where distinctly human emotions are seen as emerging from animal instincts and as being connected to pleasure or pain. From a physiological point of view, the increasing or decreasing of the processes of nutrition of organic tissues is the same both in man and in other animals. However, there is a particular factor to be considered in the human case, namely that "the vaso-motor processes that are originally called out only by the instinctive acts" are now called out by "symbolic stimuli" or "aesthetic" ones—as in the cases of war and love dances. These symbolic stimuli carry within themselves "an evaluation [by the organism] of the act before the coordination that leads to the particular reaction has been completed".[19] We could make Mead's intuition more explicit by emphasizing that a novelty here is represented by the action of cultural elements on physiological processes in the case of human emotions. Unfortunately, Mead does not tell us here anything more about what it means for a stimulus to be symbolic,[20] but we can appreciate the fact that he is trying to stress the peculiarity of human emotive behavior against the background of a basic continuity with animal life.

A second important point becomes clear in his paper *Emotion and Instinct*, where, once the basic emotional character of human interest has been recognized, the focus shifts to the differences between emotion and interest. One of the differences, according to Mead, regards their

---

[19]He will return to aesthetic stimuli of this kind in his later essay, *A Psychological Account of the Use of Stimulants*, with interesting suggestions on the origin of the arts (Mead 2011, p. 35).

[20]At this stage of his thought, Mead probably referred to Wundt's conception of symbolic stimulus. Symbolic gestures are those who mean indirectly, namely by means of a new application of an already existent sign, which is associated to the concept it represents through one or more ideas. Consequently, a symbolic gesture differs from demonstrative or imitative gestures because the latter kinds of signs are directly connected to what they signify. On this, see Wundt (1921).

position within a given act: while in most cases emotion characterizes "the immediate grasping and enjoyment of the object sought", interest involves a deliberate attempt to overcome the obstacles that impede the reaching of an end in view. Emotion seems to be connected to the immediate appropriation of the desired object (or to the avoidance of a dangerous one), and consequently seems to be almost instinctively realized, without any awareness of this pursuit of a given end by a given means: one's consciousness is rather completely absorbed in the desired or rejected object. On the contrary, an interested act is very often a conscious one, explicitly taking into account both the means and the ends in view of the action.

These suggestions are noteworthy because, although they are probably simplistic, they provide an interpretation of the connections and distinctions between emotion and cognition in a continuist perspective, where human instrumental reason seems to emerge from more instinctual modes of behavior when hindrances or inhibitions are at work. The emotional roots of cognition are not denied; on the contrary, they are conceived as basic components of intelligent behavior, which at the same time reveal themselves to be typically human forms of interaction with the environment.

Thirdly, in the paper, *The Social Character of Instinct* Mead definitely interprets the emotions from the social standpoint he had declared to be crucial in *Social Psychology as Counterpart of Physiological Psychology*. Here he states that an adequate interpretation of human emotions must not only take account of the biological factors involved but also set out from human sociality as the element in the light of which every peculiarly human phenomenon should be understood.

If we look at other human behaviors, such as curiosity, pugnacity, subjection, self-assertion and so on, we must recognize that they are basically referred to other individuals or to a structurally social environment. Mead calls these "instincts", according to William McDougall's use of this term, but they are definitely emotional behaviors.[21] Nonetheless, even if we look at human infancy, we cannot deny that the

---

[21]On the importance of McDougall's thought for Mead, see Hans Joas (1997, p. 91 and ff.).

newborn's movements are already attuned to those of its mother (or caregiver).[22] In another unpublished essay, *The Relation of the Embryological Development to Education*, Mead draws attention to John Fiske's thesis of the primary social importance of infancy: no other animal seems to have such a long period of dependence on others as human newborns, whose marked immaturity at the moment of birth makes them structurally dependent on the community taking care of them from the very beginning (Mead 2011, p. 73).[23] For this reason, Mead concludes that our "primitive consciousness even of the physical world is social, and only becomes physical consciousness with the growing power of reflections" (Mead 2011, p. 3).

It could be said that Mead was developing an idea of the socially extended mind, to adopt Shaun Gallagher's formula (Gallagher 2013). Nonetheless, I cannot fail to observe that the very concept of "extension" in connection to "mind" might have sounded strange to Mead because it seems to assume as a precondition the existence of a mind that could be further expanded into a social world. On the contrary, Mead was developing the idea that the individual mind and a mindful behavior can only emerge out of an already socially shared environment. Although Gallagher's idea of a socially extended mind is much more radical than Clark's conception of the extended mind (see Candiotto 2016), the extension vocabulary seems to pay its dues to the mind-centered approach from which Gallagher has vigorously moved away.[24]

---

[22]As noted by Cook (in Cook 1993), Mead had a strong interest in developmental psychology from both a theoretical and experimental perspective. Furthermore, he intertwined this kind of interest with his studies in animal and comparative psychology as well as with a strong attention to their evolutionary implications. This peculiar mix of interests contributed to configuring Mead's philosophical approach in a way that seems close to more recent and promising research trends at the intersection between philosophy of mind, cognitive and affective neuroscience, and developmental and evolutionary psychology.

[23]Similar reflections can be found in the first chapter of Dewey's *Human Nature and Conduct* (Dewey 1983).

[24]Analogous remarks could be made for Colombetti, regarding the connections between affectivity and sociality. In her book (Colombetti 2014), she approaches this issue from the perspective of the "feeling others" problem: in other words, sociality is not seen as a basic factor in configuring human affectivity. Her essays "Enactive Affectivity, Extended" (Colombetti 2017) seems to be more promising, even if the point in question is not fully developed.

The point is even more interesting with reference to our topic: in these texts, Mead seems to develop an idea of emotions as dynamically configured in a shared interpersonal context, as supported many years later by Fogel et al. Fogel denies that "[t]he emotion program is […] the source of the patterns that are differentially reinforced with respect to variations in the social context" (Fogel et al. 1992, p. 130). In this case, the innate emotion program in the brain would constitute "the ghost in the machine", preexisting the differentiation of emotive behaviors according to different social and cultural contexts. On the contrary, Fogel seems to develop an insight that had already been grasped by Mead, namely that biological factors and the social environment develop reciprocally and dynamically, at least in the case of human organisms. Mead's intuition was probably supported by his specific attention to the interactions between human infants and their caregivers: largely anticipating the inquiries by Daniel Stern (1985) and Colwyn Trevarthen (1979), Mead focused on mutual attunement rather than on imitation to understand the dynamic development of mother–infant interactions.

Finally, this shift from the biological dimension of animal instincts to the social one allowed Mead to find an answer to a basic gap in Dewey's argumentation, regarding the so-called expression of emotions.

We know that Dewey, together with James, rejected the traditional idea that a predefined inner state must be externally communicated by means of certain movements of the body and face. We also know that he made use of James's psychological fallacy to criticize this kind of interpretation. However, while James essentially focused only on visceral changes in the body, Dewey's functional interpretation of emotional attitudes—that is, the visible changes on the surface of the body—did not fully consider their social context. In a probably too condensed way, Mead argued that we have to recognize that emotions play a social role, not just a functional or teleological one, even if we have to reject the dualistic explanation of the perceivable bodily changes and the idea of the alleged priority of the psychical over the physical. Mead was already aware of this problem because of his proximity to Wundt's thesis of the origin of language in gestures, understood as affective expressions,

while having to avoid the recourse "to imitation and to expression of emotions" (Joas 1997, p. 103).

Mead's solution is that these emotional attitudes are immediately understood or read by the other social actors not as the outer transmission of a predefined inner state, but as gestures, that is as dispositions to act in certain ways, to which the interlocutors adjust or attune their own behaviors. In this perspective emotions are an integral and basic part of one's social conduct, namely of what later Mead will describe as the conversation of gestures, where the acts of one individual must be adjusted to others' movements. Emotional gestures are the first "means of co-ordinating social conduct" and they can be understood as signs signifying not a previous mental state, but a tendency to act or react to the movement of the interlocutor "in unreflective social conduct […]. [S]o we are continually reading from the attitude, the facial expressions, the gestures and the tones of the voices, the coming actions of those with reference to whom we must act" (Mead 2011, p. 5). This "reading" must not be regarded as a kind of explicit and articulated interpretation, or as a kind of subjective projection of one's own intentions, feelings and ideas into another's mind, but rather as a somewhat emotional tuning, which need not be conscious. In a formula, we could state with Krueger that emotions for Mead play a crucial role in shaping the mutual entrainment between mother and child as well as between adults (Krueger 2013).

From this point of view, the future key formulas for taking the role of the other seem to be rooted in a sensibility toward what the other can do—to us, with us or against us, Dewey will add. This interpretation reinforces Gary Cook's thesis that the phrase "taking the role of the other" "is a little unfortunate because it suggests an actor's attitude which is actually more sophisticated than that which is involved in our own experience" (Cook 1993, p. 78). Although it is primarily affective or based on feelings, this does not mean that it consists in the capacity to look inside the other's mind.[25] On the contrary, Mead focuses on the

---

[25]Very briefly, Mead's position is clearly far from the so-called "Theory-Theory" or "Mind-Reading" account of empathy (Stueber 2006). Maybe more interestingly, his view cannot be considered as a form of Simulation theory, if this account is understood as based on an analogy between one's own private, introspective experience and the allegedly doubtful experience of an individual different from and external to the one who would establish the analogy.

human-specific capacity to feel and imagine the possible movements of one's interlocutor, his possible (helpful or impeding) contribution to a shared action. My identity and that of my interlocutors are not previously determined before interactions take place, because it is precisely through social, affective interaction that they both develop or emerge. Besides, what situation can come before any form of social interaction? For Mead—as for Dewey—no such situation can be found in human experience, given the previous considerations about human structural embeddedness in a social environment because of man's marked immaturity at birth. Mead ultimately argues that these signs and this kind of emotional cooperative conduct constitute the basis for properly linguistic gestures, that is for symbols, and that "[t]hought and volition develop and interpret the situation that is first of all emotional". Hence, we might say that in this essay Mead finds the origin of human language in an emotional conversation of gestures, which is not substituted by verbal interaction but continues to exist within it. This is a rather complicated issue, involving—I would argue—the serious consideration of a feedback reaction (or loop effect, to use Alva Noë's words in Noë 2015) between human language and culture, on the one hand, and human sensibility, on the other. In my opinion, this hypothesis constitutes the coherent development of the classical pragmatists' conception of emotions—and maybe it could also represent a positive challenge for the most progressive trends in affective neuroscience. While it cannot further be explored here, it constitutes the focus and title of a promising forthcoming inquiry.[26]

5   Conclusion

Having reconstructed the pragmatists' account of emotion as a sort of open laboratory shared by James, Dewey and Mead, we have identified a more rounded pragmatist theory of emotion and found that these thinkers' challenging contributions to the subject can be appreciated even more in the light of current accounts of emotions.

---

[26]See Dreon 2019.

Dewey further developed and partially corrected James's original conception of emotion by emphasizing the continuity between bodily and mental aspects, as well as between emotion and cognition, sensitiveness and appraisal. Dewey also developed James's intuition about the place and role of affective facts by recognizing the pervasive aesthetic, qualitative and affective aspects of our experience. Mead mainly focused on the social dimension of emotions, conceived as basic forms of gestural communication, and his idea about a primary social configuration of emotions seems to represent a challenging issue for current inquiries in affective neurosciences.

To sum up, the pragmatists' anti-Cartesian approach to human behavior in a natural and naturally social environment, and their avoiding any form of mind-centrism—while seriously considering the specificity of mental interactions—represent a thought-provoking position for the current debate in the promising field of so-called 4E cognition. Meanwhile, a comparison with analytically detailed interpretations as well as with currently debated problems and positions proves deeply stimulating for the reconstruction of the pragmatists' account of emotions, because it leads us to clarify the different implications and consequences involved in an overall holistic approach to this issue. As Dewey says in his second essay, "[d]iscrimination, not integration, is the real problem" (Dewey 1971a, p. 179) when having to deal with a continuum.

References

Alexander, Thomas. 2002. The Aesthetics of Reality: The Development of Dewey's Ecological Theory of Experience. In *Dewey's Logical Theory: New Studies and Interpretations*, ed. Burke F. Thomas, Ester D. Micah, and Robert B. Talisse, 3–26. Nashville: Vanderbilt University Press.

Angell, James Rowland. 1904. *Psychology: An Introductory Study of the Structure and Function of Human Consciousness*. New York: Henry Holt and Company.

Baggio, Guido. 2015. *La mente bio-sociale. Filosofia e psicologia in G.H.Mead*. Pisa: ETS.

Calcaterra, Rosa. 2003. *Pragmatismo: i valori dell'esperienza*. Roma: Carocci.

Candiotto, Laura. 2016. Extended Affectivity as the Cognition of Primary Intersubjectivity. *Phenomenology and Mind* 11: 233–241.

Clark, Andy, and David Chalmers. 1998. The Extended-Mind. *Analysis* 58 (1): 7–19.

Colombetti, Giovanna. 2014. *The Feeling Body: Affective Science Meets Enactive Mind*. Cambridge, MA and London: MIT Press.

———. 2017. Enacted Affectivity, Extended. *Topoi*, 36: 445–455.

Cook, Gary A. 1993. *George Herbert Mead. The Making of a Social Pragmatist*. Urbana and Chicago: University of Illinois Press.

Cunningham, Suzanne. 1995. Dewey on Emotions: Recent Experimental Evidence. *Transactions of the Charles S. Peirce Society* 31 (4): 865–874.

Damasio, Antonio. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Penguin Books.

———. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. San Diego: Harcourt.

Darwin, Charles. 1996. *On the Origin of Species*. Oxford: Oxford University Press.

Darwin, Charles. 1998. *The Expression of the Emotions in Man and Animals*. Oxford: Oxford University Press.

De Jaegher, Hanne, and Ezequiel Di Paolo. 2007. Participatory Sense-Making: An Enactive Approach to Social Cognition. *Phenomenology and the Cognitive Sciences* 6: 485–507.

Dewey, John. 1971a. The Theory of Emotion. In *The Early Works*, vol. 4, 152–188. Carbondale and Edwardsville: Southern Illinois University Press.

———. 1971b. The Reflex-Arc Concept in Psychology. In *The Early Works*, vol. 5. Carbondale and Edwardsville: Southern Illinois University Press.

———. 1981. Experience and Nature LW. In *The Later Works*, vol. 1. Carbondale and Edwardsville: Southern Illinois University Press.

———. 1983. Human Nature and Conduct. In: *The Middle Works*, vol. 14. Carbondale and Edwardsville: Southern Illinois University Press.

———. 1984. Affective Thought. In *The Later Works*, vol. 2. Carbondale and Edwardsville: Southern Illinois University Press.

———. 1988a. Qualitative Thought. In: *The Later Works*, vol. 5, 243–262. Carbondale and Edwardsville: Southern Illinois University Press.

———. 1988b. The Vanishing Subject in William James's Psychology. In *The Later Works*, vol. 14, 155–167. Carbondale and Edwardsville: Southern Illinois University Press.

———. 1989. Art as Experience. In: *The Later Works*, vol. 10. Carbondale and Edwardsville: Southern Illinois University Press.

———. 1991. Logic: The Theory of Inquiry, LW. In *The Later Works*, vol. 12. Carbondale and Edwardsville: Southern Illinois University Press.

———. 2004. *Essays in Experimental Logic*. New York: Dover.

Dreon, Roberta. 2007. *Il sentire e la parola. Linguaggio e sensibilità tra filosofie ed estetiche del Novecento.* Milano: Mimesis.

———. 2012. *Fuori dalla torre d'avorio. L'estetica inclusiva di John Dewey oggi.* Genova and Milano: Marietti 1820. French Translation, 2017. *Sortir de la tour d'ivoire. L'esthétique inclusive de John Dewey aujourd'hui.* Paris: Questions Thèoriques.

———. 2015. Emozioni pragmatiste. In *Nuovi usi di vecchi concetti. Il metodo pragmatista oggi*, ed. Maura Striano, Stefano Oliverio, Matteo Santarelli. Milano and Udine: Mimesis.

———. 2019. Emotional and Verbal Gestures. Mead's Investigations on the Sensitive Roots of Language. *Sistemi intelligenti* 3: 115–133.

Eckman, Paul. 1999. Basic Emotions. In *Handbook of Cognition and Emotion*, ed. Dalgleish Tim and Power Mick. Sussex: Wiley.

Fogel, Alan, Nwokah Eva, Jae Young Dedo, Daniel Messinger, K.Laurie Dickson, Eugene Matsusov, and Susan A. Holt. 1992. Social Process Theory of Emotion: A Dynamic Systems Approach. *Social Development* 1 (2): 122–142.

Gallagher, Shaun. 2013. The Socially Extended Mind. *Cognitive Systems Research* 25–26: 4–12. https://doi.org/10.1016/j.cogsys.2013.03.008.

———. 2017. *Enactivist Interventions. Rethinking the Mind*. Oxford: Oxford University Press.

Garrison, Jim. 2003. Dewey's Theory of Emotions: The Unity of Thought and Emotion in Naturalistic Functional 'Co-ordination' of Behavior. *Transactions of the Charles S. Peirce Society* 39 (3): 405–443.

Goldie, Peter. 2002. Emotions, Feelings and Intentionality. *Phenomenology and the Cognitive Sciences* 1: 235–254.

Greenough, William, James Black, and Christopher Wallace. 1987. Experience and Brain Development. *Child Development* 58: 539–559.

Heidegger, Martin. 1967. *Sein und Zeit*. Tübingen: Max Niemeyer.

James, William. 1976. *Essays in Radical Empiricism*. Cambridge: Harvard University Press.

———. 1981. *The Principles of Psychology*, vol. 2. Cambridge and London: Harvard University Press.

———. 1983. *The Physical Basis of Emotion.* In *Essays in Psychology*, 299–314. Cambridge and London: Harvard University Press.

———. 1884. What is an Emotion? *Mind* 9 (34): 188–205.

Joas, Hans. 1997. *G.H. Mead. A Contemporary Re-examination of His Thought.* Cambridge, MA: MIT Press.

Krueger, Joel. 2013. Ontogenesis of the Socially Extended Mind. *Cognitive Systems Research* 25: 40–46. https://doi.org/10.1016/j.cogsys.2013.03.001.

Lange, Carl Georg. 2012. *Über Gemütsbewegungen. Eine psycho-physiologische Studie*. Barsinghausen: Unikum Verlag.

Mead, George Herbert. 1895. A Theory of Emotion from the Physiological Standpoint. *Psychological Review* 2: 399–402.

———. 1934. *Mind, Self and Society*. Chicago: University of Chicago Press.

———. 2011. *Essays in Social Psychology*, ed. M.J. Deegan. New Brunswick and London: Transaction Publishers.

Merleau-Ponty, Maurice. 1942. *La structure du comportment*. Paris: PUF.

———. 1945. *Phénoménologie de la perception*. Paris: Gallimard.

Noë, Alva. 2015. *Strange Tools: Art and Human Nature*. New York: Hill & Wang.

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

Quéré, Louis. 2013. Note sur la conception pragmatist des émotions. In *Occasional Paper 11*. Paris: Institut Marcel Mauss—CEMS.

Scheler, Max. 1923. *Wesen und Formen der Sympathie*. Bonn: Friedrich Cohen.

Stueber, Karsten R. 2006. *Rediscovering Empathy: Agency, Folk Psychology, and the Human Sciences*. Cambridge and London: MIT Press.

Stern, Daniel. 1985. *The Interpersonal World of the Infant*. New York: Basic Books.

Trevarthen, Colwyn. 1979. Communication and Cooperation in Early Infancy: A Description of Primary Intersubjectivity. In *Before Speech: The Beginning of Interpersonal Communication*, ed. Margareth Bullowa, 321–348. London, New York and Melbourne: Cambridge University Press.

Wundt, Wilhelm. 1921. *Völkerpsychologie: Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte*, vol. 1. Stuttgart: Kröner.

5

Getting Warmer: Predictive Processing and the Nature of Emotion

Sam Wilkinson, George Deane, Kathryn Nave and Andy Clark

1   Predictive Processing and Emotion: The Story So Far

Emotion and cognition are typically thought of in contrast to one another, sitting on opposite sides of a divide between passion and reason, the hot and the cold. But what does our best theory of the brain and central nervous system (CNS) tell us about the nature of emotion?

According to an increasingly popular framework in computational neuroscience, the brain is a hierarchically arranged prediction machine (Clark (2013a). Contrary to once-popular feedforward approaches, the brain does not simply take inputs from the outside world, process them, and pass

S. Wilkinson (✉)
University of Exeter, Exeter, UK
e-mail: s.wilkinson@ex.ac.uk

G. Deane · K. Nave
University of Edinburgh, Edinburgh, Scotland, UK

A. Clark
University of Sussex, Brighton, UK

them deeper and deeper into the processing economy. Instead, whenever information from the world impacts on your sensory surfaces, it is already, even at the earliest stages, greeted by a downward-flowing prediction on the part of your nervous system. This prediction comes from your brain's best model of what is going on in the world, and this model is constantly being updated by the mistakes it makes, by the so-called 'prediction error signal', which it constantly tries to keep to a minimum (Lee and Mumford 2003; Rao and Ballard 1999). In recent versions, this signal is weighted according to how reliable or salient the brain estimates the sensory information to be, relative to its best predictions. This 'precision-weighting' device operates at every level of processing. It implements attention, and allows us flexibly to balance top-down prediction and bottom-up sensory information (see Feldman and Friston 2010; Clark 2013b).

The core business of brains like ours, if these stories are on track, is the minimization of precision-weighted errors in the prediction of sensory inputs (see Friston 2005—and for comprehensive reviews, see Hohwy 2013; Clark 2013a). Importantly, the minimization of precision-weighted prediction error isn't always achieved by the brain updating its models of the world (which results in perception and belief). Instead it is sometimes achieved by bringing the world, usually the body, in line with the model (Feldman and Friston 2010; Clark 2016, Chapter 4). The result of this is bodily action.

According to early work in predictive processing (e.g. Lee and Mumford 2003; Friston 2005), what you *perceptually* experience is determined by the model that your brain adopts so as to best predict *exteroceptive* sensory signals such as incoming visual and auditory information. Building on this basic idea, it has recently been suggested (Seth 2013) that what we *emotionally* experience is determined by the model that your brain adopts so as to best predict *interoceptive* signals—signals carrying information about the states of gut, viscera, hydration, vasomotor system, air-supply, muscular system, glucose, and plasma levels, etc.

Here, the predictive processing (PP) account adds important dimensions to the well-known James-Lange model of emotional states as arising from the perception of our own bodily responses to external stimuli and events. The idea there, in a nutshell, was that our emotional 'feelings' are nothing but the perceptions of our own varying physiological responses. According to James it is our interoceptive perception of the

bodily changes characteristic of fear (sweating, trembling etc.) that constitutes the very feeling of fear, giving it its distinctive psychological flavour. From a subjective viewpoint, interoceptive awareness manifests as a differentiated array of feelings including those of 'pain, temperature, itch, sensual touch, muscular and visceral sensations…hunger, thirst, and "air hunger"' (Craig 2003, p. 500). The feeling of fear, if James is right, is thus essentially the detection of an interoceptive physiological signature that has already been induced by exposure to the threatening situation.

A popular (and useful) way to think about James' proposal is to see it as suggesting a kind of 'subtraction test'. This is a thought experiment in which you are invited to subtract all the bodily stuff (detection of your own racing heart etc.) away from the emotional experience, and ask yourself 'what would be left?'. James' claim is that you would be left with nothing that is worth counting as an experience or emotion. What an emotion really *is*, James argument suggests, is the *self-perception of changes in our own bodily states*.

But the standard Jamesian story remains somewhat inadequate. For it seems to require a one-to-one mapping between distinct emotional states and distinctive 'brute-physiological' signatures, and it seems to suggest that whenever the physiological state is induced and detected, the same emotional feeling should arise. Neither of these implications (see Critchley 2005) has been borne out by observation and experiment. The basic story can, however, be refined and extended by adding a 'predictive twist'. Thus Seth (2013) suggests that a neglected core component may be the match (or mismatch) between a cascading series of top-down predictions of our own interoceptive states, and the forward-flowing information contained in sensory prediction error. Our interoceptive predictions, this story suggests:

> arise from multiple hierarchical levels, with higher levels integrating interoceptive, proprioceptive, and exteroceptive cues in formulating descending predictions. (Seth 2013, p. 567)

A single inferential process here integrates all these sources of information, generating a context-reflecting amalgam that is experienced as emotion. Felt emotions thus integrate basic information (e.g. about bodily arousal) with higher-level predictions of probable causes and preparations for possible actions. In this way:

The close interplay between interoceptive and exteroceptive inference implies that emotional responses are inevitably shaped by cognitive and exteroceptive context, and that perceptual scenes that evoke interoceptive predictions will always be affectively coloured. (Seth 2013, p. 563)

Physiologically, the Anterior Insular Cortex is remarkably well-positioned to play a major role in such a process by encoding what Craig (2003, p. 500) describes as 'a meta-representation of the primary interoceptive activity'. Emotion and subjective feeling states arise, this story suggests, as the result of multilevel inferences that combine sensory (interoceptive, proprioceptive, and exteroceptive) signals with top-down predictions to generate a sense of how things are for us and of what we might be about to do. Such a sense of 'action-ready being' encompasses our background physiological condition, estimations of current potentials for action, and the perceived state of the wider world. This delivers a grip upon both the nature and the significance our own embodied state.

Importantly, such a grip must integrate basic information (e.g. about bodily arousal) with higher-level predictions of probable causes. This provides a very natural way of accommodating large bodies of experimental results showing that the character of our emotional experience depends both on the interoception of brute bodily signals and higher-level 'cognitive appraisals' (see Schacter and Singer 1962; Prinz 2004). An example of a brute bodily signal is generic arousal as induced by—to take the classic example from Schacter and Singer—an injection of adrenaline. Such brute signals combine with contextually induced 'cognitive appraisals' leading us to interpret the very same bodily 'evidence' as either elation, anger, or lust according to our framing expectations.

2   Emotions as "Constructs" (Models)

The account of emotion just sketched fits perfectly with the *theory of constructed emotion* (Barrett 2017). This mechanizes Barrett's preceding *conceptual act theory* (Barrett 2014) within a PP framework. The

central claim is that in each waking moment the brain is integrating past experience to generate concepts to guide actions and give meaning to sensations. When the generated concepts involved relate to physiological imperatives, your brain constructs instances of emotion.

Following from the accounts of emotion in the PP literature, each instance of an emotion arises as a categorization of bodily signals, according to context, in terms of past experiences:

> When past experiences of emotion (e.g. happiness) are used to categorize the predicted sensory array and guide action, then one experiences or perceives that emotion (happiness). (Barrett 2017, p. 9)

The theory of constructed emotion makes a sharp distinction between emotion *instances,* and emotion *categories.* An emotion *instance* is the in-the-moment construction of an emotion given the current context. What we usually describe as an emotion, (e.g. fear) is better described as an emotion 'category', which unifies diverse and highly variable instances under a single classificatory umbrella (Clark-Polner et al. 2016). Emotion categories, Barrett argues, do not exist in nature—they are assigned according to functional and socially constructed roles. Motivation for this view comes from what has been dubbed the "emotion paradox" (Barrett 2006). The emotion paradox refers to the fact that while the existence of emotions such as "sadness", "anger", "happiness" is assumed by the scientific community and supported by common sense, the empirical literature calls into question this assumption due to the absence of any signature—be it a facial expression, physiological response or neural activity—that reliably indexes *any* emotion category. This leads to Barrett's claim that emotion categories are collections of diverse instances that are clumped together in terms of their functional role, lacking dedicated facial expressions, physiological responses or neural signatures. Barrett states:

> Emotion categories are as real as any other conceptual categories that require a human perceiver for existence, such as 'money' (i.e. the various objects that have served as currency throughout human history share no physical similarities). (Barrett 2017, p. 13)

This many-to-one mapping of physical states to emotion categories—called 'degeneracy'—is the primary argument behind the lack of any kind of emotional "essence". Degeneracy is borne out by the empirical literature. A meta-analysis of facial expressions indicates that many different facial expressions can be observed for the same category, and many different emotional categories can be understood by the same facial expression (Durán et al. 2017)—the meaning of a facial expression largely depends on context. Physiological signatures for any emotion category have proved to be similarly elusive, with a recent meta-analysis (Siegel et al. 2018) showing that there are no physiological signatures that reliably correspond to any one emotion category—for instance, when you're angry, your blood pressure can go up, down, or remain the same. On Barrett's view the determining factor is what kind of *action* the brain is preparing the body for—getting ready to fight requires recruitment of different resources than some other anger-related course of action, despite the emotion categorization ('anger') being the same (Barrett 2017). Similarly, a meta-analysis on the neurophysiological basis of emotion categories are not contained within any one brain region or system, but are represented as configurations across multiple brain networks (Wager et al. 2015).

From the perspective of evolution, degeneracy in the brain makes sense as an adaptive engineering principle. A key result of degeneracy is that a single brain can create a vast number of spatiotemporal patterns. These high complexity systems are preferred by natural selection as they can reconfigure themselves into a multitude of different states (Whitacre and Bender 2010). This reconfiguration ability is what makes our brains, on this account, radically flexible according to culture and environment.

Emotions, then, are not reactions to the world, not even *interoceptively informed* reactions to the world. Rather, they are out-and-out constructions of the world. Emotions are constructed in just the same way that percepts are constructed; that is, they are predictive models of the likely causes of the sensory input, made by re-stitching together past experiences and then classifying the current experience as an amalgam of past experiences of a similar nature. These emotional predictions are made always in the service of regulating the body's internal milieu, that

is, in the service of *allostasis* (Barrett and Simmons 2015; Barrett 2017). Predictive processing, Barrett suggests, provides the mechanism underlying these categorizations.

On this more 'action-oriented' predictive processing account, the top-down flow of predictions anticipates (1) upcoming interoceptive and exteroceptive signals and (2) the best action or bodily response to deal with the upcoming sensory flow. In order to create these 'concepts' (embodied, whole-brain representations), the brain creates predictions by using past experience to answer "*What is this new sensory input most similar to?*" (Barrett 2017). The incoming sensory evidence, in the form of prediction error, helps to select and shape the distributions of predictions that are activated that best fit the sensory array, thereby minimizing prediction error—resulting in a *categorization* of the incoming sensory information in terms of past experiences (Barrett 2006). That means that the predictions activated in the present are an instance of what Barsalou refers to as '*ad hoc*' concepts (Barsalou 1983). In the brain, a concept looks like a distributed pattern of activity across populations. These ad hoc concepts or predictions, that categorize present sensory flux in terms of past experience, are the mechanism of construction of any given instance of emotion. This predictive cascade—the interpretation of the sensory flux in terms of its expected utility to allostasis—is the process of meaning-making in the brain.

Notice also that emotion and cognition are here performed in exactly the same way, that is, in reference to allostasis, and sensory inputs (prediction error) are used as information to guide the sculpting of concepts that engender adaptive action. This process is an approximation of Bayesian inference (Denève 2008) to decide among which simulation (interlocked web of predictions) should be implemented in order to maximize allostatic efficiency across multiple body systems (e.g. need for glucose, oxygen, salt, etc.), and activate appropriate metabolic expenditure in the service of action (tiger, run!).

Barrett's theory is supplemented with a compelling neurobiological implementation story, where the default mode network represents efficient, multimodal summaries, which, when activated, cascade through the entire cortical sheet, terminating in primary sensory and motor regions. The cascade as a whole is an instance of a concept, or

an emotion (Barrett 2017). That said the link between the neurobiology and the conceptual argument is not altogether clear: the empirical evidence is open to interpretation and amenable to other conceptual theories of emotion (including other conceptual theories with PP as the underlying mechanism).

The theory of constructed emotion offers a plausible account of how diverse instances of emotion come to be placed together under unifying conceptual umbrellas. It also fleshes out how emotion categories are cleaved apart according to context, and how the categories are more socially determined conceptual categories than categories existing in nature. Furthermore, the theory partially fleshes out the conception of emotion as interoceptive inference, both with a more specific mechanism of diverse instances of emotion, and in setting out how different emotion categories come to be formed.

3   From Embodied Emotion to Embodied Valence

So how do we make sense of affective value or valence? What determines the evaluative dimension of an emotion instance? Here is an initial approach we might take to accounting for valence in terms of the properties of an action-oriented predictive processing system.

The core imperative of a predictive processor is the successful prediction of incoming sensory evidence. Thus it may initially seem that the successful minimization of prediction error should be what determines an overall state of positive valence. Though this may seem promising at first, such a proposal quickly falls apart. Any account of valence that is state-based, that equates positive valence to a state of minimized prediction error, fails to do justice to the fact that prediction error minimization is necessarily a dynamic and continuous process, constantly engaging action, and designed to account for the on-going maintenance of an organism in an ever-changing world. Only from this perspective can we avoid the 'dark room' objection to predictive processing (Friston et al. 2012b). This states that if my goal is solely the minimization of

prediction error, then surely I should just seek out a dark, empty room and stay there. Perfect prediction, it seems, is attainable by avoiding action and practising sensory (and nutritional) deprivation until death. Such a policy is, of course, wholly inconsistent with the actual behaviour of living things.

An initial response to this might be that the various demands of survival (as ultimately signalled in the form of prediction error) would move you onwards. But note that even were your dark room to come equipped with a life support machine (consider an unending night in an abandoned hospital ward) it is unlikely that you would find this to be an endlessly pleasurable experience. Humans not only find a lack of novel stimulation boring, they actively seek out and take delight in a rich repertoire of aesthetic, humorous, or thrilling situations, from skydiving to stand-up, that are specifically engineered to generate a rush of prediction error through the violation of prior expectations.

A more promising strategy is as follows. Instead of tying valence to the achievement of some particular error-minimized state, Joffily and Coricelli (2013) propose a dynamic alternative in which valence is taken to be the *rate* at which this error is being reduced. In mathematical terms valence is recast as the first time-derivative of error: a matter of *velocity*, rather than position. We seek out surprising states, then, in as much as they offer us the opportunity to engage in a faster (rather than slower) rate of reduction in prediction error. Drawing on Carver and Scheier's (1990) control theoretic account of emotion, Van De Cruys (2017) improves and extends this story by suggesting that, rather than being straightforwardly a matter of a positive rate of error reduction, pleasure (positive valence) occurs when our actual rate of error reduction is higher than we had predicted it would be. If it is lower, we experience negative valence.

An upshot of explaining valence in terms of these processing characteristics, rather than specific content, is that it is no longer tied to any particular set of causes, error modality, or level of inference. We can thus describe a relationship between allostasis and valence that is not constrained (as it is in Seth 2013) to inference over the causes of interoceptive signals alone. This seems like the correct route to take. Homeostatic maintenance is served not only by the direct monitoring

and regulation of physiological variables, but also indirectly, by the anticipatory regulation of our external environment. Whether intero- or exteroceptive, persistent unreduced prediction error is a sign that we do not have a grip on our self or surroundings, and adjustments need to be made.

Furthermore, tying valence to the regulation of exteroceptive error reduction rate allows us to characterize more 'cognitive' experiences of positive valence—those that are not easily described in terms of basic physiological reactions—such as responses to art, narrative, or humour. These can now be understood as achieving their emotional effects by engineering pleasurable trajectories in the creation and violation of expectations, followed by the subsequent pleasurable release in the eventual reduction of resulting prediction error. This fits nicely with descriptions of humour, as resulting from the creation and resolution of tension (Sroufe and Waters 1976) and, as Van De Cruys and Wagemans (2011) suggest, provides a potential explanation of the failure of aesthetic principles (such as harmony, fluency, or balance) to account for the success of celebrated works of art which regularly display the deliberate violation of such rules.

4   Emotion and Cognition

Summing up the previous sections, what predictive processing reveals is a world permeated by affect—a world of opportunities for action, geared to current tasks, modulated by information about our own bodily states. But to see just how radical the PP picture turns out to be, we still need to add one final ingredient. It's that PP rejects the picture of emotion and cognition as fundamentally different kinds—at least insofar as they are causally active parts of the cognitive machinery.

According to a popular view, often associated with Hume (1739/2007), a fundamental divide among all things mental is one that divides the informational and the motivational. The former is about the organism ("coldly") coming to a view about what's going on in the world, whereas the latter is about ("hotly") driving the organism to bring about change in the world. Hume's central point was that without

the latter, without passions, no action would ever take place. A hypothetical creature only capable of having informational states would stay still, inert, and unmoved to do anything, regardless of what it learnt about the world. In this sense, according to Hume, emotion (affect, passion) broadly construed, is the driving force behind all action, but completely distinct from belief (and insulated from "reason").

The idea that informational states on the one hand, and motivational states on the other hand, are fundamentally different kinds of state whose interaction is required to bring about action, is widely embraced in daily life. It forms not only a core part of common-sense (or 'folk') psychology, but is deeply embedded in some more scientific frameworks too. Statistical decision-theory (including neuroeconomics and work on reinforcement learning) inherits this Humean picture, since in standard realizations it works with a firm separation between encodings of value or 'utility' and encodings of probability. In these frameworks, decisions are made and actions selected only when utility and probability align, revealing viable opportunities for worldly interventions that deliver weighted rewards at calculated costs (for a useful review, see Sanfey et al. 2006).

By contrast, PP posits only predictions, informed by multiple inner and outer sources of information. In PP motivational states are realized as predictions about our own future actions and states. To see how, let's return to the PP treatment of action. Action is making the world conform to some of your predictions, and is just another way of reducing long-term prediction error. At the bottom level, PP makes sensori-muscular (proprioceptive) prediction into a proxy for motor commands (Shipp et al. 2013). Predicting the flow of sensori-muscular effects that would occur if you hit the tennis ball just right actually brings the 'good hitting' about. In a little more detail, the brain predicts the flow of states of muscle spindles, tendons, and joints that the action demands, and the resulting errors (since those states are not yet actual) are systematically quashed by moving the body so as to make that flow of prediction come true. This is an elegant and economical means of delivering basic motor control (see e.g. Shipp et al. 2013).

PP deploys the same kind of story 'all the way up'. Our action-guiding proprioceptive predictions are themselves caused by

even higher-level and longer time-scale predictions—predictions about our own future behaviours and future states. These entrain actions when good opportunities arise (see Pezzulo et al. 2015). The picture is of nested beliefs that entrain actions by bringing about predicted sensory flows. For example, suppose I believe/predict that I will meet you at 7:00 at the movie-theatre. This (combined with prior knowledge and any newly gleaned information) leads me to believe/predict that I will get the 6:30 bus. That last prediction then acts as a kind of mini-policy that enslaves motor action (by means of proprioceptive predictions) when it is time to leave the house.

Simple action-entraining motor intentions here cash out as precise proprioceptive predictions, while higher-level intentions, including standing goals, are realized by higher-level predictions of whole swathes of sensory information, which likewise entrain actions (by yielding precise proprioceptive predictions) when they themselves are assigned high enough precision. These nested, interacting predictions arise and dissolve—in ways that realize the phenomenological flux of shifting drives and desires—as we move around the world, acting and harvesting new sensory information. If PP is on track, the causally potent play of human motivation is not an illusion—but it is realized using only the common currency of multilevel, multi-area prediction. In this picture, prior beliefs (resulting in predictions) combine with sensory evidence to bring about action. This is just the bedrock (Bayesian) move—one that turns everything into a form of prediction-based inference.

It has been suggested (Holton 2016; Klein 2018) that this picture is too impoverished to be a satisfying story about human minds. Part of their reasoning is roughly Humean. The Humean worry is that beliefs (or predictions) without motivations are inert, unable to mandate actions. That's already taken care of by the PP story though, since high-precision predictions that have proprioceptive (hence motoric) consequences are immediately poised to entrain actions to make themselves come true. Holton also worries that assimilating desires to predictions "doesn't do justice to the multiplicity and malleability of human desire" (2016, p. 10) noting that we need to accommodate cases where we desire X and may even do X while believing that X won't bring us happiness or pleasure. However, PP accommodates this very simply, by

separating predictions about the hedonic consequences of actions from the full set of predictions that interactively entrain actions. Specifically, the predictive processing story firmly distinguishes (Friston et al. 2012a) between sub-personal action-entraining high-precision predictions concerning what I will do and predictions of the hedonic (interoceptive) outcomes of those very actions. PP thus accommodates the fact, highlighted by Holton, that drug users often do not believe/predict that taking the drugs will actually lead to happiness. But what they do predict is seeking and ingesting the drug. PP thus easily reconstructs the useful distinction between 'wanting' and 'liking' (Berridge 2007). The PP picture thus turns out to be a neat fit with important work on the nature and mechanisms of addiction (Berridge 2007; Friston et al. 2012a). More generally, even given that the addict need not predict that the drugs will bring pleasure, PP remains poised to explore a wide variety of promising accounts in which the addict's experiences and actions are the results of interacting sub-personal (non-conscious) predictions.

This replaces Hume's two interacting kinds (reason and passion) with a picture of large numbers of subtly different and modifiably interacting elements. All of those elements are somewhat belief-like (consisting in predictions) but somewhat desire-like too (as they help select and entrain actions at multiple time-scales). So, while it may look like a simplifying move, what PP finally delivers will in fact be a far richer palette for explaining human behaviour. That palette allows a full spectrum of possibilities that reach far beyond the simple, constrained interactions suggested by crude folk psychological distinctions between 'cognition' and 'conation'.

We have seen how this collapses belief and desire, and desire is often construed as a "hot" or "impassioned" state, but it is clearly a mistake to equate emotion with desire. As several theorists have noted (e.g. Marks 1982; Oakley 1992), emotion has both belief-like and desire-like elements. Experiencing fear of the spider simultaneously tells you about the world (e.g. that there is a spider there), while also motivating you to act in a certain way (run away from said spider). But whereas the standard way of thinking of emotions is as *composed of* these belief and desire-like elements, PP construed things very differently. Just because the belief and desire-like elements can be "read off" the emotional state, it is

not to say that psychologically (or indeed ontologically) they are somehow the primitive building blocks of a hybrid and less primitive state called emotion. On the contrary, according to PP, it is emotional processes, which simultaneously inform and move, that are primitive, and, in predictive processing terms, this is all fleshed out in the common currency of predictions and predictive models: predictions generated by complex hierarchical models concerning, in an interconnected manner, the organism, the world, and the organism's place in that world.

The same point can be made in terms of "direction of fit". To use Anscombe's (1957) example, a list of items written on a piece of paper can either be seen as an inventory or as a shopping list. If it's the former, it has a list-to-world (or descriptive) direction of fit. The inventory's function is to fit the world (e.g. match what is currently in the store cupboard), and if it fails to do so, it's the inventory's "fault", so to speak. If it's the latter (viz. a shopping list), it has a world-to-list (or directive) direction of fit. The shopping list's role is not to match the world as it is, but to bring the world in line with it in the future. This notion of direction of fit has been widely applied to representations in general, and representational mental states in particular.

Whereas it has been common to think of beliefs, with their mind-to-world (or descriptive) direction of fit, and desires, with their world-to-mind (or directive) direction of fit, as being the fundamental building blocks of the mind, what is actually fundamental in the PP architecture is prediction, which can vary across a spectrum as to the extent to which it should be fulfilled by the world (perception/belief) or the self (action/desire). This means that pure belief (or cold perception), or pure desire (or blind action), is actually a theoretical ideal, one that is difficult or impossible to achieve. What we are actually left with is a wide variety of what Millikan (1995) calls "pushmi-pullyu representations" (PPRs), which simultaneously describe and direct. Millikan herself explains these as follows.

> PPRs have both a descriptive and a directive function, yet they are not equivalent to the mere conjunction of a pure descriptive representation and a pure directive one but are more primitive than either. Purely descriptive and purely directive representations are forms requiring a

more sophisticated cognitive apparatus to employ them than is necessary for these primitives. Purely descriptive representations must be combined with directive representations through a process of practical inference in order to be used by the cognitive systems. Purely directive representations must likewise be combined with descriptive ones. The employment of PPRs is a much simpler affair. (1995, p. 186)

There are many examples of PPRs, especially in the domain of animal signals. The hen's food call to her brood simultaneously says, "There's food here" and "Come get it". We see the general insight, namely, that PPRs are primitive, as fitting extremely nicely with our idea that emotions, as viewed within the predictive processing framework, are just these kinds of bi-directional PPRs, and are similarly primitive. The immediate grip of fear tells you simultaneously "There is danger" and "Run!". This also fits nicely with pre-theoretical intuitions about the primacy of emotion and affect. Both phylogenetically and ontogenetically, emotion and affect comes first, whereas reason (however that is cashed out) comes much later. If the basic predictive processing mechanism is truly how our embodied nervous systems work, then this is all pleasantly unsurprising.

5   Concluding Remarks

Emotions, we have argued, are built from predictions. They reflect inner and outer sources of information, combined in flexible ways, and are answerable to the full world knowledge (generative model) of an agent. But they are not a special cognitive kind. Instead, they are part and parcel of an integrated processing system whose core functionality is to reduce precision-weighted prediction error by maintaining dynamic engagements with the world. These engagements display trajectories both marked and determined by valence, where positive valance reflects better-than-predicted slopes of error minimization. What emerges is a picture of mind as an action-oriented system all of whose states are somewhat belief-like, and somewhat desire-like too.

Another way of looking at this is as follows. In so far as full-blown emotions as we typically understand them are the most prominent and consciously detectable (and hence categorized) of these action-oriented states, one could say that PP renders emotion, construed more broadly to include even the very subtlest of these, ever-present in cognition. In other words, the embodied predictive mind is, by necessity, an emotional mind.

References

Anscombe, G.E.M. 1957. *Intention*. Cambridge: Harvard University Press.

Barrett, L.F. 2006. Solving the Emotion Paradox: Categorization and the Experience of Emotion. *Personality and Social Psychology Review* 10: 20–46.

———. 2014. The Conceptual Act Theory: A Precis. *Emotion Review* 6 (4): 292–297.

———. 2017. The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization. *Social Cognitive and Affective Neuroscience* 12 (1): 1–23.

Barrett, L.F., and W.K. Simmons. 2015. Interoceptive Predictions in the Brain. *Nature Reviews Neuroscience* 16: 419–429. https://doi.org/10.1038/nrn3950.

Barsalou, L.W. 1983. Ad hoc Categories. *Memory and Cognition* 11: 211–277.

Berridge, K.C. 2007. The Debate Over Dopamine's Role in Reward: The Case for Incentive Salience. *Psychopharmacology (Berl)* 191: 391–431.

Carver, C.S., and M.F. Scheier. 1990. Origins and Functions of Positive and Negative Affect: A Control-Process View. *Psychological Review* 97 (1): 19–35.

Clark, A. 2013a. Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioral and Brain Sciences* 36 (3): 181–204.

———. 2013b. Are We Predictive Engines? Perils, Prospects, and the Puzzle of the Porous Perceiver. *Behavioral and Brain Sciences* 36 (3): 233–253.

———. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

Clark-Polner, E., T.D. Johnson, and L.F. Barrett. 2016. Multivoxel Pattern Analysis Does Not Provide Evidence to Support the Existence of Basic Emotions. *Cerebral Cortex* 27: 1944–1948. https://doi.org/10.1093/cercor/bhw028.

Craig, A.D. 2003. Interoception: The Sense of the Physiological Condition of the Body. *Current Opinion in Neurobiology* 13: 500–505. https://doi.org/10.1016/s0959-4388(03)00090-4.

Critchley, H.D. 2005. Neural Mechanisms of Autonomic, Affective and Cognitive Integration. *Journal of Comparative Neurology* 493: 154–166. https://doi.org/10.1002/cne.20749.

Denève, S. 2008. Bayesian Spiking Neurons I: Inference. *Neural Computation* 20: 91–117. https://doi.org/10.1162/neco.2008.20.1.91.

Durán, J.I., R. Reisenzein, and J.-M. Fernández-Dols. 2017. Coherence Between Emotions and Facial Expressions: A Research Synthesis. In *The Science of Facial Expression*, ed. J.-M. Fernández-Dols, and J. A. Russell, 107–129. New York, NY: Oxford University Press.

Feldman, H., and K.J. Friston. 2010. Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience* 4: 215. https://doi.org/10.3389/fnhum.2010.00215.

Friston, K. 2005. A Theory of Cortical Responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 360: 815–836. https://doi.org/10.1098/rstb.2005.1622.

Friston, K., T. Shiner, T. FitzGerald, J.M. Galea, R. Adams, et al. 2012a. Dopamine, Affordance and Active Inference. *PLoS Computational Biology* 8 (1): e1002327.

Friston, K., C. Thornton, and A. Clark. 2012b. Free-Energy Minimization and the Dark-Room Problem. *Frontiers in Psychology* 3: 130. https://doi.org/10.3389/fpsyg.2012.00130.

Hohwy, J. 2013. *The Predictive Mind*. UK: Oxford University Press.

Holton, R. 2016. Review of Surfing Uncertainty: Prediction, Action, and the Embodied Mind. *Times Literary Supplement* 7 (October): 10–11.

Hume, D. 1739/2007. *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. Oxford: Oxford University Press.

Joffily, M., and G. Coricelli. 2013. Emotional Valence and the Free-Energy Principle. *PLoS Computational Biology* 9 (6): e1003094.

Klein, C. 2018. What Do Predictive Coders Want? *Synthese* 195 (6): 2541–2557.

Lee, T.S., and D. Mumford. 2003. Hierarchical Bayesian Inference in the Visual Cortex. *Journal of the Optical Society of America A. Optics and Image Science* 20: 1434–1448. https://doi.org/10.1364/JOSAA.20.001434.

Marks, J. 1982. A Theory of Emotion. *Philosophical Studies* 42 (1): 227–242.

Millikan, R. 1995. Pushmi-Pullyu Representations. *Philosophical Perspectives* 9: 185–200.

Oakley, J. 1992. *Morality and the Emotions*. London: Routledge and Kegan Paul.

Pezzulo, G., F. Rigoli, and K. Friston. 2015. Active Inference, Homeostatic Regulation and Adaptive Behavioural Control. *Progress in Neurobiology* 134: 17–35.

Prinz, J. 2004. *Gut Reactions: A Perceptual Theory of the Emotions*. Oxford: Oxford University Press.

Rao, R.P., and D.H. Ballard. 1999. Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-Field Effects. *Nature Neuroscience* 2: 79–87. https://doi.org/10.1038/4580.

Sanfey, A.G., G. Loewenstein, S.M. McClure, and J.D. Cohen. 2006. Neuroeconomics: Cross-Currents in Research on Decision-Making. *Trends in Cognitive Sciences* 10: 108–116. https://doi.org/10.1016/j.tics.2006.01.009.

Schachter, S., and J. Singer. 1962. Cognitive, Social, and Physiological Determinants of Emotional State. *Psychological Review* 69 (5): 379–399.

Seth, A.K. 2013. Interoceptive Inference, Emotion, and the Embodied Self. *Trends in Cognitive Sciences* 17 (11): 565–573.

Shipp, S., R. Adams, and K.J. Friston. 2013. Reflections on Agranular Architecture: Predictive Coding in the Motor Cortex. *Trends in Neurosciences* 36 (12): 706–716.

Siegel, Erika H., Molly K. Sands, Wim Van den Noortgate, Paul Condon, Yale Chang, Jennifer Dy, Karen S. Quigley, and Lisa Feldman Barrett. (2018). Emotion Fingerprints or Emotion Populations? A Meta-Analytic Investigation of Autonomic Features of Emotion Categories. *Psychological Bulletin* 144 (4): 343–393.

Srofe, L.A., and E. Waters. 1976. The Ontogenesis of Smiling and Laughter: A Perspective on the Organization of Development in Infancy. *Psychological Review* 83 (3): 173–189.

Van de Cruys, S. 2017. Affective Value in the Predictive Mind. In *Philosophy and Predictive Processing*, Chapter 24. ed. T.K. Metzinger, and W. Wiese. Frankfurt am Main: MIND Group.

Van de Cruys, S., and J. Wagemans. 2011. Putting Reward in Art: A Tentative Prediction Error Account of Visual Art. *i-Perception* 2 (9): 1035–1062.

Wager, Tor D., Jian Kang, Timothy D. Johnson, Thomas E. Nichols, Ajay B. Satpute, Lisa Feldman Barrett, and Jörn Diedrichsen. (2015). A Bayesian Model of Category-Specific Emotional Brain Responses. *PLOS Computational Biology* 11 (4): e1004066.

Whitacre, J., and A. Bender. 2010. Degeneracy: A Design Principle for Achieving Robustness and Evolvability. *Journal of Theoretical Biology* 263 (1): 143–153.

6

Emotional Reflexivity in Reasoning: The Function of Describing the Environment in Emotion Regulation

Dina Mendonça and João Sàágua

1 Introduction

The chapter shows how meta-emotions are crucial for understanding the role of emotions in reasoning. The fact that emotions can be about emotions gives rise to meta-emotions and makes reflexivity of emotions a part of the emotional landscape (Mendonça 2013). Meta-emotions have yet to be fully explored and consequently their role in reasoning has not been widely debated. Once reflexivity of emotions (the way in which a subject can feel emotions about their other felt emotions) is theoretically recognized, it becomes clearer that part of the regulatory processes are performed by emotions in addition to the way in which reason and thinking regulate emotional processes. The suggestion of the chapter is that to fully understand the role of emotion in its regulation requires an accurate description of the emotional situation that includes the presence of emotions and meta-emotions in connection to reasoning because only this complete description can provide a coherent whole for integrating various interpretations of a situation. Once the

D. Mendonça (✉) · J. Sàágua
IFILNOVA, FCSH NOVA, Universidade Nova de Lisboa, Lisbon, Portugal

fully integrated picture of a situation is attained, it makes understanding how it is subsequently modified clearer, and reveals some of the more intricate details of how the emotional regulatory mechanisms work.

Our conception of rationality would be insane if it did not include feelings, emotions, and sentiments (Williams 1981, p. 29). Though traditionally the capacities for deliberation and judgment have been taken as more rational than the capacity for emotion because it is assumed that "any other mental state (such as the emotions) that conflicts with the outcomes of deliberation and judgment must *ipso facto* be irrational," (Helm 2000, p. 4), it is nowadays completely clear that the exercise of reasoning is no longer in opposition to emotion (Elster 1996; Greenspan 2004; Damasio 2003; Kirman et al. 2010), and "there is also growing evidence that judgments and decisions based on integral affect also tend to be more consistent interpersonally" (Pham 2007, p. 165).

Nevertheless, to acquire a complete picture of the nature of reasoning given this new position of emotion and reason it is not sufficient to simply add emotions to rationality, and it is important to acknowledge that "any categorical statement about the overall rationality or irrationality of emotion may be simplistic and misleading" (Pham 2007, pp. 155–156). The integration of emotion in rationality requires a reconceptualization of rationality and a deeper understanding of how cognition incorporates and interacts with emotion. However, it is also the case that we can still find in the literature examples of a certain kind of opposition between emotion and rationality (Leahy 2002, pp. 182–183), and consequently to totally overcome this misconception the final integrated picture of rationality will also have to provide an explanation for the long historical legacy of thinking of emotions as opposed to reason, as well as a justification to why the "emotional person is not irrational, but ecologically rational" (Pham 2007, p. 172).[1] This means the outcome will provide an insightful description as to why emotions sometimes appear

---

[1]The notion of ecological rationality takes rationality to be the result of the adaptive fit between the human mind and the environment. More specifically, "The concept of ecological rationality suggests three basic tenets regarding decision making. First, the mind's decision strategies are adapted to particular environments. (…) Second, in certain environments, simple decision strategies are able to compete with complex strategies—less is (sometimes) more. Third, humans largely respond adaptively to task and environment characteristics" (Mata et al. 2012, p. 1).

to be in conflict with some types of thought processes, and why some systems can be inherently selfish while others are distinctively social and moral.

One way to contribute to the reconceptualization of rationality is by identifying the complexity of emotions and rationality. While some researches identify different notions of rationality at play such as the notion of ecological rationality (Pham 2007, pp. 156–159), other researchers aim to capture the way "emotions are complex, multidimensional, and multilayered" (Hofmann 2013, p. 94). The present chapter adds another step in this theoretical effort by showing how emotions and meta-emotions function in reasoning and decision-making, which provides added dimensions to be described in processes of reasoning and decision-making. This added layer of descriptive power empowers and refines processes of emotion regulation while it promotes a self-corrective procedure for future experience.

2 Meta-Emotions

When emotions are about emotions, they are layered instead of sequential (Pugmire 2005, p. 174). Emotions are layered when they take other emotions as their object. For example, when someone is sad about their own shame, their sadness is a meta-emotion and the emotion of sadness is an added emotional layer on the emotional landscape of the situation. If sadness did not have shame as an object of the emotion, and was instead an emotion experienced after the experience of shame, sadness would appear in a sequential way, and though it might nevertheless interfere with an experience of shame, it would not be about it. Thus, emotions can appear in a sequence when they are about different things as when someone feels shame about something they have done and then feel sad about the subsequent consequences of their action; and emotions can appear in a layered manner, then emotions are about emotions.

The concept of meta-emotion first appears in Gottman, Katz, and Hoven's work to refer to parents' emotions and beliefs regarding their own and their children's emotions (Gottman et al. 1996, 1997). Since then researchers have defined meta-emotions in a variety of ways

(Haradhvala 2016, p. 1). Some define meta-emotions as the beliefs about emotions about primary emotional processes (e.g., Beer and Moneta 2010; Ferrari and Koyama 2002), others define them as emotions which have other emotions as their object (Jäger and Bartsch 2006; Mendonça 2013), others label them as evaluative cognitions and emotions about one's primary emotions (Bartsch et al. 2008, 2010), others as a pre-emotion, which underlies all of the emotional landscape and is a condition for all emotions felt (Belli and Broncano 2017), and others identify them as a set of strategies used to act on emotional information (Koven 2011). The concept has been present in the literature and we can find various references to them and as Hoffman writes, "emotional experience have long been recognized and described as meta-experiences of emotions" (Hofmann 2013, p. 95). Meta-emotions have been employed in conceptual discussions and can also be found in the context of therapy, such as in Greenberg's Emotions Focused Therapy (Greenberg 2011) which distinguishes between primary and secondary emotions, where "[s]econdary emotions are responses to those primary emotions or cognitions rather than to the eliciting situation or event per se," (Hofmann 2013, p. 95) which enables the distinction between emotions triggered by trauma or traumatic events, and those emotions that are felt about the emotional reaction toward such events.

In the philosophical literature the concept first appears in a paper by Annette Baier "What Emotions Are About" when she writes that, "[e]motions with this sort of very comprehensive object will be meta-emotions, evaluative reactions to the sum total, the sequence or accumulation to date of one's admirations, amusements, envies, indignations, loves, griefs" (Baier 1990, p. 24). Also, in "The Pleasures of Tragedy" Susan Feagin (1995) further explores meta-emotions by pointing out the difficulty of distinguishing emotions from meta-emotions because the vocabulary does not separate the different layers and the different layers are phenomenally intertwined. Meta-emotions intentionality is also analyzed in the philosophical literature (Elster 1999a, pp. 106, 116, 255, 316; 1999b, p. 109) showing that "when an emotion generates a meta-emotion that induces a subsequent belief change" (Elster 1999a, p. 316).

Though the existence of meta-emotions is undeniable and they have been identified in the literature at least since the 1990', attention to the

phenomenon of meta-emotion has only recently become a strong focus of interest for philosophers and psychologists (Jäger and Bartsch 2006; Mitmansgruber et al. 2009; Hoffmann 2013; Mendonça 2013; Jäger and Bänninger-Huber 2015; Norman and Furnes 2016; Howard 2017). Despite many problems and issues surrounding the understanding of meta-emotions (Mendonça 2013, p. 396),[2] such as what are their limits and boundaries and how they are indistinguishable from first order emotions in language and in phenomenology, and how they relate to consciousness (Mendonça 2013), the recent focus on meta-emotions has shown that they provide crucial insight about emotional processes, and that they cannot be simplistically taken as a special case of emotion that simply happens to have another emotion for its object. The reason for this is that reflexivity modifies the nature of our emotional world (Mendonça 2013), and no theory of emotion is complete without acknowledging and working out and incorporating the meta-emotional layer (Mendonça 2013; Howard 2017; Belli and Broncano 2017).

Reflexivity of emotions changes the meaning and value of the first order emotions: being angry about being sad and being proud about being sad is a completely different emotional experience of sadness. Since meta-emotions have an impact on the value of the first-order emotion which modifies them and promotes a change in the whole emotional experience (Mendonça 2013, p. 394), their importance lies partly on the way in which they influence and mold the impact of the first order emotional experience (Mendonca 2013, p. 390). That is, the added significance brought by the meta-emotion is not an addition to the meaning of the first-order emotion because it has a transformative effect and as a result the "information obtained with the description of meta-emotions is not simply a matter of having more information about the experience; the extra knowledge we get from meta-emotions may change the meaning of the experience altogether" (Mendonça 2013, p. 394).

---

[2]We would like to thank Phillip Gerrans for raising the problematic issue of amount of layers of meta-emotions' layers and the necessity to establish a limit, and Pascal Engels for the suggestion that the issue may be handled similarly to belief revision at the question period of the Conference "Feeling Reasons" at the University of Edinburgh (May 2017).

Both philosophers and psychologists identify meta-emotions as human mechanisms to establish a healthy emotion regulation making them a crucial element for emotion regulation. Meta-emotions have been shown to be important in considering regulation of emotion because "[p]erceiving an emotion as problematic, aversive, or unacceptable, instead of normal, comprehensible, and acceptable can influence the way a person regulates the emotional state itself" (Couyoumdjan et al. 2016, p. 1), and this is why they are consequently a decisive part of self-appraisal (Jager and Banninger-Huber 2015). However, it is also the case that meta-emotions "might produce vicious circles and rebound effects" (Mitmansgruber et al. 2009, p. 449), and this can ultimately exert a powerful influence on emotional health and equilibrium though not necessarily on the direction of desirable mental health (Mitmansgruber et al. 2009). Thus, though meta-emotion can easily be interpreted as strategies and tools for emotional health and emotion regulation, they can also be "prime contributors to the mess inside us" (Howard 2017, p. 20). Howard describes how

> iterated intentional turnabouts can rapidly lead to metaemotional pileups: situations in which one's emotions about one's emotions have become so self-devouring and clogged with rationalizations that their intentional and rational structure is impossible to untangle. (Howard 2017, p. 16)

That is, recognizing meta-emotions as strategies for healthy emotion regulation also means accepting that there is no trait in reflexivity that guarantees the adaptive and healthy outcome, and that the relationship between emotions and meta-emotions can be far more complex and negative that the usual positive regulative connections (Howard 2017, pp. 11–15), making it crucial to identify the ways in which regulation can be maladaptive.

3   Meta-Emotion and Emotion Regulation

Even though reflexivity does not necessarily award a positive self-corrective direction, just like thinking about thinking does not, it is a privilege ground for instances of regulatory mechanisms because just

as thinking about thinking can guide and correct thinking, emotions about emotions can refine and correct feeling. In order to show this, we will briefly describe emotion regulation and the ways in which these same mechanisms can be understood in more detail. It is also important to note that emotion regulation is more apparent in researchers who adopt a basic emotion perspective or an appraisal perspective as opposed to theorists who see emotions more as a social and psychological constructionist perspective (Gross and Barrett 2011, p. 8). Even though Gross and Barrett (2011) do not think the difference is a necessary state of affairs, it shows the importance of becoming aware of how the different theoretical perspectives condition and format certain aspects of emotion research because becoming aware of the theoretical bends and tendencies enables theorists to be less at the mercy of the consequences of their theoretical assumptions and more able to incorporate areas of emotion research that appear within other theoretical perspectives.

Emotion regulation originally appeared in developmental psychology (Gaensbauer 1982) and many researchers consider emotion regulation central to psychopathology (Gross 1998, p. 274). Emotion regulation is distinct from emotion generation because the first refers to ways in which people influence "which emotions they have, when to have them and to experience them and how to express them" (Gross 1998), while emotion generation concerns what triggers emotions to appear and, according to some perspectives the distinction exists "on the assumption that the two are biologically distinct" (Gross and Barrett 2011, p. 4). While anger will be generated by an unfair situation, the regulation refers to all the types of processes a person can experience to deal with the anger already in place in face of an unfair situation such as controlling expression of anger, revaluating the situation to verify its unfairness, avoiding future instances of confrontation with aspects of the unfair situation. However, "the distinction between these two processes still remains the source of debate" (Aldao et al. 2010, p. 218), even though the process of emotion regulation has been conceptualized as a distinct process from emotion generation (Gross and Thompson 2007; Rottenberg and Gross 2003).

There are a variety of ways in which emotion regulatory processes can be said to occur because emotion regulation can be either automatic

or controlled, it can either appear at the conscious or unconscious level, and the regulatory process may happen in one go and all at one or appear in several points in the emotion generative process and its development. Thus, emotion regulation involves changes in various ways (latency, rise time, magnitude, duration, and offset of responses in behavioral, experiential, or physiological domains) (Gross 1998, p. 275). It is distinguished from other psychological processes such as coping because coping refers to "cognitive and behavioral efforts to deal with specific demands" (Lazarus and Folkman 1984, p. 141), as well as "mood regulation, defense, and affect regulation" (Gross 1998, p. 271). Consequently, this means that emotion regulation can include various actions (including nonemotional) (Gross 1998, p. 275).

Gross describes how emotions can be regulated at five different moments of the emotion process: "(a) selection of the situation, (b) modification of the situation, (c) deployment of attention, (d) change of cognitions, and (e) modulation of responses" (Gross 1998, p. 271). This means that the therapeutic interventions to help individuals, couples, and families also aim to modify ineffective patterns of emotion regulation (Gross 1998, p. 280), and reinforces how psychotherapy targets ways to influence situations and ways to modify the way the situations are construed, in addition to searching for ways to alter the emotional response to the situation itself (Gross 1998, p. 280). Accordingly, the five ways of emotion regulatory processes described by Gross are constantly challenged in order to disentangle emotion generative from emotion regulatory processes (Gross 1998, p. 286). In addition, the multidimensional of emotions' various components (experiential, behavioral, physiological) and their multilayered complexity makes emotion regulatory processes hard to capture, identify and master. Since emotions be seen to serve numerous functions (evolutionary, social, communicative, decision-making), it is clear that emotion regulation is a social and cultural process, which always needs to be taken into consideration in reflection and research upon emotion regulation (Gross 1998, p. 279).

Finally, emotion can be adaptive and maladaptive, and consequently emotion regulation can also sometimes be in line with an adaptive pattern of responses just like meta-emotions can sometimes increase health, and at other times need to be regulated in order to maintain health.

In sum, though psychologists and philosophers usually illustrate meta-emotions as strategies for healthy emotion regulation, there is no trait in reflexivity that grants healthy regulation, and consequently, the relationship between emotions and meta-emotions can be far more negative that the usual positive regulative connections that often appear in the theoretical examples (Howard 2017, pp. 11–15). Nevertheless, we aim to show that in its connection to reasoning we can see that it is a privilege ground for instances of regulatory mechanisms because the reflexivity of emotions is what enables emotions to become reasons for action and choices, which then enables them to be part of a critical sphere in which they can be further evaluated as good, less good, or bad reasons within deliberation and decision-making.

In addition to the previous problem, Howard describes how "simple and paradigmatic cases of meta-emotion forces us to modify an intuitive thesis about the cognitive basis of any emotions' object-directedness (Howard 2017, p. 2). Just like emotion regulation asks us to become aware of our assumptions about emotions (Gross and Barrett 2011, p. 8), so meta-emotion requires a revision of how we think about emotions' intentionality. According to Howard, the natural interpretation about intentionality of emotions, which he calls the tripartide model, views that the relation of emotion to "its object is mediated by another mental state such as a belief, perception, memory, or imagining, that represents the object in question" (Howard 2017, p. 3). This provides an explanation of the intentionality of emotions and the cognitive bases of emotions because it justifies how we are aware of the object of an emotion by way of other states (Deonna and Teroni 2012, p. 85). As Howard well describes, the tripartide model is not problematic and can smoothly enable us to spell out how emotions work and feel almost beyond question because it explains how objects of emotion are internal to the subject (Howard 2017, p. 4). In addition, the model enables treating the cases in which the object of an emotion is another mental state in a similar fashion for "when an emotion is about another mental state as such, that state is represented as content by a third mental state, which is the emotion's cognitive base" (Howard 2017, p. 4), making it similar to how it happens with mind-external objects of emotion. So, when meta-emotions are primarily a response to an autobiographical

belief or aspects of a narrative notion of self, or as a sense of image that others have about the self, the tripartide model works without any problem. However, Howard argues, the tripartide story becomes less able to deal with certain simple and direct cases of meta-emotions, and they "turn out to undermine the tripartide model as a thesis about the necessary structure of emotional intentionality" (Howard 2017, p. 6), because meta-emotions do not seem to require a third mediating element. Howard illustrates this with a case of hating his own jealous thoughts,

> when we recognize that the typical target of a metaemotion is an emotional *experience* – e.g. that what I loathe are the jealous thoughts and feelings I am experiencing – it becomes clear that the two-part analysis is sufficient to account for metaemotional intentionality. For it is in the nature of *having* jealous thoughts and feelings that one is ordinarily aware of them. And if my jealousy, *qua* experience, is already characterized by such self-awareness, then it does not stand in need of further representation if it is to be target of another mental attitude, such as metaemotion. (Howard 2017, p. 7)

Howard concludes that the way in which we tend to think about emotions' intentionality as "mediated by a separate state that represents the object such as perception or a belief" (Howard 2017, p. 2) cannot be applied and transferred to many meta-emotions, and that we can also see the shortcomings of the tripartide model with meta-emotions when we can compare it with other instances of emotions about other states (such as a sensation) in which there is also no need for a mediating third mental state. Accordingly, one can have an emotion about pain and "this pain, a feeling, is already something that I am aware of, and thus it does not need to be object of a further representation in order to be the object of an emotion" (Howard 2017, p. 7).

Even though, as Howard points out, Robert Gordon "has warned that 'reference to the relation between an emotion and its object' quickly proves to be a nonstarter" (Gordon 1987, p. 45), intentionality is at the center of the affective turn. It is the recognition that emotions have intentionality that made Kenny distinguish emotions from

perceptions and sensations showing that "the dissimilarities between emotions and perceptions are more significant than the similarities" (Kenny [1963] 2003, p. 39), and that "[t]he most important difference between a sensation and an emotion is that emotions, unlike sensations, are essentially directed to objects" (Kenny [1963] 2003, p. 41). The fact that emotions have intentionality sets them apart from sensations, and introduces the cognitive element in their nature giving the cognitive theories of emotion a new strength and offering a new difficulty for feeling theories of emotions. It is beyond the scope of this paper to examine emotion intentionality and how the cognitive element of emotions is handled by the different theories of emotion, though the present analysis sufficiently points out why meta-emotions stand as an ultimate test to the available emotion theories as well as a test for the more general assumptions of theories of emotions.

In sum, though the tripartide model works in general, it does not seem to be able to deal with some cases of the emotional life such as some meta-emotions and emotions about other states like sensations because these instances of emotion do not require a third mediating element. In the next section, we want to suggest that adopting the already available Situated Approach to Emotions is well-suited for dealing with the issues raised above because it does not necessarily imply the tripartide model of emotion, and it enables a way to explain the effect of refinement and complexity on emotion regulation.

4   Meta-Emotion in a Situated Approach to Emotions

The Situated Approach to Emotions has been on the rise in the last years (Griffiths and Scarantino 2008; Stephen 2012; Mendonça 2012; Stephen et al. 2014) and offers several advantages. First, it is in line with the situated perspective of the mind and its advantages. For emotion theory, it offers an important specific advantage: given that the various problems of emotion theory leave the field with no other choice but to accept that, for the time being, a theoretical pluralism

(Griffiths and Scarantino 2008, p. 449). The Situated Approach to Emotions can be coordinated with various other theoretical backgrounds because "the situationistic perspective is not, in principle, incompatible with other existing theoretical approaches" (Griffiths and Scarantino 2008, p. 449). Consequently, the Situated Approach to Mind provides a privilege terrain for emotion theory because, while maintaining theoretical pluralism, it may also offer the needed thinking tools to deal with the role of emotion in rationality, thinking, and decision-making.

The pragmatists are among the various influences of the Situated Cognition and the Situated Approach to the Mind (Clancey 2009; Gallagher 2009, p. 35). Our specific take on the Situated Approach to Emotions is developed upon only one of the pragmatists: John Dewey's philosophical work. We extend to Emotion Theory Dewey's insightful concept of situation and building upon it (Mendonça 2012) take up the way in which Dewey explains that, "an emotion is to or from or about something objective. An emotion is implicated in a situation" (Dewey, LW 10:72).

As Gallagher writes, Dewey is not just part of an historical background but by thinking about his contributions, we can go further. Namely, when we read Dewey "[w]e should add the important point that the situation should be defined as inclusive of the inquirer. It is not I as a cognitive inquirer confronting a situation; the situation surrounds and includes me" (Gallagher 2009, p. 47). Thus, recovering Dewey's description of what is a situation enables us to move away from the colloquial use that takes situation to be independent of the subject of experience, and able to adopt a philosophical technical meaning that incorporates the subject of experience into the notion of situation.

One frequent objection to Dewey's proposal is that his description of a situation leaves unanswered, the question "who is the subject of experience?" The objection often focuses on the notion of problematic situation and questions how a situation can be problematic without a subject to have the experience and identify its problematic character (Mackay 1942, p. 394). If the problematic situation does not require the subject to disclose it, then we are forced to wonder why Dewey chooses such terminology; for the word "problematic" implies someone to find

it such. If the problematic character of the situation is dependent on a subject, then it will vary according to the different subjects who experience it. This will make what is problematic to be mainly subjective and greatly depend on the abilities of the subject to identify problems. Dewey thinks the objection confuses the importance of the subject in experience with a bias toward objectivity. He writes,

> Experience shows that as a matter of fact objective reference precedes subjective reference. Reference to a subject instead of an object is extrinsic and reflective. It is indeed only another mode of objective reference; that is, some tediousness of the object is accounted for in terms of an unusual state of the object. Otherwise to Say "I am bored" and "It is tedious" are merely two phases to express exactly the same fact. (Dewey LW2:91)

Thus, when a situation is taken to be external to the subject's feelings while it simultaneously incorporates the subject, it follows that when a situation is of a certain kind it must affect the subject who experiences it accordingly. For example, if the situation is dangerous, the subject feels fear. And this explains why we think something is wrong, for instance, when a scary situation does not frighten the subject. In addition, Dewey thinks that this confusion between the importance of the subject and object seems to imply that ownership is a characteristic that makes rational discourse unattainable because it suggests that something being mine, yours, or of a specific subject "so permeates the properties of being a house that nothing intelligible can be said about the latter" (Dewey LW1:179).

When we take up John Dewey technical meaning of situation and understand that no such situation can be thought without a subject to experience because to understand properly, Dewey's notion of situation means that "[w]e should add the important point that the situation should be defined as inclusive of the inquirer. It is not I as a cognitive inquirer confronting a situation; the situation surrounds and includes me" (Gallagher 2009, p. 37), we acquire other theoretical gains. For example, it establishes a pertinent link between emotions and moods because "the hypothesis that emotions are best understood by the notion of situation, while moods by the notion of context, would

provide an interpretative tool to equate their interconnection by under-standing the links between contexts and situations" (Mendonça 2017, p. 8). In what concerns meta-emotions, it enables a better way to under-stand how meta-emotion contributes to emotion regulation because it shows that an accurate description of the situation is needed to fully understand the role of emotions and meta-emotions in a situation. In addition, the way the subject perceives and is able to describe the situ-ation will modify how emotion regulation is available for the subject of experience.

That is, when the Deweyan technical notion of situation is adopted, it is then possible to show that to fully understand emotions and meta-emotions requires an accurate description of the situation in which they occur or that they qualify. For example, if someone acts out of fear of water when one is three years old may not require more than a description of fear of the ocean to pinpoint the situation. We easily take fear as something that happens to that person, but if someone acts out of fear when one is much older, say thirty or forty or fifty, then it may raise curiosity and require that a more complex story be told because we think the description of the situation is incomplete and we look for a reason for fear of water to have such an impact at an older age. The same emotion (fear of water) requires a more detailed description of the situation when the subject who experiences is different because the sub-ject stands is part of the situation.

To conclude this section we can say that adopting a Deweyan con-ception of situation means that we add the subject of experience into its description and can contrast it with a more general context in which situations occur, similarly to the way emotions occur within a more general mood. This different picture of emotional entities (emotions) within the scope of more general emotional entities (moods) offers a new way to understand the way meta-emotions belong to the general emotional landscape. Thus now, when we consider meta-emotions, it is now crucial to understand how they change the situation within a more general contextual whole. The Situated Approach to Emotions proposal thus changes the focus on the tripartide model in what type of third mediating element is necessary because it focuses on the more general dynamics of the emotional landscape in which different

dimensions dynamically interact, and meta-emotions can now be described as another extra type of dynamical emotional movement. In this way the Situated Approach to Emotions dodges Howard tripartide model criticism because it puts forward a more complex picture in which a situation is not simply described by "belief, perception, emotion" and, since it includes the subject of experience, it requires a more detailed narrative that captures first order emotions and meta-emotions acknowledging that it may be the case that more than one emotion is at play in a situation, and consequently that issues such as goal of action, intention, moods, layers of emotion play a role in the experience of a situation (Mendonça 2012). The conclusions of this section suggest that in light of the Deweyan concept of situation, it is now possible to provide a richer description of situations, the role of emotions and meta-emotions and their role in decision-making and rationality.

5 Describing and Interpreting Situations: Emotions as Causes and Reasons

Describing a situation within this Deweyan framework means that a proper understanding of the role of emotion in it may require telling a story. This is perhaps why we find the frequent statement from emotion theorists that emotion has a narrative structure (De Sousa 1987, p. 183; Baier 1990, p. 18; Goldie 1999, p. 395; Nussbaum 2001, p. 236). Martha Nussbaum explains that certain aspects of emotions can only be understood in light of the narrative underlying their experience writing that "[t]his is what Proust meant when he claimed that certain truths about the human emotions can be best conveyed, in verbal and textual form only by a narrative work of art" (Nussbaum 2001, p. 236).

We want to add that the narrative sometimes required to properly describe an emotional situation means being able to indicate the way in which meta-emotions color and influence first-order emotions, and provide the interpretative reading of when the different emotional levels require different responses. For example, someone who has been taught to not get angry may feel shame and hide anger because this person

will have difficulty in acting in a strategic and appropriate way that encompasses both emotional states (Jones and Botker 2001, p. 240). This means that while describing a situation, one needs to apprehend the details that give away which story is behind the emotional experience. When someone is four years old and scared of the ocean the story might not need more details than the age of the subject of experience, while when someone is embarrassed of being scared, we might need more specific details to understand what type of situation is giving the experience of fear and of embarrassment besides the age of the subject.

If our suggestion is correct it means that describing the situation is crucial for working out how emotions and meta-emotions work in reasoning because to the extent that we get a description, we attain the various layers of emotions and moods that guide and format the description, and it is then easier to better understand how interpretation of situation may modify the way a situation is experienced providing ways to establish the regulatory mechanisms identified by Gross. For instance, Jones and Botcker argue that becoming aware of our meta-emotions may be a crucial way to change perspective and adopt a more collaborative way to deal with a conflict situation (Jones and Botker 2001, p. 240).

We think that the interpretative task is fundamental for understanding the role of emotion in reasoning and deliberation because it provides a better understanding the complexity of emotional experience showing the plural role of emotions in deliberation (Mendonça and Sàágua 2018) and the source of its normative strength. For example, it shows how certain meta-emotions can become first-order emotions as, for example, when someone predicts the possibility of fear, they might feel embarrassed before its occurrence making embarrassment a first-order emotion because of its second-order genealogy. In addition, it reveals that emotions have a regulatory character because they are sometimes best captured as reasons for action and other times best captured by the description of causes of action. For example, when someone is at the mercy of fear, the emotion is best interpreted in a determinate causal form. However, when someone is embarrassed about their fear, we may no longer describe the fear in a causal manner, and the added layer of

emotion—embarrassment—that occurs may be best captured by a reason for action in a determinate way. That is, while a description of the situation might focus on emotion of fear—he was scared but nevertheless controlled it to attain the desired goal—the interpretative explanation of the situation would indicate how embarrassment stood as a reason to control fear—he was scared but because of his embarrassment, he was able to control his fear and attain the desired goal. Thus, emotional reflexivity grants an added ability of control and regulate emotion, and being able to identify meta-emotions and their regulatory role also enables to better grasp how emotions work as reasons.

The suggestion that sometimes emotions are causes of action and other times are reasons for action is not meant to introduce a qualitative difference in emotions. We follow Davidson in thinking that rationalization is a "species of causal explanation" (Davidson 1963, p. 691), and hope to further reinforce his claim with the reflexivity of emotion and its regulatory role since the reflection of the paper shows that when someone provides an interpretation of their action, they have given us a new description of what happened by introducing an added element when they described the situation with different emotional layers. To use the previous example of the fear of water, when we asked someone why they looked so tense near the pool and they say that they are embarrassed about their fear of water "we have an interpretation, a new description of what he did which fits it into a familiar picture" (Davidson 1963, p. 691). Introducing an emotion in a process of reasoning may be what is required to shift the description to a re-description of the situation and enable other actions.

6   Concluding Remarks

It is clear that the recognition and inclusion of meta-emotions is decisive for an updated picture of rationality that incorporates the emotional input in decision-making. Nevertheless, far more work is required besides including emotional reflexivity in order to attain a new picture of

rationality. As Davidson put it, "[t]alk of patterns and contexts does not answer the question of how reasons explain actions, since the relevant pattern or context contains both reason and action" (Davidson 1963, p. 692). Therefore, though emotional reflexivity shows how emotions stand as reasons and provide a new description of situations, the suggestion needs to be further explored and analyzed. For example, it will be crucial for the completion of this analysis to examine the connection between normative and descriptive. If meta-emotions can be shown to have an impact in reasoning because they add items to the situations, as Jones and Botker suggest (2001), and if descriptions are already loaded with meta-emotional layers of experience, then it is important to verify how the new descriptions modify the overall situation. This also highlights how it is central to explore the link between language and emotions because if statements re-describe a situation, this may explain why language provides an added form of modifying situations. In addition, the way language expresses emotions may carry the complexity of emotional reflexivity and show why sometimes it is a key for regulatory mechanisms. In sum, whatever the outcomes of future research might be, reflexivity promises to be insightful.

References

Aldao, A., S. Nolen-Hoeksema, and S. Schweizer. 2010. Emotion Regulation Strategies Across Psychopathology: A Meta-Analytic Review. *Clinical Psychology Review* 30: 217–237.

Baier, A. 1990. What Emotions Are About. *Philosophical Perspectives* 4: 1–29.

Bartsch, Anne, Peter Vorderer, Roland Mangold, and Reinhold Viehoff. 2008. Appraisal of Emotions in Media Use: Toward a Process Model of Meta-Emotion and Emotion Regulation. *Media Psychology* 11 (1): 7–27.

Bartsch, A., M. Appel, and D. Storch. 2010. Predicting Emotions and Meta-Emotions at the Movies: The Role of the Need for Affect in Audiences' Experience of Horror and Drama. *Communication Research* 37 (2): 167–190.

Beer, N., and G. Moneta. 2010. Construct and Concurrent Validity of the Positive Metacognitions and Positive Meta-Emotions Questionnaire. *Personality and Individual Differences* 49: 977–982.

Belli, S., and F. Broncano. 2017. Trust as a Meta-Emotion. *Metaphilosophy* 48 (4): 430–448.

Clancey, W.J. 2009. Scientific Antecedents of Situated Cognition. In *The Cambridge Handbook of Situated Cognition*, ed. P. Robbins and M. Aydede, 11–34. Cambridge, UK: Cambridge University Press.

Couyoumdjian, Alessandro, et al. 2016. Reducing the Meta-Emotional Problem Decreases Physiological Fear Response During Exposure in Phobics. *Frontiers in Psychology* 7: 1105.

Damasio, A. 2003. *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. London: William Heinemann.

Davidson, D. 1963. Actions, Reasons, and Causes. *The Journal of Philosophy* 60 (23): 685–700.

Deonna, J.A., and F. Teroni. 2012. *The Emotions*. New York: Routledge.

De Sousa, R. 1987. *The Rationality of Emotion*. Cambridge, MA: MIT Press.

Dewey, J. 1988–1989. *John Dewey: The Later Works: 1925–1953*, ed. J. A. Boydston. Carbondale: Southern Illinois University Press.

Elster, J. 1996. Rationality and the Emotions. *Economics Journal* 106: 1386–1397.

———. 1999a. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.

———. 1999b. *Strong Feelings: Emotion, Addiction, and Human Behavior*. Cambridge, MA: MIT Press.

Feagin, S.L. 1995. The Pleasures of Tragedy. In *Arguing About Art*, ed. A. Neill and A. Ridley, 204–217. New York, NY: McGraw-Hill.

Ferrari, M., and E. Koyama. 2002. Meta-Emotions About Anger and Amae: A Cross-Cultural Comparison. *Consciousness & Emotion* 3 (2): 197–211.

Gaensbauer, T.J. 1982. Regulation of Emotional Expression in Infants from Two Contrasting Caretaking Environments. *Journal of the American Academy of Child Psychiatry* 21 (2): 163–170.

Gallagher, S. 2009. Philosophical Antecedents of Situated Cognition. In *The Cambridge Handbook of Situated Cognition*, ed. P. Robbins and M. Aydede, 35–51. Cambridge: Cambridge University Press.

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

Greenberg, L.S. 2011. *Emotion-Focused Therapy: Theories of Psychotherapy Series*. Washington, DC: American Psychological Association.

Greenspan, P. 2004. Practical Reasoning and Emotion. In *The Oxford Handbook of Rationality*, ed. A. Mele and P. Rawlings. New York: Oxford University Press.

Griffiths, P., and A. Scarantino. 2008. Emotions in the Wild: The Situated Perspective on Emotion. In *The Cambridge Handbook of Situated Cognition*, ed. P. Robbins and M. Aydede, 437–453. Cambridge: Cambridge University Press.

Goldie, P. 1999. How We Think of Other's Emotions. *Mind and Language* 14: 394–423.

Gottman, J.M., L.F. Katz, and C. Hooven. 1996. Parental Meta-Emotion Philosophy and the Emotional Life of Families: Theoretical Models and Preliminary Data. *Journal of Family Psychology* 10 (3): 243–268.

———. 1997. *Meta-Emotion: How Families Communicate Emotionally*. Mahwah: Lawrence Erlbaum Associates.

Gross, J.J. 1998. The Emerging Field of Emotion Regulation: An Integrative Review. *Review of General Psychology* 2 (3): 271–299.

Gross, J.J., and Lisa Feldman Barrett. 2011. Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View. *Emotion Review* 3 (1): 8–16.

Gross, J.J., and R.A. Thompson. 2007. Emotion Regulation: Conceptual Foundations. In *Handbook of Emotion Regulation*, ed. J.J. Gross, 3–24. New York, NY: Guilford Press.

Helm, B.W. 2000. Emotional Reason: How to Deliberate About Value. *American Philosophical Quarterly* 37: 1–22.

Jäger, C., and A. Bartsch. 2006. Meta-Emotions. *Grazer Philosophische Studien* 73: 179–204.

Jäger, C., and E. Bänninger-Huber. 2015. Looking into Meta-Emotions. *Synthese* 192: 787–811.

Jones, T., and A. Bodtker. 2001. In Theory Mediating with Heart in Mind: Addressing Emotion in Mediation Practice. *Negotiating Journal* 3: 217–244.

Haradhvala, Natasha. 2016. Meta-Emotions in Daily Life: Associations with Emotional Awareness and Depression. Arts & Sciences Electronic Theses and Dissertations. 97. http://openscholarship.wustl.edu/art_sci_etds/978. Accessed 1 December 2018.

Hoffmann, S.G. 2013. The Pursuit of Happiness and Its Relationship to the Meta-Experience of Emotions and Culture. *Australian Psychologist* 48 (2): 94–97.

Howard, S.A. 2017. Metaemotional Intentionality. *Pacific Philosophical Quarterly* 98: 406–428.

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

Kirman, A., P. Livet, and M. Teschl. 2010. Rationality and Emotions. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences* 365 (1538): 215–219.

Koven, N.S. 2011. Specificity of Metaemotion Effects on Moral Decision Making. *Emotion* 11 (5): 1255–1261.

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

Leahy, R.L. 2002. A Model of Emotional Schemas. *Cognitive and Behavioral Practice* 9: 177–190.

Mackay, D.S. 1942. What Does Mr. Dewey Mean by an Bindeterminate Situation? Volume 15: 1942–1948 Essays, Reviews and Miscellany. In *John Dewey: The Later Works—1925–1953*, ed. J.A. Boydston, 393–401. Carbondale: Southern Illinois University Press. 1991.

Mata, R., T. Pachur, B. von Helversen, R. Hertwig, J. Rieskamp, and L. Schooler. 2012. Ecological Rationality: A Framework for Understanding and Aiding the Aging Decision Maker. *Frontiers in Neuroscience* 6: 19.

Mendonça, D. 2012. Pattern of Sentiment-Following a Deweyan Suggestion. *Transactions of Charles Peirce Society* 48 (2) (Spring): 209–227.

———. 2013. Emotions About Emotions. *Emotion Review* 5 (4) (October): 1–7.

———. 2017. Situating Moods. *Philosophia* 45: 1453–1467.

Mendonça, D., and J. Sàágua. 2018. The Plural Role of Emotions in Ethics. In *Essays on Values and Practical Rationality: Ethical and Aesthetical Dimensions (IFILNOVA Yearbook)*, ed. António Marques and João Sàágua, 53–70. Bern: Peter Lang.

Mitmansgruber, H., T.N. Beck, S. Höfer, and G. Schüßler. 2009. When You Don't Like What You Feel: Experiential Avoidance, Mindfulness, and Meta-Emotion in Emotion Regulation. *Personality and Individual Differences* 46: 448–453.

Norman, E., and B. Furnes. 2016. The Concept of "Metaemotion": What There Is to Learn from Metacognition? *Emotion Review* 8 (2) (April): 187–193.

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

Pham, M.T. 2007. Emotion and Rationality: A Critical Review and Interpretation of Empirical Evidence. *Review of General Psychology* 11 (2): 155–178.

Pugmire, D. 2005. *Sound Sentiments: Integrity in the Emotions*. Oxford: Oxford University Press.

Rottenberg, J., and J. Gross. 2003. When Emotion Goes Wrong: Realizing the Promise of Affective Science. *Clinical Psychology: Science and Practice* 10: 227–232.

Stephan, A. 2012. Emotions, Existential Feelings, and Their Regulation. *Emotion Review* 4 (2): 157–162.

Stephan, A., S. Walter, and W. Wilutzky. 2014. Emotions Beyond Brain and Body. *Philosophical Psychology* 27 (1): 65–81.

Williams, B. 1981. Moral Luck. In *Moral Luck*, 20–40. Cambridge: Cambridge University Press.

Part IV

The Epistemic Value of Emotions in Self-Understanding

7

Moving Stories: Agency, Emotion and Practical Rationality

Dave Ward

1 Introduction

How are agents made? Which psychological components, assembled with which structure, do we need to engineer a system that acts for reasons, rather than one that is merely shunted around by external forces? One kind of answer to these questions, associated with G. E. M. Anscombe (1957), identifies the crucial ingredient as epistemic—the qualitative difference between agents and mere movers or behavers depends on a distinctive kind of knowledge that agents have of their own actions. As an agent, your knowledge of mere happenings (like your stomach involuntarily rumbling) is observational, whereas your knowledge of your voluntary actions (like your going to the kitchen to make a snack) is non-observational—you have practical knowledge of your own activity through your very act of carrying it out. Let's assume this framework for now (we'll work through arguments in its favour below).

D. Ward (✉)
University of Edinburgh, Edinburgh, UK
e-mail: Dave.Ward@ed.ac.uk

How should the kind of practical knowledge distinctive of agency be analysed? In particular, what role if any do *emotions* play in enabling knowledge of this kind?

Over the past 30 years, J. David Velleman (1989, 2006a, 2009, 2014) has developed a sophisticated view of practical rationality which aims to address these questions. For reasons we'll set out below, he argues that the kind of practical rationality distinctive of agency depends on the motivational force of a drive towards self-consistency. To be an agent is to be a system that is driven to act in ways that make sense to oneself. Articulating the nature and operations of this drive towards self-consistency thus becomes the key task for a theory of human agency and practical reason. The main question with which this chapter is concerned is: What role do emotions play in structuring such a drive and its operations? As we will see, there are two interpretations of the nature of the drive towards self-consistency in Velleman's work, not always happily related. Agents might aim towards self-consistency by striving to act in ways that make causal-psychological sense—roughly, by acting in ways that maximize the coherence of their beliefs, desires, plans and other psychological states. Or they might aim towards self-consistency by striving to act in ways that make *narrative* sense—acting in ways that make their lives conform to the beats of familiar story structures. Velleman provides compelling reasons to think that emotions are implicated in narrative sense-making—the beats of familiar story structures are *felt* rather than thought through, so agency can consist in shaping your life in order to feel a particular way about it. How are causal-psychological consistency and narrative consistency related? As we'll see below, Velleman initially thought of the latter as an appealing way of redescribing the former, but later came to argue that each kind of consistency conveyed a fundamentally different kind of understanding. I will agree with Velleman that causal-psychological and narrative understanding are distinct, but argue that Velleman's view of their relations is unsatisfactory. Velleman comes to prioritize causal-psychological over narrative self-understanding in a way that jeopardizes the naturalistic credentials of his framework and undermines his claim that a drive towards narrative coherence can constitute a legitimate source of practical rationality. The positive suggestion this chapter aims to motivate

is that the problems facing Velleman's view can be overcome by reversing his order of dependence between causal-psychological and narrative self-consistency—instead of viewing narrative self-understanding as a sophisticated achievement resting on folk-psychological self-understanding, we should understand our drive towards rational self-consistency—and thus our status as agents—as resting on an emotionally structured bedrock of narrative competences.

2   Narrativity and Normativity

In 'The Self as Narrator', Velleman (2005) proposes a connection between narrativity and a kind of normativity which he thinks is distinctive of practical rationality. A self-narrating system—one that issues a running commentary on its own activities—comes to exhibit a form of practical rationality when its past self-narrations constrain its present action selection in a particular way. This is the sort of constraint that happens when, weighing up whether to go for a run or to the pub, I remember saying to myself that morning that I was turning over a new leaf—one involving less pub-going and more running—and factor that into my decision. A bit of self-narration from my past has come back to haunt my present deliberation.

Velleman's guiding example, adapted from Dennett's (1991) reflections on the links between selfhood and narrativity, concerns Gilbert, an intelligent robot that churns out a running commentary on his activities as he goes about his business. Dennett (op. cit., 1991; Dennett and Humphrey 1998) thinks that such epiphenomenal, stream-of-consciousness self-narration is all there is to selfhood. Consequently, Dennett holds that we should understand the self—both in Gilbert's case and our own—as a *narrative fiction*. The self is just the protagonist described or implied by the epiphenomenal tale spun by a self-narrating system. But this protagonist is fictional. When we look beyond or behind the narrative, Dennett argues, we will find nothing but a patchwork of special purpose subsystems, cleverly stitched together to issue in behaviour and self-report that gives the illusory appearance of a unified 'self' working behind the scenes.

In response, Velleman points out that self-narration need not be epiphenomenal. There are cases, including our own, where 'an autobiography and the behaviour that it narrates are mutually determining' (Velleman 2005, p. 211), such that a narrative produced at one time can have repercussions for subsequent deliberation and action selection. The narrative Gilbert produced about his activities in the morning might shape his action selection in the evening. If Gilbert self-narrates that he will soon go into the supply closet for fresh batteries, the memory of this proclamation might tip the balance in a subsequent competition between closet-going and energy-conserving subroutines vying for control of Gilbert's behaviour. In the same way, my morning proclamation that I am now more of a runner than a pub-goer can help determine my behaviour in the evening. According to Velleman, this kind of system or creature—one that is driven to shape its current action selection to cohere with its past acts of narration—is a genuine *agent*. And this specifies a sense in which such a system or creature has, or is, a *self*—as we will see below, Velleman argues that some of their behaviours qualify as exercises of agency by being genuinely *autonomous* or *self-governed*.

One intuitive way of differentiating actions from mere happenings is via appeal to the distinctive way in which the past shapes the present and the future in exercises of agency. Every happening in the universe unfolds according to causal laws, whereby the past and present shape the future. But some happenings are also *acts*, in which an agent can properly be held responsible for the way things unfold. In cases of agency the past and present shape the future not only via causal laws, but also by furnishing an agent with *reasons* to act in one way rather than another.[1] Velleman's suggestion is that we can often understand these reasons in terms of the way in which potential states of an agent's future stand in relations of *narrative coherence* to states of their past. When weighing up whether to run or to pub, my deliberation is shaped

---

[1]On Velleman's understanding this way of drawing the act/happening distinction coincides with Anscombe's epistemic strategy as outlined in the introduction. As we will see, for Velleman the reason-governed way in which the past shapes the future in agency is a function of the distinctive epistemic relation in which an agent stands to her acts.

not only by causal processes determining my current levels of thirst and fatigue, but also by the fact that I *told myself* this morning that I was now the sort of person who goes for a run more often than they go to the pub. Choosing the pub over the run would mess up the coherence of this self-conception. According to Velleman, the fact that these considerations of coherence are in play in motivating my choices can mark me out as an *agent* who is responsive to *reasons*, rather than a mere lump of organic matter, shunted around by external forces.

What exactly, you might be wondering, is narrative coherence? More on this soon. For now, note that feeling the motivational sway of considerations of narrative coherence is, for Velleman, sufficient but perhaps not necessary to qualify as an agent. This is because his earlier work argues, persuasively and influentially, for understanding agency in terms of susceptibility to the motivational sway of considerations of *causal-psychological* coherence—roughly, being motivated to act in a *rationally* coherent way. We will look at this view in a little more detail below. For now, the basic idea is that someone motivated by a drive towards causal/psychological coherence is thereby disposed to avoid acting in ways that are inconsistent with their plans, projects and self-conception, and to ensure that the motives that drive their eventual action are those that they, as a rational agent, have the strongest reasons to endorse.[2] It thus appears, on Velleman's picture, that there are two ways in which a system can qualify as autonomous, and thereby as an agent. Agents can be motivated by a drive towards narrative coherence, as when they try to act so as to make a coherent story out of the unfolding events of their lives. Or they can be motivated by a drive towards causal-psychological coherence, as when they try to maximize the rational consistency of their current activity with their plans, projects and self-conceptions. How are these two ways related to each other? On the face of it, Gilbert's declaration that he's about to go into the closet, and my declaration that I will err towards running over pubbing, could be understood as constraining our respective futures via a drive towards either kind of coherence. Indeed, Velleman initially supposed

---

[2]See e.g. Velleman (2006b, 2009, 2014).

that narrative coherence was simply an appealing way of redescribing causal-psychological coherence:

> My hope was that narrative self-understanding would be less objectionable to those who had previously criticized my view of autonomy as overly intellectualist, because it requires the agent to psychologize about himself. Nobody wants to think that he is studying himself like a rat in a maze; but everybody likes to think of himself as the novelist of his life and hence as the protagonist of his own story. (2007, p. 284)

We all know, however, that great stories can contain gaps and inconsistencies. And any student of philosophy knows that great reasoning doesn't always make for great reading. Velleman thus rightly came to think that narrative coherence is distinct from causal-psychological coherence, and that the pursuit of each kind of coherence involves striving to confer a distinct kind of intelligibility on ourselves and our activities—more on which below. This complicates Velleman's view of the kind of practical rationality that is constitutive of agency:

> The upshot is that practical reasoning is fragmented into the pursuit of two distinct modes of self-understanding. I think that we aim to make sense of ourselves not only in the mode of causal explanation but also in the mode of storytelling. (ibid., p. 285)

How do these two modes of self-understanding relate to each other? And what, a reader of this volume might be wondering, does either of them have to do with emotions? While our ultimate goal is to make progress on the first question, the second question is easier—so let's start there.

3 Understanding via Emotional Cadence

As noted above, Velleman initially thought of narrative self-understanding simply as a folksy way of describing the causal-psychological self-understanding he thought was constitutively involved in practical

rationality. As we also noted, however, the virtue of sheer rational coherence looks different from the virtue of being a good story. Velleman (2003) argues that this is because stories convey understanding in way that is different from a mere description of causally connected events, and do so because they arouse and complete an *emotional cadence*. A description of a sequence of causally connected events becomes a story when it has a beginning, middle and end, with each of these stages understood in terms of their emotional structure. A story's beginning arouses an emotional reaction, which is developed or complicated by its middle, and resolved by its end. It thus conveys an understanding of its events by bringing its audience to *feel* a particular way about them:

> the emotion that resolves a narrative cadence tends to subsume the emotions that preceded it: the triumph felt at a happy ending is the triumph of ambitions realized and anxieties allayed; the grief felt at a tragic ending is the grief of hopes dashed or loves denied. Hence the conclusory emotion in a narrative cadence embodies not just how the audience feels about the ending; it embodies how the audience feels, at the ending, about the whole story. (ibid., p. 19)

A happy ending can help the audience to view the protagonist's trials and tribulations as character-building episodes on the road to her inevitable triumph; a tragic ending can help the audience to view the same struggles as hopeless attempts to delay a foregone conclusion. Emotions thus impart narrative understanding by shaping the audience's experience of events as they unfold, and by imposing a stable and coherent interpretation of events via the anticipation or undergoing of a concluding emotion that provides a unified perspective on the story's contents.

This view of narrative understanding relies on a pair of relatively uncontroversial claims about emotion that I will assume to be correct. The first is that emotions have an intentional component—they involve a way of being directed towards specific objects, or towards the world at large (see e.g. Lazarus 1991; Goldie 2000; Prinz 2004; Slaby 2008). This is one reason why the emotional character of stories imparts understanding—feeling emotions involves having our minds preferentially directed to particular aspects of situations and events. The second claim

is that at least some emotions (those implicated in narrative understanding) have an essentially diachronic and relational character (see e.g. Lewis 2005; Goldie 2012; Barrett 2017). That is, they are partly defined by the ways in which they unfold over time and relate to situations, behaviour, and other emotional states:

> Fear and anger, for example, can be elicited out of the blue… whereas disappointment, gratification and grief must develop out of some antecedent attitude that can be disappointed, gratified, or aggrieved… Emotions like hope, fear, and anger are by nature unstable, because they motivate behavior, or are elicited by circumstances, that ultimately lead to their extinction… By contrast, grief and gratification are stable, because their eliciting conditions and resulting behaviors are not conducive to change. (ibid., p. 15)

As a result, the emotional cadences that are constitutive of narrative understanding have important constraints on how they can and cannot unfold. Not just any patchwork of affective responses will do—the concluding emotion must be one that is capable both of evolving out of and resolving the range of emotions engendered by the story, and of providing a coherent perspective on the story's situations and events. We exercise narrative understanding when our grasp of the elements of a situation or sequence of events is shaped by an emotional cadence—either via occupying a stable emotional state contributing to a perspective which makes sense of a prior sequence of events or feelings, or anticipating how unfolding events and feelings will resolve into such a perspective.

Velleman (2003, 2009) has much of interest to say about how this view relates to theories of literary narrative, and the particular cadences distinctive of various literary and dramatic genres. Our focus here, though, is on Velleman's claim that we usually understand *ourselves* in narrative terms. This amounts to the claim that we are motivated to grasp the activities and episodes that make up our lives in terms of the kinds of familiar emotional cadences that we use to understand stories. In arguing that there is a narrative component to our self-understanding, Velleman is claiming that in making sense of our lives we look for

the kinds of emotional linkages and resolutions that are at work in successfully making sense of a story. To illustrate, consider the ways in which many puzzling patterns of apparent practical irrationality can be explained by distinguishing between these modes of self-understanding. Velleman suggests, for example, that we can account for the 'sunk cost' fallacy—'in vernacular terms, throwing good money after bad' (2007, p. 285), in terms of our propensity to seek emotionally meaningful payoffs to our projects and commitments:

> As any A.B.D. student knows, an endeavor abandoned in mid-course yields no emotional resolution, not even the resolution of disappointment. Continually abandoning one relationship or project for another would leave us not knowing (as we might say) what some parts of our lives had been about. And to know what parts of our lives have been 'about' is just to know how they fit into a story with an emotionally intelligible arc. (ibid., p. 286)

Likewise, Velleman suggests, our propensity to see misfortune as something that can be learned from can be understood in terms of a drive towards narrative self-understanding:

> If life hands you a lemon, the instrumentally rational course may be to throw it away and look for a kumquat instead. Why feel obliged to make something out of a lemon just because you've been handed one? The answer is that a misfortune can be given meaning by a narrative that incorporates it into the remainder of your life, during which its bitterness is still detectable but somehow sweetened. (ibid.)

In each of the above cases, an apparent failure of practical rationality is explained by supposing that our choices aim at an outcome that completes a familiar emotional cadence, thereby providing a perspective according to which prior events are rendered intelligible in terms of their place in a narrative structure. If our overriding practical interest were in totting up utility points then ditching a failing relationship or research project might often be our best option. But if we have a practical interest in attaining a kind of self-understanding that involves

threading the events of our lives into a unified narrative, then the time and suffering costs incurred by persevering needn't put us off. For if we can successfully fight our way to feeling good about our relationship or research, this affords us a perspective from which the struggles along the road are revealed as necessary landmarks or obstacles in the journey of our life. This is the sort of story we all know and understand, even if its instances don't always make strict logical sense.

Let me note a final point in favour of Velleman's view of the role of narrative understanding in our lives. Velleman's reflections on the narrative structure of practical reason can be placed within a long tradition of arguing that important aspects of our selfhood, or self-understanding, must be accounted for via appeal to narrativity (MacIntyre 1981; Taylor 1985; Schechtman 1996; Korsgaard 2009; Rudd 2012). As Jongepier (2016) notes, however, narrative accounts of selfhood face a dilemma—they appear to be caught between *chauvinism* on the one hand, and *triviality* on the other hand. The claim that humans necessarily understand their lives as a story courts accusations of chauvinism—critics of narrative accounts argue that other ways of living and understanding one's life are available, and that these may be more psychologically or morally healthy (Strawson 2004; Lamarque 2007). Galen Strawson, for example, claims 'I have no sense of my life as a narrative with form, or indeed as a narrative without form. Absolutely none' (2004, p. 433). Given the apparent existence of such non-narrative psychologies, Strawson and others argue that claims that people do or should understand their lives as they would understand a story 'are not universal human truths… Their proponents, the narrativists, are at best generalizing from their experience in an all-too-human way' (Strawson 2017, p. 124). In light of such accusations of chauvinism, a natural response is to claim that narrative self-understanding need only be *implicit*—somehow evinced in the structure of our lives or psychology, rather than being a self-consciously held schema for making sense of ourselves (e.g. Stokes 2010; Davenport 2012; Schechtman 2014). However, such accounts face the challenge of spelling out a non-trivial sense of narrativity that nonetheless accommodates the sincere reports of those who find no trace of a story-like structure in their self-understandings. If the narrative structure of Strawson's self-understanding

is buried so deeply that he can't find it within himself in almost 20 years of trying (Strawson 1999, 2017), why suppose it's there at all?

Velleman's account simultaneously avoids the charges of chauvinism and triviality. Conceiving of narrative understanding in terms of grasping emotional cadences suggests how someone might exercise practical rationality in a way that aims at narrative coherence, but do so without having an explicit sense of their life as a continuous unfolding story. If Velleman is right then the question raised by those, like Strawson, who find a narrative self-conception implausible is whether their practical reasoning is often or always motivated by a drive to resolve an emotional cadence. We might be convinced by Velleman's substantive arguments that such drives play an important role in an agent's psychology while denying that this fact must be introspectively obvious to us. The claim that the involvement of emotional cadences is what distinguishes narrative understanding from mere causal-nomological understanding has likewise required substantive argument. Velleman's claim that human agency involves narrative understanding in the form of a drive to act in ways that resolve emotional cadences is thus non-trivial. It is non-chauvinist insofar as it is compatible with granting full human agency to those who have simply failed to notice this structure in their own psychology, so long as that structure does in fact obtain. Velleman's view does imply, however, that the absence of this narrative strand of practical rationality constitutes a defect:

> [C]aring about narrative coherence is part of human rationality. Someone who manages his life as he would a stock portfolio is not just unsentimental; he's lacking in a virtue of practical reason that's available to human beings, who can understand their lives in both causal-psychological and narrative terms. (2007, p. 286)

But this position rests on more than the hasty extrapolation from quirks of individual psychology which grounds Strawson's accusations of 'chauvinism'. If Strawson or a kindred spirit finds no trace of a drive towards narrative coherence in their psychology, the burden of proof now appears to rest with them—they must explain away the appearance that there is something deficient, with respect to the virtues that characterize

human practical rationality, in managing the projects and commitments of one's life like a stock portfolio.[3]

So the sincere self-reports of Strawson and his ilk need not count against Velleman's claims about the role of narrative understanding in our lives. Moreover, as we have begun to see above, and will now consider further, Velleman argues persuasively that such understanding plays a crucial role in a form of autonomous agency that is plausibly a constitutive aspect of human personhood.[4]

* * *

So far I have mostly been approvingly summarizing some lines of Velleman's thinking about practical rationality. In a moment I want to begin make some trouble for him. Recall that Velleman originally began appealing to narrative self-understanding as a homely redescription of the causal-psychological self-understanding he argued elsewhere was a constitutive motive of practical rationality. But, as we saw in the last section, there are good reasons to think that understanding events, including the unfolding events of our lives, in narrative terms is distinct from understanding them in terms of mere causal and rational consistency. Velleman thus comes to think that the psychology of human agents is fractured into two 'independent and potentially competing modes of practical reasoning' (2007, p. 287), aimed at two distinct modes of self-understanding. In itself, there's nothing wrong with this. Perhaps it's not nice to be told that we have a fractured psychology, but sometimes the truth hurts. And given the widespread appearance

---

[3]Velleman (2009) appears to have changed his mind on this score, claiming that the drive towards narrative self-understanding is an optional supplement to the drive towards causal-psychological self-understanding for practical reasoners, 'at least to some extent' (p. 204). Below I argue that this is a mistake—it is more plausible to construe causal-psychological understanding as built on a foundation of narrative understanding.

[4]The degree of compatibility between Velleman's views on practical rationality and selfhood and the narrative theories cited above is a delicate issue. A key tenet of Velleman's work on personhood is that 'self' is a multi-purpose reflective pronoun, rather than the designator of a single entity or topic of enquiry (Velleman 2006a). In this chapter our concern is with the mode of reflexive self-understanding that Velleman argues is constitutive of agency. But Velleman is at best agnostic about the relationship between the self-understanding constitutive of agency (understood, roughly, in terms of a drive to act in ways that make sense to ourselves) and the kinds of self-conception with which narrative theorists are frequently concerned (e.g. conceiving of oneself as a loving spouse, committed parent, or hopeless loser). See e.g. Velleman (2001a, 2007, 2009).

of practical irrationality, perhaps the diagnosis of a fracture should not surprise us.

However, the admission that we're dealing with two modes of practical rationality rather than one appears to undermine some important aspects of Velleman's view. First, once it becomes clear that narrative self-understanding is distinct from causal-psychological self-understanding, the appeal to narrative self-understanding can't serve its original purpose of assuaging the worries of critics who accuse Velleman's causal-psychological account of practical reasoning of an implausible intellectualism. Second, given that narrative self-understanding was initially supposed to be equivalent to causal-psychological self-understanding it could borrow its plausibility as a source of practical rationality from Velleman's earlier work on the role of causal-psychological self-understanding in practical rationality. Now that we've denied the equivalence of these modes of self-understanding, that borrowed plausibility must be returned and narrative self-understanding must pay its own way as a legitimate source of practical reason. Can it do so? Not, I will argue, unless we understand causal-psychological understanding as build upon a bedrock of narrative self-understanding. Narrative understanding, it will emerge, should be seen as essential to practical rationality, and the regular cadences that pattern the ebb and flow of our emotional lives are essential to narrative understanding. To see why this is so, however, we first need to know a little more about the motivations for Velleman's causal-psychological view.

4   Three Ways to Build an Agent

Consider (following Velleman 2001b) three ways in which we might go about building an agent. Let's give ourselves a head start by beginning with a creature that is a bundle of competing drives or motivations. How should we modify such a creature to make it capable of agency like our own? One simple strategy is suggested by Donald Davidson's (1963, 1980) influential causal theory of action—for a happening to qualify as an action, it is necessary and sufficient that it be caused in an appropriate way by an intention to act. Your reading these words is an exercise

of your agency because it is a happening caused by your intentions. Your yawning, rumbling stomach, or sagging eyelids are not exercises of agency because they are not so caused. To build an agent, then, perhaps we need only to take a creature that can have intentions and move around, and ensure it is wired up such that some of its movements will be appropriately caused by its intentions. Those movements will qualify as actions.

But note the looming difficulties of spelling out what it is for an action to be *appropriately* caused by an intention. Harry Frankfurt (1975, 1988) points out that our activities can be caused by our desires and intentions without our feeling that those activities are expressive of our will—we can be *alienated* from some of the forces that motivate us. This morning I desired to hit the snooze button on my alarm for a third time and formed a corresponding intention that caused me to do so. But my desire for a third snooze was competing with other, worthier desires: to get up, go for a run, get working and generally achieve great things. As often happens, the desire to snooze won out. But its winning out was, so it seems, an imperfect reflection of my will and agency— if *I* had been allowed to pick the winning desire, I would have gone with 'achieve great things'. Even if we don't wish to say that agency was wholly absent in my snooze-button pressing, it looks like a poor paradigm case for understanding what's distinctive about human action. In fleshing out what it is for an action to be 'appropriately caused' by an intention, our blueprint for designing an agent should allow for the distinction between akratic actions like my snooze-button pressing and those that are more authentic expressions of the agent's will. Frankfurt's hierarchical model of agency tries to do this by appealing to higher-order desires or volitions. This morning my first-order desire to snooze trumped its worthier competitors. But it was at best a partial exercise of agency, since I had a *second*-order desire that my first-order desire to get up and achieve great things should win out. On Frankfurt's hierarchical model, the happenings that are paradigmatic exercises of agency are those caused by a first-order desire that the agent *identifies with* via having a second-order desire to be actuated by it. Our second agent-building strategy, then, involves adding a layer of second-order volitional states to our creature's psychology which can harmonize

(or clash) with the first-order states that actuate its behaviour. When such harmony obtains, we have an exercise of agency.

This looks better—but if we were dissatisfied with our first strategy, we shouldn't accept our second as it is. We objected to Davidson's causal model by noting that we can feel alienated from the intentions that end up moving us—we do not recognize them as authentic expressions of our wills. Velleman (1992, 2001a, drawing on Watson [1982] and Bratman [1999]) notes that we can likewise feel alienated from our second-order desires or identifications. I might find myself—in fact, most mornings *do* find myself—giving a second-order endorsement to a first-order desire for further snoozing that I find mildly shameful. It appears that some second-order endorsements of first-order desires can reflect an agent's will better than others. A second-order endorsement might be a product of ennui, depression, inebriation or mind-control. An account of agency should try to accommodate a sense in which the behaviours flowing from such endorsements are imperfect exercises of agency. Simply appealing to higher levels of reflective endorsement won't help, unless we are given a reason why the endorsements at some particular level can't be produced in the defective ways suggested above. We need, it seems, an account of what it is for an *agent* to identify with an actuating desire, as opposed to that identification being something that simply befalls them.

This leads to our third strategy for agent-building—Velleman's own. Velleman traces the failures of our first two strategies to the appearance that the psychological dynamics they suggested as constitutive of agency can obtain without the agent being suitably involved. The Davidsonian account understands agency as the government of your movements by intentions—but what if those intentions aren't the ones that *you,* the agent, want to be governed by? The Frankfurtian account understands agency as movements that are governed by intentions which have received higher-order endorsement—but what if those endorsements are not ones that *you,* the agent, recognize as expressive of your will? Velleman's strategy is to specify the psychological dynamics of a kind of motivated activity that cannot, by definition, unfold independently of the agent's involvement. Such dynamics would need to encompass whatever state, process, or organization, we thought

constituted agency. So what *does* constitute agency? Our objections to the previous agent-building strategies have relied on understanding the agent as capable of standing back and surveying possible determinants of behaviour (intentions, motives, higher-order volitions), and taking a stance on which should prevail. This surveying and stance-taking, thinks Velleman, is simply the functional role of an agent: 'that of a single party prepared to reflect on, and take sides with, potential determinants of behavior at any level in the hierarchy of attitudes' (1992, p. 19). Given this functional characterization of agency, we can't identify behaviour-governing intentions or second-order desires with the agent, since these are aspects of the agent's psychology which the agent can scrutinize and take a stance on. So, what plays the role of the agent?

> What mental event or state might play this role of always directing and never merely undergoing such scrutiny? It can only be a motive that drives practical thought itself. That is, there must be a motive that drives the agent's critical reflection on, and endorsement or rejection of, the potential determinants of his behavior, always doing so from a position of independence from the objects of review. Only such a motive would occupy the agent's functional role, and only its contribution to behavior would constitute his own contribution. (Velleman 1992, p. 119)

Velleman argues that the motivation constitutive of agency is a drive to act in accordance with reasons (to 'act reasonably' for short).[5] You can stand back from your desires and critically reflect on them; you can also stand back from this process of critical reflection and reflect on whether you agree with its procedure or results. But you, as an agent, cannot disassociate yourself from the drive to act reasonably that informs the results of any such standing back. Suppose that, in your practical deliberation, you find yourself reflecting on your drive to act reasonably, and somehow decide that it is a bad thing—starting from now, you will no longer allow the drive to act reasonably to motivate you. This resolution

---

[5]Note that an agent might be motivated by this drive without having the concept of acting for reasons. It suffices that the drive in fact propels the agent to act reasonably, whether or not the agent understands themselves in these terms (Velleman 1992, pp. 120–121; 2001b, pp. 26–32).

is impossible to put into practice. In trying not to be motivated by the drive to act reasonably you will be trying to do what you judge yourself to have best reason to do—you will be trying to act reasonably, and your attempt will have been self-undermining. Suppose instead, then, that you arrange to have yourself undergo targeted brain damage, removing only your drive to act reasonably. Would you have succeeded in your resolution? It seems not—*you*, the agent that made the resolution, are no longer around to have your actions be motivated, or fail to be motivated, by any drive. In your place is a creature who can still step back, survey the various potential determinants of its behaviour, and ponder what it would be most appropriate to do in light of this survey. But the results of this deliberation need bear no systematic relation to what the creature goes on to do, lacking as it does the drive to act reasonably. What this creature goes on to do is, in the sense relevant for agency, no longer under its control—however its competing drives and impulses are channelled into a single coherent course of action, we have no reason to think that this channelling is a product of an aspect of its psychology with which we can identify the creature qua agent. In giving up your drive to act in accordance with reasons you gave up the motive in virtue of which your activity could intelligibly be traced back to you as an agent.

On Velleman's model, then, adding the motivational force of the drive to act reasonably to one among the agent's competing motives ensures that the resulting behaviour constitutes an exercise of agency:

> [W]hen a desire appears to provide the strongest reason for acting, then the desire to act in accordance with reasons becomes a motive to act on that desire, and the desire's motivational influence is consequently reinforced… This latter contribution to the agent's behavior is the contribution of an attitude that performs the functions definitive of agency; it is therefore, functionally speaking, the agent's contribution to the causal order. (1992, p. 121)

In place of a Frankfurtian higher-order endorsement, it is the addition of the motivational force of the drive to act reasonably to one of your other motivations that constitutes your 'identifying' with it—in

the sense of its belonging to you as an agent. We now have our third and final blueprint for building an agent. We began with a bundle of competing drives and motives and, following Davidson, hooked things up such that our creature's intentions can function as the causes of its behaviour. Following Frankfurt, we then added a capacity for the creature to step back and survey the potential psychological determinants of its behaviour, and give or withhold its endorsement of them. Finally, following Velleman, we added the background drive to act reasonably. This drive can now determine how exercises of the capacity to step back and survey are translated into behaviour, by joining its motivational force with that of the course of action that best accords with its aims. My evening decision to eschew the pub and go for a run, and Gilbert's decision to head to the closet for fresh batteries are, for Velleman, exercises of agency because our prior resolutions are modulating the competition between our present motives in a way that is guided by a drive to act reasonably that is a constitutive requirement of agency.

5   Two Sources of Practical Reason

The above sketch of Velleman's interrelated conceptions of agency and practical reason brings some problems for his account of the relationship between causal-psychological and narrative self-understanding into view. We have already noted that Velleman revised his earlier view that narrative understanding was a homely redescription of causal-psychological understanding, now holding that 'causal explanation and storytelling convey fundamentally different modes of understanding' (2009, pp. 185–186), and thus that practical rationality is driven by the pursuit of two potentially competing drives towards distinct ways of making our lives intelligible to ourselves. Velleman is thus unable to appeal to narrative self-understanding to allay worries about the intellectualist, or cognitively demanding character of his causal-psychological account of practical rationality. Is this a problem? That depends, I think, on our explanatory interests. Velleman's chief concern is with the logical structure of mature human practical rationality. Maximizing one's causal-psychological intelligibility to oneself is posited as the ideal at which

such practical rationality aims. The fact that this ideal is one that few people would consciously recognize as animating their behaviour, and is difficult or impossible to obtain, need not count against Velleman's characterization. I, for example, could not tell you the statistical rules and regularities which govern optimal betting patterns in poker, and am confident that no game of poker I ever play will approximate those patterns—but it might nonetheless be true that my poker playing activity qualifies as an attempt to play poker insofar as it attempts to track those rules and regularities, and can be assessed as going better or worse insofar as it succeeds or fails at this.

However, if our goal is to naturalistically explain the capacity for practical rationality as it is realized in and exercised by humans, we should be less sanguine about the idealizing and intellectualist features of Velleman's view. Consider how we might attempt to put the previous section's blueprints for building an agent into practice. Given current cognitive-scientific knowledge, it's plausible that we could indeed engineer systems with competing sets of motives, and even with the higher-order meta-motives to which Frankfurt appeals. But Velleman's suggested addition of a global background motivation to act reasonably that modulates the interactions of a system's existing motives and behaviours is more mysterious—this is not a component that engineers of real or artificial systems have lying around in their workshops. As a piece of construction advice it is little more helpful than being told to add an immaterial soul, or a perpetual motion machine. We might thus agree that Velleman is right about the psychological structure that our completed agent will instantiate, while objecting that his account does nothing to help us with the naturalistic goal of understanding how agents can be constructed out of simple worldly materials.[6] Velleman's shift from describing his view of practical reason in terms of mature rational agents and idealized psychological theories to simple robots and

_____

[6]In the interests of simplicity I'm writing as if we can legitimately pull apart the explanatory projects of characterizing the constitutive structure of some phenomenon and spelling out how that phenomenon does or could arise in nature. But note that many will be inclined to resist this separation, holding that empirical inquiry and data are our best guides to metaphysical structure. Such folks should be especially sceptical of Velleman's intellectualism about practical reason.

viscerally understood stories looked like a promising step towards this naturalistic goal. But the admission that this shift represents a change in topic rather than a redescription shows that no progress has been made. If we care about understanding how agency emerges in the natural world then while Velleman might show us where our account must end up, he can't tell us how to get there.

Intellectualist worries about Velleman's view thus resurface once we sever the link (or come to believe that there was no link to be severed) between causal-psychological and narrative self-understanding. Severing this link also appears to undermine the claim that a drive towards narrative self-understanding can make a constitutive contribution to agency. We saw in the last section that Velleman posits the drive towards causal-psychological self-understanding as the crucial ingredient of agency because he thinks it can be plausibly identified with the functional role of the agent—it is this drive that animates the activity of weighing up and selectively modifying the force of the motives vying for control of behaviour. And this activity can be plausibly understood as the essential form of agency. What would happen if this drive towards causal-psychological self-understanding were replaced with a drive towards narrative self-understanding? The latter drive is equally capable of adjudicating between, and modulating the influence of, competing motives on behaviour. The only apparent difference is that this adjudication and modulation is driven by standards of narrative, rather than causal-psychological coherence. Perhaps, then, we simply have two candidate occupants of the functional role of the agent instead of one—and this dual occupancy is what accounts for the fractured nature of practical rationality. But this is too quick. Consider why the drive towards causal-psychological self-understanding seemed like a plausible candidate for the role of the agent in the first place. Velleman clarifies:

> Why would behaviour produced by such a mechanism be any more attributable to the person than that produced by other causes? The answer is that a person is somehow identified with his own rationality. As Aristotle put it, 'Each person seems to be his understanding.' Hence causation via a person's rational faculties qualifies as causation by the person himself. (2001b, p. 17)

This quote reveals the rationalist intuitions at work in Velleman's account of agency. The fact that adjudicating between, and modulating the motivational force of, competing determinants of behaviour is governed by rational standards is what makes it an intuitively plausible candidate to be cast in the role of the agent, qua rational intellect, thinks Velleman. In defence of this intuition we might note, as we did above, that an account of agency should explain the difference between behaviour governed by mere causal forces and behaviour guided by reasons. Velleman's account of agency straightforwardly accomplishes this by understanding the mechanisms that determine an agent's behaviour in terms of rational standards applied by rational faculties. This link does not appear to obtain if we try to construe a drive towards narrative coherence as a motivating force of agency. After all, our initial characterization of the distinction between narrative and causal-psychological coherence relied on stressing that a sequence of events can make narrative sense—via establishing and completing an emotional cadence—without being fully rationally intelligible. Think, for example, of the way in which myths and fantasies with supernatural elements, or narratives consisting of series of causally and spatiotemporally dislocated vignettes, can convey understanding. It thus appears on Velleman's view that we have no reason to attribute behaviour stemming from the motivational force of a drive towards narrative understanding to the agent qua rational intellect, or to view this behaviour as guided by reason in the way characteristic of agency. A drive towards narrative understanding no longer looks like a plausible source of practical rationality.

Velleman (2009) appears to modify his earlier view of the relationship of narrative and causal-psychological intelligibility in ways that bear on these complications. At the end of §3 we saw Velleman (2007) claim that lacking a drive towards narrative self-understanding (as might be manifested by managing one's life like a stock portfolio) amounts to lacking a virtue of practical rationality. But Velleman (2009, p. 204) describes narrative self-understanding as a supplement to causal-psychological self-understanding that is 'to some extent' optional, citing self-descriptions such as Strawson's (2004, 2017; §3, above) as evidence in favour of such a view. And he takes pains to emphasize that he now views narrative understanding as based on causal-psychological understanding:

> I assume that narrative intelligibility requires psychological intelligibility, but not vice versa. That is a story requires action, and action has to be intelligible as caused by the attitudes and attributes of a character, lest it count as no more than mere behavior. (2009, p. 185)

I will try to motivate an alternative conception of the relation between narrative and psychological intelligibility in the chapter's final section, below. For now, note that these qualifications to Velleman's view further undermine his claim that narrative intelligibility can function as a constitutive motive of practical rationality. Velleman now views the drive towards causal-psychological self-intelligibility as sufficient for practical rationality, with a drive towards narrative self-intelligibility as an optional supplement. Now that Velleman insists that narrative intelligibility presupposes psychological intelligibility, we can specify why there should be a link between behaviours motivated by a drive towards narrative self-intelligibility and agency: the ability to make narrative sense to oneself presupposes the ability to make psychological sense to oneself which Velleman identifies with the agent, qua rational intellect, and thereby qualifies as an exercise of practical rationality. But this is not because the drive towards narrative self-intelligibility is an independent source of practical rationality—it is because it is an optional way of manifesting the psychological drive with which Velleman originally identified the agent.

Here, then, is the situation. Velleman began with a plausible but demanding intellectualist account of the logical structure of practical rationality. He then added reference to a drive towards narrative intelligibility to his account in a way that appeared to make it less intellectualist, more naturalistically tractable, and better able to accommodate the apparent psychological diversity of ways in which we experience our activities as making sense to us. The cost of this was an apparent fracture in the structure of practical rationality. Upon the closer inspection we have given Velleman's views in this section, however, all these appearances are illusory. Narrative intelligibility, according to the most recent incarnation of Velleman's views, depends on the same capacities of the rational intellect he has always emphasized. Appealing to it thus moves us no closer to the goal of a naturalistically tractable account of

the constitution of agency. And when we act in ways that make more narrative than causal-psychological sense (as when we persist in the face of adversity) we are failing to properly realize our status as practical reasoners and thus as agents, since our actions fall short of the only standards that really matter for agency—those of causal-psychological consistency. The silver lining is that practical rationality no longer appears fractured—it has the same monolithic, intellectualist structure as in Velleman's earlier work.

6   A Strange Inversion of Practical Reasoning

The appropriate response to these problems, I think, is to reverse the relations of dependence Velleman sees between capacities for rendering events intelligible in causal-psychological and narrative terms. Velleman (2009) argues that understanding a narrative presupposes the capacity to understand events in terms of the inferentially related psychological states that caused them:

> Narrative understanding is not a substitute for causal-psychological self-understanding. The tale of Oedipus makes for a good story even if we don't understand why the prophecy of Tiresias came true; but the tale would make no sense of any kind if we didn't understand, from one scene to the next, why Oedipus was behaving as he did… Hence practical reasoning aimed at narrative self-understanding can supplement but not replace reasoning aimed at self-understanding in causal-psychological terms. (p. 203)

But this line of thinking undersells the scope of Velleman's original account of narrative understanding. That account specified a possible way of understanding the relationship between *any* sequence of events, in terms of the way in which its constituents hang together according to a familiar emotional cadence. Velleman rightly notes above that, in some cases, we must exercise relatively sophisticated folk-psychological capacities in order to grasp the events we aim to make intelligible before we can perceive any emotional cadence that unites them. Fully grasping

the emotional beats of a complex espionage thriller or Shakespearean farce requires some prior understanding of the psychological states of the characters and the ways in which they are revealed or hidden by their activities. But relations between non-psychological states can also be grasped according to their place in a familiar emotional cadence. The very locution of 'cadence' draws attention to the fact that we can exercise narrative understanding in experiencing the affective structure of non-psychological states or events such as a series of musical notes. In such a case we, the subjects of understanding, are ushered through a series of affective psychological states that determine our perspective on the object of understanding—but this process need not involve imputing psychological states to ourselves, the object of understanding, or an author behind it. Even when the object of our narrative understanding is ourselves or other agents, some behaviour can be rendered intelligible via its fit with an emotional cadence without deploying the sophisticated folk-psychological capacities emphasized by Velleman. Think of an infant learning to enjoy the activity of peek-a-boo, or simple turn-taking games. Children who are too young to reliably parse sequences of behaviour into intentional actions, or to manifest sensitivity to the potential divergence between behaviours and their psychological causes, can nonetheless come to grasp ritualized play events and the expressive behaviours they involve in terms of their place in a familiar emotional cadence. This kind of understanding is manifested by the way in which infants anticipate the next stage in the play, and their reactions when such anticipations are confirmed or frustrated. To understand events in this way is not yet to understand a narrative, or the psychological states of any of the parties involved—but it nonetheless requires exercising the mode of understanding relations between events which Velleman argues we employ when grasping the events in a narrative. The positive suggestion with which I close this chapter is that such exercises of the capacity for narrative understanding, when suitably enculturated and scaffolded, form the bedrock for the mature folk-psychological capacities which Velleman makes central to practical rationality.

Luckily, this is not a case I need to make from scratch. The past 20 years or so have seen a host of important work that aims to question

nativist, modular and cognitivist accounts of capacities for folk-psychological understanding by emphasizing the role of embodied, emotional and situational factors, and by drawing attention to various social scaffolds and dynamics that support and constrain our understanding of ourselves and others (e.g. Hurley 1998; McGeer 1996, 2007; Mameli 2001; Sterelny 2003, 2012; Zawidzki 2008, 2013). A key strand of that literature argues that our folk-psychological apparatus for interpreting behaviour serves not only descriptive and explanatory purposes, but also plays a *regulative* role. This kind of regulation is at work when Gilbert's avowal that he plans to go to the closet, or my avowal that I intend to go for a run this evening, shape our subsequent deliberation. Our respective bits of self-directed folk-psychology function not only as descriptions, but as motivations to conform to those descriptions in our future behaviour. The same regulative pressure is at work, these theorists argue, when I interpret others or when they interpret me—a community of sociable, eager-to-please, but quick-to-censure intentional agents will give rise to an emergent pressure to make one's own activities and those of one's peers conform to shared standards of intelligibility. To see why, and the role that narrative understanding might play here, let's return to our attempt to specify a blueprint for building an agent.

Suppose once again that we begin with a creature that is a bundle of competing drives and motivations. Add the capacity for that creature to occupy emotional states, understood as valenced bodily states that preferentially direct attention to specific aspects of the creature's situation. Now add the capacity for those states to link together in recognizable sequences or cadences which the creature can grasp via anticipation or retrospection (the kind of primitive capacity for narrative understanding which I suggested above may be at work when young children grasp the structure of events in simple play activities). Suppose also that the range of emotional cadences via which we can understand events is developmentally plastic—while the emotional cadences that structure understanding of some kinds of events (such as those involved in a game of peek-a-boo) might be innate, exposure to the right events and associated bodily states can inculcate a grasp of new cadences. The emotional beats that structure our understanding of complex murder mysteries or espionage thrillers, for example, are unlikely to be innately understood.

But a subject with a developmental trajectory that includes repeated exposure to the right books or films can come to anticipate or retrospect the relationships between events in such plots as unthinkingly as a child anticipates the next event in a game of peek-a-boo. Two final suppositions about our creature—first, it has an innate appetite for intelligibility, and distaste for unintelligibility. In this context, this simply means that our creature likes it when events fit familiar emotional cadences—it's usually happier when its anticipations pan out, or when its retrospection fits recent events into a familiar emotional structure (babies, for example, prefer it when games of peek-a-boo follow established protocol). Second, our creatures are sociable—they like to please, don't like to displease, and are sensitive to indications of pleasure or displeasure from their fellow creatures. Such a creature will be driven to understand its world via the emotional cadences with which it is familiar, and will acquire familiarity with new emotional cadences via repeated exposure over the course of its development.

Now, suppose further that we scaffold our creature's development in a particular way—we surround it with other sociable creatures who likewise crave intelligibility and approval, and who exhibit patterns of behaviour that is intelligible in folk-psychological ways. Suppose also that our creature's developmental trajectory involves regular interaction with caregivers who repeatedly expose it to affectively engaging narratives conveying stereotypical folk-psychological wisdom via the emotional beats of their story structures (Sterelny 2003, 2012; Hutto 2008). When our creature's development is scaffolded in this way, it will acquire a drive to make its own behaviour intelligible in terms of whatever emotional cadences structure the understanding of its fellow creatures. Perceptible regularities in the behaviours of its fellow creatures will eventually be narratively understood in terms of the reliably similar structured patterns of embodied affective reactions—emotional cadences—they produce for our creature. A helping hand in learning the most important of such cadences is provided by the creature's caregivers, who supply a regular and easily digestible diet of tried-and-tested narratives and interactive practices that highlight patterns which are particularly important for understanding and predicting behaviour in their community. Crucially, because our creatures are sociable and

intelligibility-craving, they will acquire a drive to render *themselves* intelligible in terms of the structure and patterns that characterize their understanding of their peers. They themselves are objects whose behaviour their peers are attempting to understand; as intelligibility-craving creatures, their peers will be happier when the creature behaves intelligibly by their lights; as a sociable creature, our creature has a drive to make its peers happy, and will thus acquire a motivation to behave intelligibly by the lights of their peers. For the kinds of creatures we have described, intersubjective understanding thus acquires a *regulative* as well as a *predictive* role (McGeer, Mameli, op. cit.). Given this, creatures with the features and social scaffolding described above will eventually acquire a drive to act in ways that conform to whatever standards of intelligibility are prevalent in their community.[7]

This is how a drive towards folk-psychological intelligibility could be built out of a suitably scaffolded drive towards narrative intelligibility. We have reversed the relations of dependence that Velleman (2009) argues obtain between these drives. In doing so, we have resolved the problems I raised for his account in §4. While the blueprint for creature-design just presented is more complex than the Vellemanian blueprint of §3, it is also more naturalistically tractable—it is easier to see how we could realize that blueprint with cognitive structures and processes whose operations we broadly understand. Where Velleman speaks of bolting a drive to act reasonably on to a bundle of competing motivations, I have suggested how such a drive could gradually emerge from a suitably scaffolded, developmentally plastic, motivated creature that can make sense of events via emotional cadences. While defending the

---

[7]One question that a fully developed version of the present proposal should address concerns the origins of the practices of intersubjective understanding that scaffold development. Given the design specifications above, a suitably developmentally plastic creature will acquire a drive to act intelligibly by the standards of whatever practices of intersubjective understanding are present in its community—hence sophisticated folk-psychological understanding can emerge from primitive narrative understanding. But how do sophisticated practices of intersubjective understanding come to exist in the scaffolding environment in the first place? A naturalistic story here must show how folk-psychological practices like our own can bootstrap themselves into existence from humbler beginnings given the existence of a community of creatures with the psychological structure described above. Sterelny (2003, 2012) and Zawidzki (2013) in particular have provided plausible, empirically informed suggestions about how such bootstrapping could occur.

empirical credentials of this picture is a task for another time, I think the prospects are good. In addition to the work by proponents of the regulative dimension of folk-psychology cited above, work on embodied resonance and interpersonal coordination (as suggested by e.g. Hurley 2008) and constructionist approaches to emotion (e.g. Barrett 2017) yields many suggestions about how we humans might instantiate the blueprint sketched in this section. Viewing the drive towards folk-psychological intelligibility as a culturally scaffolded drive towards narrative intelligibility helps restore the credentials of the latter drive as a legitimate source of practical rationality. On Velleman's (2009) view, the drive towards narrative intelligibility can qualify as a source of practical rationality only because it is an optional way of manifesting a drive towards causal-psychological intelligibility. But the blueprint above shows how we can agree with the rationalist intuitions motivating Velleman's account of practical rationality (whereby the drive to act reasonably is plausibly identified with the agent because there is a privileged link between personhood and rational faculties) while still holding that the drive toward narrative intelligibility is essential to agency. This is because the latter drive makes an essential contribution to the motive force of the drive to act reasonably—creatures that meet the design specifications above acquire a motive to act reasonably only because their drive towards narrative understanding takes on a particular, culturally inculcated form whereby their sociable and intelligibility-craving nature motivates them to self-regulate their behaviour according to whatever folk-psychological principles structure the intersubjective understanding of their peers. More tentatively, the blueprint above might suggest how particular instances of a drive towards narrative intelligibility can qualify as a source of practical rationality in its own right. On certain permissive social practice accounts of reasons (e.g. Rorty 1979, 1999), a good reason is simply one that most of our peers are happy to accept. Such views would allow the norms of practical rationality to be fixed by the practices of intersubjective understanding that a given community actually employs, even when those standards fall short of strict standards of calculative rationality. As we saw above (§2), an appealing feature of a narrative conception of practical rationality is that it makes sense of the appearance that we do indeed make

sense of our own activity and that of others in diverse ways—persisting in the face of adversity can appear a reasonable course of action to us even when it is likely to fail to maximize the satisfaction of an agent's projects and preferences. If the norms of practical rationality are fixed via reference to actual practices of intersubjective understanding then the blueprint above suggests how any such norms can acquire motive force for a suitable enculturated creature, regardless of how closely they align with the norm of maximizing causal-psychological intelligibility.

I have argued, then, that we can save the best features of Velleman's view of practical rationality by inverting it—rather than founding the capacity for narrative understanding on sophisticated capacities for causal-psychological understanding of ourselves and our peers, I have sketched how the latter capacities might be built out of socially scaffolded capacities for narrative understanding, and how such scaffolding can imbue the principles that govern the frameworks we use to make sense of each other with the motive force that Velleman argues is constitutive of practical rationality. On this picture, practical rationality is essential to agency, narrative understanding is essential to practical rationality, and the regular cadences that pattern the ebb and flow of our emotional lives are essential to narrative understanding. It is in virtue of narrative cadences that move us emotionally that we can be moved agentially. That's my story, and I'm sticking to it.

References

Anscombe, G.E.M. 1957. *Intention*. Oxford: Blackwell.

Barrett, L.F. 2017. The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization. *Social Cognitive and Affective Neuroscience* 12 (1): 1–23.

Bratman, M. 1999. Identification, Decision, and Treating as a Reason. In *Faces of Intention: Selected Essays on Intention and Agency*, 185–206. Cambridge: Cambridge University Press.

Davenport, J. J. (2012). *Narrative Identity, Autonomy, and Mortality: From Frankfurt and MacIntyre to Kierkegaard*. New York: Routledge.

Davidson, D. 1963. Actions, Reasons and Causes. *Journal of Philosophy* 60 (23): 685–700.

———. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.

Dennett, D., and N. Humphrey. 1998. Speaking for Our Selves. In *Brainchildren: Essays on Designing Minds*, 185–206. Cambridge, MA: MIT Press.

Frankfurt, H. 1975. Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68 (1): 5–20.

———. 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.

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

———. 2012. *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford: Oxford University Press.

Hurley, S. 1998. Vehicles, Contents, Conceptual Structure and Externalism. *Analysis* 58 (1): 1–6.

———. 2008. The Shared Circuits Model (SCM): How Control, Mirroring, and Simulation Can Enable Imitation, Deliberation, and Mindreading. *Behavioral and Brain Sciences* 31 (1): 1–22.

Hutto, D. 2008. *Folk-Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge, MA: MIT Press.

Jongepier, F. (2016). Towards a Constitutive Account of Implicit Narrativity. *Phenomenology and the Cognitive Sciences* 15 (1): 51–66.

Korsgaard, C. 2009. *Self-Constitution: Agency, Identity and Integrity*. Oxford: Oxford University Press.

Lamarque, Peter. 2007. On the Distance Between Literary Narratives and Real-Life Narratives. *Royal Institute of Philosophy Supplement* 60: 117–132.

Lazarus, R.S. 1991. *Emotion and Adaptation*. New York: Oxford University Press.

Lewis, M.D. 2005. Bridging Emotion Theory and Neurobiology Through Dynamic Systems Modeling. *Behavioral and Brain Sciences* 28 (2): 169–193.

McGeer, V. 1996. Is 'Self-knowledge' an Empirical Problem? Renegotiating the Space of Philosophical Explanation. *Journal of Philosophy* 93 (10): 483–515.

———. 2007. The Regulative Dimension of Folk Psychology. In *Folk Psychology Reassessed*, ed. D.D. Hutto and M. Ratcliffe, 137–156. Dordrecht: Springer.

MacIntyre, A. 1981. *After Virtue: A Study in Moral Theory*. London: Duckworth.

Mameli, M. 2001. Mindreading, Mindshaping, and Evolution. *Biology and Philosophy* 16: 597–628.

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

Rorty, R. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

———. 1999. Ethics Without Principles. In *Philosophy and Social Hope*, 185–206. London, UK: Penguin Books.

Rudd, A. 2012. *Self, Value and Narrative: A Kierkegaardian Approach*. Oxford: Oxford University Press.

Schechtman, M. 1996. *The Constitution of Selves*. Ithaca: Cornell University Press.

———. (2014). *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford University Press.

Slaby, J. 2008. Affective Intentionality and the Feeling Body. *Phenomenology and the Cognitive Sciences* 7 (4): 429–444.

Sterelny, K. 2003. *Thought in a Hostile World*. Oxford: Blackwell.

———. 2012. *The Evolved Apprentice*. Cambridge, MA: MIT Press.

Stokes, P. (2010). What's Missing in Episodic Self-Experience? *Journal of Consciousness Studies* 17(1–1): 119–143.

Strawson, G. 1999. The Self and the SESMET. *Journal of Consciousness Studies* 6 (4): 483–518.

———. 2004. Against Narrativity. *Ratio* 17 (4): 428–452.

———. 2017. *The Subject of Experience*. Oxford: Oxford University Press.

Taylor, C. 1985. Self-Interpreting Animals. In *Philosophical Papers I: Human Agency and Language*. Cambridge: Cambridge University Press.

Velleman, J. David. 1989. *Practical Reflection*. Princeton: Princeton University Press.

———. 1992. What Happens When Someone Acts? *Mind* 101 (403): 461–481; Reprinted in Velleman (2014), pp. 100–124.

———. 2001a. Identification and Identity. In *The Contours of Agency: Themes from the Philosophy of Harry Frankfurt*, ed. S. Buss and L. Overton. Cambridge, MA: MIT Press; Reprinted in Velleman (2006a), pp. 330–360.

———. 2001b. Introduction. In *The Possibility of Practical Reason*; Reprinted in Velleman (2014), pp. 11–46.

———. 2003. Narrative Explanation. *Philosophical Review* 112 (1): 1–25.

———. 2005. The Self as Narrator. In *Autonomy and the Challenges to Liberalism: New Essays*, ed. J. Anderson and J. Christman. Cambridge: Cambridge University Press; Reprinted in Velleman (2006a), pp. 203–223.

———. 2006a. *Self to Self: Selected Essays*. New York, NY: Cambridge University Press.

———. 2006b. The Centered Self; Velleman (2006a), pp. 253–283.

———. 2007. Reply to Catriona MacKenzie. *Philosophical Explorations* 10 (3): 283–290.

———. 2009. *How We Get Along*. New York, NY: Cambridge University Press.

———. 2014. *The Possibility of Practical Reason*, 2nd ed. Ann Arbor, MI: Maize Books.

Watson, G. 1982. Free Agency. In *Free Will*, ed. G. Watson, 205–220. Oxford: Oxford University Press.

Zawidzki, T. 2008. The Function of Folk Psychology: Mind Reading or Mind Shaping? *Philosophical Explorations* 11 (3): 193–210.

———. 2013. *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press.

8

Disorientation and Cognitive Enquiry

Owen Earnshaw

1   Introduction

At first sight it may seem obvious that the experience of disorientation would be highly disruptive to an enquiry and may very well lead to the abandonment of cognitive projects. However, there is a growing accumulation of research now emerging, which suggests that the experience of disorientation or bewilderment is somehow central to at least certain cognitive inquiries. Mulhall (2012) gives the example of Heidegger's (1962) *Being and Time* where a sense of disorientation is elicited and intentionally maintained throughout the book. Mulhall claims this is necessary in order to acknowledge that 'Dasein displays itself as a limited whole only against an ungraspable background' (Mulhall 2012, p. 129). He also gives the example of Wittgenstein who famously claimed that 'A philosophical problem has the form "I don't know my

O. Earnshaw (✉)
Durham University, Durham, UK
e-mail: owen.earnshaw@dur.ac.uk

way about"' (Wittgenstein 1963, §123), which again seems to suggest that disorientation has an intimate relation to the ability to proceed with a philosophical enquiry. Ami Harbin (2016) argues in a related vein that certain events such as illness, trauma, racism and experiences of queerness can cause in us the emotion of disorientation and this emotion can have the beneficial moral effect of tenderizing us to the vulnerabilities of our self and others. Finally, there is evidence in the psychological literature of a common process of post-traumatic growth that can occur after a devastating life crisis involving disorientation and can lead to an enhanced appreciation of life (Tedeschi and Calhoun 2004), suggesting that disorientation can be existentially beneficial. All these examples highlight that disorientation can be pivotal in our sense-making relation to the world. It therefore seems appropriate to examine the relation of disorientation to cognition in more detail to see if this once neglected emotion can be of positive value in our inquiries.

In this chapter, I will argue that the experience of the emotion of disorientation should be a background affect in any intellectual enquiry, both motivating the enquiry and being necessary to instill certain epistemic virtues in the inquirer and can also play the role of an indicator of when the project threatens to traverse the boundary of sense. I will firstly elaborate how disorientation can be understood as an emotion and the type of emotion it is, namely what aspect of the world it makes salient. I will argue that it is an emotion that is evoked through the encounter with what we might want to call 'mystery'. I will then expand on my claim that disorientation has a role in cognitive enquiry as an indicator of where the boundary of sense has been overstepped by looking at disorientation, mystery and nonsense. It will then be necessary to look at how an enquiry can maintain a relation to the possible interruption of disorientation and what epistemic virtues it is necessary to be open to and responsive to from the experience of disorientation when following through a line of enquiry. Lastly, I will discuss the practical consequences of this study outlining what this perspective on disorientation means for carrying out philosophical studies, how it should inflect our educational practices and what lessons can be learnt in terms of psychopathology and recovery from trauma.

2   Disorientation as a Neglected Emotion

In this section I will extend Mulhall's and Harbin's analyses of disorientation and attempt to elucidate the emotion that is at the heart of these experiences and show how this emotion is one which takes 'mystery' as its object. Disorientation can be phenomenologically characterized as a feeling of losing one's coordinates for action. The feelings associated might include unease, anxiety, vertigo and a loss of a sense of familiarity. As Otto describes it referring to what he calls the *mysterium*,

> It is through this positive feeling-content [disorientation] that the concepts of 'transcendent' and 'supernatural' become forwith designations for a unique 'wholly other' reality and quality, something of whose special character we can *feel*, without being able to give it clear conceptual expression. (Otto 1950, p. 30)

The key point from the quote is that the experience of disorientation leaves us inarticulate initially as we feel unable to apply concepts to the experience, hence, leaving us without coordinates for going on. The emotion described here is also typically evoked by circumstances that exhibit something highly unexpected that is emotionally salient in some way. Trauma can be a common way that disorientation is induced. A couple of examples that we might look at that exemplify the emotion of disorientation are firstly, patients finding out that they have been diagnosed with cancer or secondly, the experience of those suffering from schizophrenia. In each case the everyday is stripped of its normal background meaning for a time and the person experiences an intense feeling of being lost. One patient who received the diagnosis of cancer describes the experience,

> When the doctor confirmed my illness, I didn't know what to do. I was totally shocked and my mind went blank. I couldn't believe it and I cried a lot. (Tang et al. 2007, p. 87)

This highlights that disorientation can be a shock and leaves us without support in the face of the disorientating event. The usual coordinates,

that structure our world, seem to be disabled by such devastating news and we are left at a loss as to how to go on. This description might be seen as the extreme pole of disorientation that can be elicited in many other scenarios, however, this example might be seen as a paradigm for the experience of disorientation. Noticeably it is an experience of the mind going blank for a while that is the main component of the experience i.e. disorientation is always experienced with an initial inarticulacy and finding words that are fitting to the experience is the important work of using the emotion of disorientation.

Moving on to our second example, Ratcliffe gives the example of the change in the structure of experience that might be found in schizophrenia,

> Suppose, for example that one lost the usual sense of structured anticipation, the background that is more usually taken for granted. All objects of experience might then appear as "unexpected", with surprise and bewilderment being the all-encompassing form of one's world rather than occasional responses to specific occurrences in the world. (Ratcliffe 2012, p. 483)

This suggests that the experience of disorientation could be generalized and become a regular aspect of the world in conditions like schizophrenia. The loss of coordinates might then become a regular experience leaving the door open to the formation of a delusional reality at odds with everyday reality. I will return to the example of trauma and schizophrenia specifically in the final section of the chapter and look at how they can be interpreted in more depth. Although these experiences of disorientation may be paradigmatic they seem distant from common experience and so it may be helpful to relate this characterization of disorientation to more typical experiences that can be found in carrying out an enquiry. One intellectual enquiry that seems pertinent here in looking at philosophical projects is examining what our own death means to us. As Mulhall puts it,

> [I]f being mortal is a matter of every moment of our lives being internally related to utter nullity or annihilation, that amounts to their being internally related to that which is beyond our comprehension… [D]eath is our

utter nonexistence: so death is not an event in life, not even the last. But the human capacity to comprehend anything is (on Heidegger's account) allowing it to manifest itself to us as it is in itself; so if our death is not something we encounter, it is something we cannot possibly comprehend. (Mulhall 2012, p. 129)

If we embark upon an enquiry and find that we run up against something we cannot possibly comprehend we are left feeling disorientation; a sense of losing our coordinates as to how to go on. The example that Mulhall elaborates spells out that some inquiries bring us to the edge of a mystery and the emotion of disorientation alerts us to the fact of this. Relating this to earlier examples of trauma and mental illness we can see that they too involve a relation to mystery. In the case of the cancer patient the mystery can be similarly that of one's own death and how to go on in the face of the certainty that one is dying. The case of schizophrenia can be seen as the encounter with mystery in even the most commonplace experience; the world shows up as being essentially mysterious.[1]

To bring this back to cognitive enquiry in general we can find ourselves at the edge of mystery whenever we come across a question that seems unanswerable. Some questions in philosophy that can leave us at a loss as to how to respond include, how does the mind relate to the body? How can we answer the skeptic? How can we justify inductive practices? And so on. Philosophical research seems to occur in the midst of mystery and to Mulhall disorientation is the appropriate emotion that acknowledges a discipline that deals with trying to give answers to seemingly unanswerable questions. If any discipline asks questions of increasing generality about their subject matter they will eventually come across mystery and disorientation. So, to summarize this discussion, we might agree that disorientation is an emotion that registers when we have reached the limit of an enquiry. It is found in experiences when the facts of a situation force us to confront an aspect of the

---

[1]For a discussion of emotion in relation to psychosis and how to find the words to express this emotion, see Earnshaw (2018).

world that seems to exhibit mystery. It is now necessary to look at what role this emotion can play in our ongoing attempts to make sense of the world and the importance it plays in avoiding illusion and error in our inquiries.

3   Feeling the Limits of Sense

We now need to chart the role disorientation can play in cognitive enquiry. If we look at the structure of questioning, the very act of asking a question would seem to have a relation to disorientation and, indeed, some questions can leave us feeling disorientated if we don't immediately know the answer. These questions lead us to a point we have never traversed before and so we do not know our way about in this particular intellectual space of reasons. We are at a loss and the emotion itself gives us the experience of having no immediate support for answering the question intelligibly. One way to deal with the disorientation of such a question is to backtrack to find where we made the wrong turning into unfamiliar territory and another way is to try to relate the question to something we do know. However, both of these ways of dealing with disorientation are ways of covering it over. This may seem like a sensible strategy in the face of disorientation and a loss of bearings, but other options are available and it will be worth exploring these to see what role disorientation can play in our inquiries.

The way to acknowledge disorientation is to say 'I don't know' and this confession is an admission that at this point we have come up short at the boundary of mystery. In seeing where our sense making has run out we are enabled to start again with an appreciation of where mystery lies. Mystery is intimately related to nonsense. A mystery is something we cannot further articulate and in trying to say something about it we invariably produce nonsense propositions. This is not to say that the mysterious is inherently trivial or of little value but that rather we have difficulties expressing anything about it even though we feel there is an immense amount of value in the direction we cannot quite point at. And this highlights the importance of disorientation. Not having our bearings in a certain intellectual space means that we have to forge

new paths so that the mysterious can be registered through what we say about something else or through what we do not say.[2] Mystery can only be shared in an emotional attunement to certain features of the world found in disorientation. We can try to say something about this attunement, which necessarily results in nonsense or else we can find a new starting point with the newly found awareness of the limits of our enquiry.

Nonsense is constant threat to inquiries and it is shown to be present in the features of discourses found in contradictions, aporias and fantasy. Each feature can be brought to our attention through disorientation. Wittgenstein's method in philosophy was very sensitive to these features and he highlighted that nonsense can very often be disguised (Wittgenstein 1963, §464). If a feeling of disorientation helps us to pinpoint where in our enquiry we have drifted into nonsense, we must be careful about how we articulate the insight it gives us. To return to Heidegger's (1962) discussion of our own death, the disorientation here alerts us to the threat of nonsense.[3] We can look at Heidegger's way of dealing with this material to see if he successfully avoids unintelligibility. Firstly he acknowledges that our relation being-toward-death is a relation to an impossible possibility. That being the case it is not something we can straight forwardly have a concept of. Heidegger avoids nonsense by suggesting that our relation to this mystery should inflect how we live our lives. That is, he does not try to articulate the mystery directly, but implies that how we relate to our lives can express the existence of a mystery here. In other words we articulate the mystery through making something else the subject matter of our concern in accordance with a particular treatment of that subject matter that aims to highlight that there is a mystery in the region. His particular account

---

[2]'So we grope for metaphors and analogies…that give an intimation at least of the contours of mystery' (Cooper 2017).

[3]Heidegger (1962) makes a convincing case that in any inquiry our attunement [befindlichkeit] through our emotions and moods provides the orientation for the inquiry. In the case of disorientation we might want to say that the mood provides us with the sense that we have come across a mystery so the average, everyday understanding of the phenomenon is not straightforwardly applicable.

is that it is necessary to register the mystery of our own death by living a life that is truly individual, a life that is chosen in all its particularity. By enjoining us to lead exemplary lives that register the ways we are different from others he claims we will be maintaining a relation to the mystery of death. Whether this successfully avoids nonsense can be seen in whether such an authentic way of living seems possible to others as an intelligible way of negotiating our relation to our own death. This highlights that intelligibility is an intersubjective affair (more about this below) and when we express ourselves about mystery it is necessary to see if others can also find sense in our expressions. Logic by itself cannot ensure that we stay this side of intelligibility. Finding mutual recognition of sense is necessary for finding how to go on in our inquiries. It is necessary next to look at the potential benefits of maintaining an awareness of the feeling of disorientation to help advance our inquiries. As we have found that it can help us negotiate the nonsensical, we can now ask whether there is a further role for the emotion in guiding our inquiries.

4   The Benefits of Bewilderment

One result of recognizing our feelings of disorientation is that if we find that mystery is an ever-present possibility this will inflect our enquiry with the idea that it is only ever provisional. This in itself can enable us to maintain an openness to other positions and allow us the space to be ever ready to start anew on an enquiry even given a lot of time and energy spent following one particular branch of reasoning or empirical study. It will also dispel the illusion that we have a bedrock of certainties to rely on and this agnosticism in relation to inquiring will have the effect of inducing an epistemological humility in us. This virtue is essential to making us responsive to reality. Iris Murdoch claims,

> We are not isolated free choosers, monarchs of all we survey, but benighted creatures sunk in a reality whose nature we are constantly and overwhelmingly tempted to deform by fantasy. (Murdoch 1997, p. 293)

If it is the case that we tend to deform reality through fantasy, suggested by our inability to tell nonsense from sense as Wittgenstein claims (Wittgenstein 1963, §38), we would do well to try to practice the virtue of humility in our inquiries and developing a sensitivity to disorientation would help us with this. If enquiry starts when we admit to ourselves that we are lost, this shows that disorientation cannot be separated from our intellectual projects. It may be argued that perhaps disorientation motivates the need for an enquiry at the beginning, but we can leave it behind as we are progressing and it would just serve to slow down or disrupt an enquiry that has already started. In response to this we could say that an awareness of and a regular return to the motivating emotion of disorientation will enable us to renew and concretely assess the progress of an enquiry as well as helping us ascertain if we have been led into nonsense. Humility and motivation can both be seen as fruits of developing an awareness of disorientation in our cognitive projects.

Another result of disorientation is to also allow us to see that we are dependent on others for sense-making activities, as we do not have all the cognitive resources to carry out such a project on our own and sense making inherently involves the interests and capabilities of others. Disorientation shows us that enquiry is constitutively an intersubjective affair, akin to a game where it is necessary that others make moves along side us to keep the enquiry going. Disorientation highlights where we are trying to make a move that will prevent us making sense to another person and so points up where and when we need to find a recalibration with the responses of our fellow inquirers in order to go on.[4] This is given an explanation by Cavell when he says,

> [T]he degree to which you talk of things, and talk in ways, that hold no interest for you, or listen to what you cannot imagine that talker's caring about…is the degree to which you consign yourself to nonsensicality…. (Cavell 1979, p. 95)

---

[4]The disorientation we feel can be understood as a contagion that starts with the person making the communication and infects others through the communication so that the sense of the communication is lost. For more on the contagion of emotions, see Candiotto (2015).

Shared interest gives us the coordinates for making sense of something collectively. In looking for and finding where our co-inquirers are situated evaluatively in relation to our own position and interest we learn something about how to remain intelligible to others and ourselves. From Cavell it can be seen that a major issue is whether something matters to us and if it does and also matters to the other person we will find ways to overcome nonsense. If nonsense persists it merely shows that the issue does not matter enough to any of the inquirers involved. It is through our interests and care for a subject matter that we find the resources to deal with mystery.

It also becomes clear through our disorientation in relation to death, as we have seen, that the inheritance of sense-making projects from others is an important part of enquiry and this is highlighted if we think about religious ways of understanding death. Religions themselves can be understood as sense-making projects that have been handed down through the generations in order for us to have something to say to each other about mystery and ameliorate the loss of coordinates that is found in the face of mystery. Dealing with our death is one way that religion can help us deal with disorientation but there is also the mystery of our being born. The fact of our birth is dealt with in the way that we pass cognitive projects on to further others and this is brought out clearly when we examine the aporias of education given an airing in Plato's discourse on *anamnesis* (recollection) as the solution to the puzzle of how it is possible to learn something one did not already know (*Meno 81d*). Religion and stories such as Plato's account of the soul although perhaps contentiously related to our modern cognitive inquiries provide us with ways of giving sense to resistant parts of life in the face of an abyss of mystery that is found even when following through a strictly scientific investigation.

Of course scientific knowledge, as well, is handed down from generation to generation and is not simply generated in the mind of one particular person. This very process of inheriting sense-making projects from the past highlights the fragility of the project. The project can be seen to be fragile in that projects must be handed down through intense years of schooling and training in the techniques of such projects and the decline of the modern world could lead to the loss of these projects

as has happened in the past when civilization has encountered so-called dark ages. To put it another way, cognitive enquiry is a socially realized practice performed over what might be understood as an abyss of mystery and our sense of disorientation at points helps us to remember this and keeps us aware of the value and vulnerability of the various interlocking projects that make up our intellectual culture. This understanding of enquiry inflected constantly by disorientation should foreground the fact that we are stewards of knowledge with all the concomitant duties this assigns to us. We are reminded that we are flawed knowers, that is, not self-sufficient, always dependent on others and also that we tend to deform reality as Murdoch insists and seems to be confirmed through our experience. Our knowing is limited through, not only our flaws, but also the conditions that limit what is possible for us to know. These conditions are the basis for our ability to inquire and are also the subject of our enquiry, in that an enquiry tests the boundaries of what it is possible for us to know. This leads us to the idea that any and all inquiries search to push back the frontier of mystery and so accordingly disorientation must be central to all our investigations in searching out where mystery lies.

5   Maintaining a Relation to Mystery

As shown above disorientation is the basis for a deepening of an understanding of our limits as human beings. This idea of limits corresponds to such facts as our mortality, our vulnerability, our dependence on others and our failures. Living a life where these facts are salient leads to a sense of humility, which can be developed to allow a cognitive appraisal of the equal value of others and a sense of our dependence on others and tradition for our ongoing cognitive interests. The value of this emotion to cognition is that it allows us to pierce through the fantasies that we normally live in, as Murdoch described our situation earlier, by bringing back to us forcefully our limits and in allowing us to starkly apprehend the reality of those limits and judge our abilities in the face of these limits. By allowing us access to reality the cognitive benefits

of disorientation imply that it is an emotion that can be thought of as central to cognitive inquiries in the sciences and the humanities. It is necessary to the performance of accurate inquiries to remain sensitive to disorientation and incorporate the virtues gained from it in ongoing investigations. Moreover, there are also specific implications for particular domains of enquiry, three of which I shall address here in brief as a sketch of further research areas that could be taken in future.

Firstly, disorientation can be seen as central in carrying out a philosophical project where a sense of mystery is essential to developing and defending a position. The acknowledgment that any position is liable to refutation allows the recognition, particularly pertinent in philosophy, that enquiry is a matter of sense making over an abyss of nonsense. This recognition enables practitioners of philosophy to affirm the tentativeness of their endeavor but also the importance of the attempt to overcome the tendency toward speaking nonsense. The project of philosophy can be seen as one that is regulated by using feelings of disorientation to signal when we come across a disguised piece of nonsense. Keeping open to a sense of disorientation allows the practitioner to find ways to communicate intelligibly. Although there are a variety of motives involved in philosophizing, one particular aim must be to express oneself coherently about very general matters and this relies on coming up with strategies to avoid nonsense. We looked at one particular strategy of Heidegger's mentioned earlier in the chapter, that of living a life in its full particularity as a way of acknowledging the mystery of death. Other strategies include skepticism i.e. maintaining a relationship to the idea that we can never really know anything, or the Socratic strategy of avoiding holding a position but rather subjecting the opinion of others to scrutiny. If we look at the Wittgenstein's method, remembering that he was very sensitive to our penchant to talk nonsense, we can see a strategy of coming up with thought experiments to test when our sense of disorientation kicks in and using this sense to tell us when our imaginary scenarios have led us to emptiness. His solution to nonsense was to see what we do with words in unproblematic everyday contexts and then to look at how our projections into the context of mystery lead us to say things without meaning. As he remarks,

> Where does our investigation get its importance from, since it seems only to destroy everything interesting, that is, all that is great and important?…What we are destroying is nothing but houses of cards and we are clearing the ground of language on which they stand. (Wittgenstein 1963, §118)

We can take from this that the ever-present temptation to talk nonsense ('…everything that is interesting…') must be overcome lest we eventually lose all coordinates for how to go on in our investigations. Philosophy, being the most general of disciplines, is undoubtedly the most prone to this problem being closest to mystery and so a sensitivity to disorientation should be the most important attribute of a philosopher. It is this sensitivity that allows her to arrogate a voice in pronouncing on matters associated with the whole gamut of academic disciplines and it is this ability that is most uniquely hers.

Secondly, moving on to education in general, it is necessary to help students develop a sense of the mysteriousness of the discipline they have set out to study as the bedrock of encouraging their own inquiries as they mature. In philosophical education this tends to be a commonplace given that most positions seem to be open to refutation and this leads the student to doubt that any position can be seriously defended from all criticism giving the sense that mystery is ubiquitous. However, this can lead to the feeling that philosophy is a waste of time if the value of this encounter with mystery is not underlined. For other subjects a certain stress can be put on the fragility of the results achieved in the discipline in a similar way to the philosophy seminar possibly through looking at more general question the discipline might lead to, but also through a genealogical approach to the discipline's main theories and central hypotheses. Developing and maintaining the mysteriousness of a discipline for the student is essential to the inheritance of the practices of the discipline in giving the student the sense that they are a steward in an ongoing project and require humility to carry out this stewardship as outlined above.[5] Such a sense of stewardship can be built on

---

[5]For more discussion about the humility necessary in learning both on the side of student and educator, see Earnshaw (2014).

the wonder that many children seem to feel about the world and this natural wonder should be attended to and shared in. Disorientation, although seemingly more negative an emotion than wonder can be seen to be closely related in that wonder is the experience of not knowing how to categorize something and therefore being similar, in the sense that there is the feeling of a lack of coordinates or lostness. It could be argued that disorientation is a degrading of natural wonder, that is, it is wonder without the affirmation of the experience. However, disorientation could be understood to be more primordial than wonder in that it puts into question not only the experience of the world but the value of that experience and so allows evaluation itself to be questioned. As well as the child's sense of wonder there can also be found, in the course of an elementary education, certain moments of bewilderment and imparting to the child that these moments to are important in exploring the world is highly pertinent to giving them a sense that it is worthwhile to continue to try to sense-making activities in the face the mystery of the world. Story-telling is a paradigm example, open to children, of the activity of sense making in the face of mystery and should be introduced as a way of dealing with mystery as children encounter it.

Finally, I will now look at a particular discipline where the development of a sensitivity to disorientation is particularly relevant, namely, in understanding mental health conditions. In the study of psychopathology, an acknowledgment of disorientation can help with developing a fellow feeling and an empathy with people suffering from mental health problems.[6] This is accomplished by recognizing that they are not living outside the gamut of human experience but rather in an existential situation where the inherent mystery of the world presents itself relentlessly. Feelings of disorientation in people unaffected by mental health problems give a window into what the world can be like for sufferers. Bringing the emotion of bewilderment to the fore in discussions of

---

[6]Memoirs of the mental health problems can help reduce the stigma of these conditions through allowing an insight into the life experiences of those suffering from a severe mental illness alongside an interpretation of the experience. For one example of this, see Greene-McCreight ([2006] 2015). It should be noted that many of the memoirs explicitly highlight the disorientation felt in such conditions.

mental health could be an important contribution to combatting the stigma of such conditions by showing how the erratic behavior of people suffering from such conditions may be understood through an emotion that we all experience at times. To give a characterization of how disorientation functions in terms of such conditions as schizophrenia and bipolar disorder it can be said that in these conditions there can be found a continual surprise and bewilderment elicited by the facts of the world. This phenomenological description can be understood given that disorientation is a possibility when confronted with any seeming fact about the world and because the world as such is inherently pregnant with mystery. Looking further at depression and anxiety, these conditions can be interpreted as expressing the darker side of mystery where estrangement from familiar practices through disorientation leads to a disengagement from the world or to a world in which there is felt a constant sense of threat. To turn next to traumatic experiences that elicit disorientation, we can even see the tantalizing prospect of existential benefits proceeding from disorientation as shown in the research on post-traumatic growth. Importantly research needs to be done on the relation between post-traumatic growth and the empirical benefits of disorientation on cognition related to academic investigations and research. As there is such a lack of empirical psychological research on the relation between cognition and disorientation we can only extrapolate from what has been suggested in this essay namely that the epistemic virtues that grow out of disorientation, namely, humility and a sensitivity to mystery through being able to identify nonsense could be the basis for substantially improving the cognitive functions of inquirers. This, in turn, could be a partial basis for the findings that lives can be significantly improved after a traumatic event, given that improved cognitive performance may enhance life decisions. All this suggests that the emotion of disorientation is central to our inquiries in psychopathology and even in the parallel study of improving well-being and that this emotion has been unfairly overlooked by the literature on mental health and well-being.

6   Conclusion

To sum up, the emotion of bewilderment or disorientation can be seen as motivating cognitive enquiries as a form of collective sense making against the backdrop of an abyss of mystery. This emotion regulates enquiries by allowing an awareness of when nonsense threatens to encroach on particular lines of enquiry as is found in the phenomena of contradictions, aporias and fantasies that detach us from a relation to the real. Disorientation also highlights the communal nature of enquiry as the stewardship of knowledge by exposing us to our vulnerability to nonsense that can be ameliorated by the help of others bridging generations. Finally, disorientation helps us develop a particular epistemic virtue in having a sense of humility in our inquiries. For these reasons disorientation should be given further attention as an emotion that is the basis for the possibility of intellectual enquiry and is highly relevant to philosophical, educational and psychopathological research in particular.

References

Candiotto, L. 2015. Aporetic State and Extended Emotions: The Shameful Recognition of Contradictions in the Socratic Elenchus. *Ethics & Politics* XVII (2): 219–224.

Cavell, S. 1979. *The Claim of Reason: Wittgenstein, Skepticism, Morality and Tragedy*. Oxford: Oxford University Press.

Cooper, D.E. 2017. *Senses of Mystery: Engaging with Nature and the Meaning of Life*. London: Routledge.

Earnshaw, O. 2014. Learning to Be a Child: A Conceptual Analysis of Youth Empowerment. *Educational & Child Psychology* 31 (1): 13–21.

———. 2018. Mood, Delusions and Poetry: Emotional 'Wording of the' in Psychosis, Philosophy and the Everyday. *Philosophia* 45 (4): 1697–1708.

Greene-McCreight, K. [2006] 2015. *Darkness Is My Only Companion: A Christian Response to Mental Illness*, 2nd ed. Grand Rapids, MI: Brazos Press.

Harbin, A. 2016. *Disorientation and the Moral Life*. Oxford: Oxford University Press.

Heidegger, M. 1962. *Being and Time*, trans. J. Macquarrie and E. Robinson. Oxford: Blackwell.

Mulhall, S. 2012. Wonder, Perplexity, Sublimity: Philosophy as the Self-Overcoming of Self-Exile in Heidegger and Wittgenstein. In *Practices of Wonder: Cross-Disciplinary Perspectives*, ed. S. Vasalou. Eugene, OR: Pickwick Publications.

Murdoch, I. 1997. *Existentialists and Mystics*. London: Penguin Books.

Otto, R. 1950. *The Idea of the Holy: An Inquiry into the Non-rational Factor in the Idea of the Divine and Its Relation to the Rational*. Oxford: Oxford University Press.

Plato. 2002. *Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo*, 2nd ed., trans. G. M. A. Grube. Indianapolis: Hackett Publishing.

Ratcliffe, M. 2012. Phenomenology as a Form of Empathy. *Inquiry* 55 (5): 473–495.

Tang, V.Y.H., A.M. Lee, C.L.W. Chan, P.P.Y. Leung, J.S.T. Sham, J.W.C. Ho, and J.Y.Y. Cheng. 2007. Disorientation and Reconstruction. *Journal of Psychosocial Oncology* 25 (2): 77–102.

Tedeschi, R.G., and L.G. Calhoun. 2004. Posttraumatic Growth: Conceptual Foundations and Empirical Evidence. *Pyschological Inquiry* 15 (1): 1–18.

Wittgenstein, L. 1963. *Philosophical Investigations*, trans. G. E. M. Anscombe. Oxford: Blackwell.

Part V

The Epistemic Value of Negative Emotions and Suffering

9

Learning from Adversity: Suffering and Wisdom

Michael S. Brady

It is commonplace, in philosophy and in everyday life, to think that suffering, understood as a kind of negative affective experience, is bad. What else might we hold, traditional thinking suggests, about pain, misery, despair, disappointment, grief, hunger, fatigue, anxiety, and shame? Nevertheless, the case can be made that suffering, in certain instances and circumstances, has considerable value. Indeed, it seems plausible that we would be considerably worse off if we *didn't* experience things like pain and remorse, hunger and shame. Those who are insensitive to pain don't live very long, after all. And those who are incapable of feeling negative emotions such as guilt and shame will find it very difficult to form and function in social relationships that are central to a happy life. By the same token, we think that we ought to experience other negative emotions: we ought to grieve when a loved one dies, and that despair is appropriate when our dreams are shattered. Reflection on

M. S. Brady (✉)
University of Glasgow, Glasgow, Scotland, UK
e-mail: michael.brady@glasgow.ac.uk

these cases might incline us to the view that suffering is not always and everywhere bad, but can in many instances be good.

In this paper I want to extend this positive perspective on the value of suffering, by examining the idea that suffering is necessary for *wisdom*. This idea has precedents. It is, for instance, a tenet of Buddhist thinking, expressed in the legends about Gautama Buddha. These explain the ways in which the experience of suffering is necessary for moral development and enlightenment.[1] But experiencing and transcending suffering is also vital for attaining wisdom, since such things are necessary if we are to follow the Noble Eightfold Path and thereby achieve *nirvana*. John Bowker writes: "The Eightfold Path is … right view (understanding), right thought, right speech, right action, right livelihood, right effort, right mindfulness, right concentration. Together those eight factors form the three foundations of Buddhist life: wisdom, ethical conduct, and mental discipline. Right view and thought constitute wisdom; right speech, right action and right livelihood constitute ethical conduct; right effort, right mindfulness and right concentration constitute mental discipline" (Bowker 1970, p. 255). The idea is also prominent in Greek thought. Thus Aeschylus writes in the "Hymn of Zeus" in the *Agamemnon*: "Zeus, who guided men to think, who has laid it down that wisdom comes alone through suffering" (Aeschylus *Ag.* 176–178, in Lattimore 1953, pp. 39–40). And the idea that suffering promotes understanding and wisdom is also central to Nietzsche's thinking. He writes, in typical style: "The discipline of suffering, of great suffering – do you not know that only this discipline has created all enhancements of man so far? That tension of the soul in unhappiness which cultivates its strength, its shudders face to face with great ruin, its inventiveness and courage in enduring, persevering, interpreting, and exploiting suffering, and whatever has been granted to it of profundity, secret, mask, spirit, cunning, greatness – was it not granted to it through suffering, through the discipline of great suffering?" (Nietzsche [1886] 1992, p. 225).

---

[1]These biographical details, sometimes referred to as the 'legend' of the 'historical Buddha', and presented by the poet Ashvaghosha in the *Buddhacarita* or *The Acts of the Buddha*, relate a number of occasions where Gautama Buddha's suffering has significant moral effects. I've discussed these details, and the general line in this chapter, in more detail in Brady (2018).

Rather than examining these historical resources, I want instead to investigate the link between suffering and wisdom by focusing on lines of thought in contemporary philosophy and psychology, and in particular on the idea that facing and overcoming adversity is positively correlated with growth in wisdom. Section 1 I'll say a little about what wisdom is, and then proceed to explain why negative affect and adversity are necessary for its development.

1   The Nature of Wisdom

Reflecting on and writing about wisdom is difficult. One reason for this is that there seem to be different kinds of wisdom, and so wisdom is not a unitary concept. Aristotle distinguishes *sophia*, or theoretical wisdom, and *phronesis*, or practical wisdom.[2] Some psychologists have differentiated "personal wisdom", which involves insights about the wise person's own life, and "general wisdom", where insights are about the lives of others, or about important aspects of the world as such (Staudinger and Glück 2011). Often wisdom is taken to have a spiritual or transcendental aspect. Those who take this line are more likely to identify figures such as Jesus or the Buddha as (perfect) examplars of wisdom. There are others who emphasize the critical or advisory aspects of wisdom, and who might be inclined to list Socrates or Shakespeare's Polonious as ideals (Ferrari and Weststrate 2013). And there are others still who maintain a more down-to-earth or folk conception of wisdom: here the model of the wise person is more likely to be the wise grandmother or sagacious teacher. Given these different perspectives and exemplars, it is unlikely that we can identify any neat list of necessary and sufficient conditions for wisdom. Nevertheless, both folk and expert ideas of wisdom suggest certain components that seem to be present in paradigmatic cases of wisdom. I propose to outline these, and then proceed

---

[2]For an excellent account of Aristotle on *phronesis*, see Russell (2009, Ch. 1). Many of the paradigmatic elements of wisdom to be discussed below resemble the parts or capacities that Russell thinks are components of *phronesis* on Aristotle's view.

to argue that suffering plays a necessary role in the development and exercise of these paradigmatic elements. In particular, I claim that suffering is central to the kind of *reflection* that philosophical thinking, and empirical evidence, suggests is at the heart of wisdom.

(**i**) *Understanding.* The wise person is thought to possess deep understanding of important aspects of herself and her world; she has particular insight into, and a grasp of, such things. (Staudinger and Glück 2011; Baehr, forthcoming). Of particular importance here is general knowledge or understanding of *value*, and of universal human problems and central spheres of experience (Baehr, forthcoming; Walsh 2015).

(**ii**) *Decision-making.* Wisdom is thought to involve the skill and ability to make good choices or decisions about how to respond to one's existential situation, an idea that is central to Aristotle's account of *phronesis* (Russell 2009). Wisdom requires excellence in making important decisions, solving difficult problems, and—as a corollary—giving advice to others as to how they should respond to their situation (Weststrate and Glück 2017; Tiberius and Swartwood 2011). So the capacity to make good decisions and give good advice, based upon one's deep understanding of and insight into what is valuable, is essential to wisdom.

(**iii**) *Experience.* The understanding and ability to make good choices that are characteristic of the wise person are grounded in lived experience. As Jason Baehr writes: "For this reason, we are reluctant to consider young persons wise. Rather, we think of wisdom as something that tends to be acquired over the course of a lifetime" (Baehr, forthcoming).[3] The relation between wisdom and (old) age captures the idea that wisdom requires a whole range of experience, some positive but (especially) some negative.

---

[3]See also Staudinger and Glück (2011) and Westrate and Glück (2017).

(iv) *Reflectiveness.* The wise person is held to be suitably deliberative or reflective about her experiences and existential challenges, an aspect of wisdom that is also central to Aristotle's account of *phronesis*. The wise person isn't someone who is inclined to make quick or snap judgements; rather, the wise person thinks and reflects deeply upon the world and themselves (Staudinger and Glück 2011). Wisdom is therefore associated with reflection and reflectiveness, both in psychological models and folk accounts (Weststrate and Glück 2017).[4]

(v) *Humility.* It is widely agreed that the wise person is not boastful or self-aggrandizing, but is instead humble and possessed of humility. Wise people know their own limitations, both epistemic and practical, and acknowledging this plays a role in her decision-making and advice-giving (Bortolotti 2011; Tiberius 2016).

(vi) *Compassion.* There is also widespread agreement that the wise person is not self-centred, but instead cares for and is concerned about the well-being of others. The wise person transcends or overcomes their own interests, and is sensitive, generous, and helpful (Staudinger and Glück 2011).[5]

Although these are distinct elements, they are—in wisdom—closely related to each other. As noted, experience of value is necessary if we are to understand and gain insights into its nature. In addition, such understanding is arguably essential if we are to choose well and advise others to choose well. An understanding of the relative value of one's plans and projects and achievements would seem to be a central part of humility. Further, some degree of care or compassion and selflessness is required if one is to be in the business of giving good advice, and so these elements of wisdom are also arguably connected.

However, the component of wisdom that ties all of these together would seem to be *reflectiveness*, and I hope to explain in the rest of this

---

[4]For the idea that reflection is necessary for wisdom, see also Webster (2003, p. 14) and Ardelt (2003, p. 278).

[5]See also Meeks and Jeste (2009) and Weststrate and Glück (2017, p. 804).

paper. It seems to me that reflectiveness is at the heart of wisdom, since that there are two important connections between suffering and reflectiveness, connections which will illuminate the centrality of suffering to the cultivation and exercise of wisdom. The first is that suffering motivates the kind of reflection that generates understanding of value. The second is that reflection *on* negative experiences is positively correlated with the growth of wisdom. In the following two sections I'll consider these ideas in turn.

2   The Importance of Suffering for Understanding

Why think that suffering is important for understanding of value? We can begin by noting a number of ways in which suffering is important from the standpoint of our epistemic goals or aims.[6] One way in which suffering has worth here is through being necessary for the appreciation of particular values. Consider, to illustrate, those values which contrast with suffering. The thought here is that we only really understand or grasp the goodness of a warm house and a full stomach if we have been cold, homeless, and hungry; we only fully appreciate love and companionship after we have experienced heartbreak and loneliness. The claim that suffering is important for such appreciation and understanding can itself be understood in two ways. On one reading, the claim is that the experience of certain pleasures *requires* the experience of suffering, since the pleasure itself can be defined simply in terms of *relief from* suffering. Think, for instance, of the relief one gets when one no longer suffers toothache, or when a neighbour's loud music finally stops, or when one scratches that really irritating itch. As Siri Leknes and Brock Bastian note, "The contrast afforded by pain and other aversive experiences is closely associated with the subjective experience of relief" (Leknes and Bastian 2014, p. 65). And "pain affords an effective contrast to many

---

[6]Cf. Davies (2012). He writes: "Schopenhauer said that any increase in our awareness is paid for with suffering", p. 128.

non-painful experiences, which can appear relatively pleasant or rewarding if they occur after the pain has ended. For instance, fruit flies approached odours associated with pain offset even when these odours were initially mildly aversive … Similar effects have been observed in rats … and humans" (ibid.). On another reading, the claim is that a prior experience of suffering *intensifies* our experience of some value, so that the latter is experienced as more pleasant as a result. The pleasure of cold beer is heightened and intensified if one is thirsty and overheated on a blistering day. Leknes and Bastian again: "Although introspection suggests to many people that relief and pleasure are easily dissociated, most primary rewards are intensified by relief. Food and drink taste better when providing relief from hunger or thirst … And where would the pleasure in going to bed at night be if we were not so tired, our muscles weary and aching?" (ibid.).

There is another way in which experience of suffering is epistemically important; for it seems that some forms of suffering are essential for our access to certain values, and hence essential for our understanding of these. Mark Johnston, for instance, argues that negative feeling or affect is the way in which certain negative values are *disclosed* to us (Johnston 2001). Without "affective engagement", he thinks that we would be blind to or ignorant of the relevant values. Johnston makes his case for the necessity of affective engagement by focusing on a particular class of values, which include, on the positive side, "the beautiful, the charming, [and] the erotic", and on the negative "the banal … the horrific and the plain old … repellent" (Johnston 2001, p. 182). Johnston thinks that "[i]f one has never been moved or *affected* by the determinate ways in which things are beautiful or charming or erotic or banal or sublime or appealing, then one is ignorant of the relevant determinate values" (Johnston 2001, p. 183). If this is right, then suffering is necessary for knowledge of a particular class of negative values. Without suffering, our knowledge of the world of values would be impoverished.

A third, and perhaps the most important, contribution that suffering can make to our epistemic lives is that it facilitates our *understanding* of a wide range of negative values, such that without suffering our understanding of value would be impoverished. I have elsewhere explained how emotions in general enable us to understand a range of values, by

focusing our attention onto evaluative situations, and motivating the search for reasons that bear on whether things are as they are emotionally presented as being (Brady 2013). I'll now give a brief overview of this account, tailored to highlight the specific epistemic importance of forms of suffering. So let us start with an obvious truth: there are very close links between suffering and *attention*. When I grieve I'm focused on and attentive to my loss, at the expense of other possible objects of attention; when I'm disappointed my attention is locked onto the disappointing event. Indeed, emotional effects on the focus of attention are particularly pronounced with negative emotion, with considerable evidence indicating that "negative" emotion focuses attention more narrowly on the relevant object or event than positive emotion, facilitating a "local" rather than a "global" appraisal of that object or event.[7] What is true of negative emotion is equally true of other forms of suffering; we saw earlier that physical pain makes salient potential bodily damage.

Suffering is not simply a source of salience, however; emotions such as fear and shame do not just direct and focus attention. In addition, these and other negative emotions tend to *capture* and *consume* attention. To say that attention is captured and consumed by emotional objects and events is to say that such objects and events hold sway over us, often making it difficult for us to disengage our attention and shift focus elsewhere. So fear and jealousy and guilt and disappointment stay with us; they are not simply short-term reflexive interruptions to our mental life, but often persist and dominate that life so that we remain focused on and attentive to danger, infidelity, wrongdoing, and frustrated dreams.

In my view, one of the important things that attentional persistence can do is to motivate reflection on the relevant objects and events, which itself has two important outcomes. First, such reflection enables us to discover reasons which bear on the accuracy of our initial emotional appraisals; second, through reflecting we determine

---

[7]The many studies that support this line. See, for instance, Wells and Matthews (1994), Isen (2000), and Fredrickson and Branigan (2005).

an appropriate behavioural response to the object or event.[8] In other words, the persistence of attention in experiences of suffering can motivate the search for and discovery of reasons, and in so doing can facilitate a judgement as to whether emotional appearance matches evaluative reality. Second, attentional persistence motivates reflection upon and a decision about the best way to respond to our existential situation.[9] So emotional persistence can enable the subject to determine whether what appears to be dangerous or shameful really is dangerous or shameful, through keeping attention fixed on these questions and promoting critical reflection about them; and it (thereby) facilitates a decision on how best to deal with the danger or shamefulness.

The idea that suffering motivates the search for reasons and coping strategies is well-supported by reflection on our own experience, by philosophical theorizing, and by empirical evidence. Consider first phenomenological evidence: we often *feel the need* to discover reasons and evidence when suffering. Think, for instance, of the person suffering great disappointment who feels motivated to discover whether she should be as disappointed as this, or whether indeed her response is unwarranted. Or think of the terrified person who feels a strong need to seek evidence that could confirm (or hopefully, disconfirm) his appraisal that he is in danger. This is not only the point that when emotional we feel the need to discover reasons of either kind; it also seems true that when we are no longer emotional we usually *lack* the motivation to assess the accuracy of our initial emotional appraisals or think about ways of dealing with our situation. If, for instance, I no longer feel afraid, then it is unlikely that I'll bother myself much with seeking evidence as to whether or not I'm in danger, and as to how I might therefore cope. Here again—as with the cases discussed in the previous chapters—it is negative affect that is central to the motivational story

---

[8]See Ronald de Sousa: "[P]aying attention to certain things is a *source* of reasons" (de Sousa 1987, p. 196).

[9]See Clark and DavidWatson, who write that "triggered by environmental events, emotions act as salient internal stimuli that alert the organism to *the need for further information gathering and action*" (Clark and Watson 1994, p. 131; My italics). On my view, this need is best served through the consumption of attention in negative emotional experience.

we tell. If the prospect of danger or frustrated expectation didn't *feel bad*, then the chances are that I wouldn't be (as) bothered to deal with it, and would thereby fail to arrive at the appropriate understanding and awareness of the best coping strategy.

Phenomenological support for the idea that suffering facilitates reappraisal, through effects on attention, fits in nicely with views in psychology which suggest that appraisal and reappraisal is an ongoing process in emotional experience. Klaus Scherer, for instance, argues that "emotion *decouples* stimulus and response", allowing a "latency period between stimulus evaluation and reaction" (Scherer 1994, p. 128). On his view, "the first major function [of the latency period] is the ongoing analysis of the stimulus event, which allows the organism to arrive at a more detailed or more realistic conclusion and may lead to a re-evaluation and consequently a revision of the original appraisal" (Scherer 1994, p. 128).

The idea that suffering facilitates reappraisal through the capture of attention also finds philosophical support in the writings of Thomas Reid. Reid claims that "[i]t requires a strong degree of curiosity, or some more important passion, to give us that interest in an object which is necessary to our giving attention to it. And, without attention, we can form no true and stable judgement of any object" (Reid 1969, pp. 184–185). And: "[a]ttention may be given to any object, either of sense or of intellect, in order to form a distinct notion of it, or to discover its nature, its attributes, or its relations and so great is the effect of attention, that, without it, it is impossible to acquire or retain a distinct notion of any object of thought" (Reid 1969, pp. 76–77). So for Reid, emotion (or "passion") is necessary for us to pay attention to some object or event, and paying attention is necessary for us to form an accurate ("a true and stable") judgement about that object or event. This is as true for emotional suffering as it is for forms of emotional experience in general. But it also seems true of physical suffering: our pain keeps our attention focused on potential damage, both so that we get a better idea of the nature and extent of the damage, and an idea of how we might cope with it. If Reid is right, then without negative affect, we would lack the relevant forms of evaluative knowledge.

Finally, there is neuroscientific evidence that speaks in favour of the proposal. A central part of this evidence rests upon the idea that emotions involve increased *cortical arousal*, and that this is a central element in attentiveness. For as Joseph LeDoux has argued, increased cortical arousal is central to increased processing of information about emotional objects and events, and to retrieving memories about those objects and events. A result of this is an enhanced or more accurate representational picture (LeDoux 1996).

There is, therefore, a raft of evidence supporting the idea that suffering, both physical and emotional, can facilitate reappraisal of our situation, by motivating a search for reasons that bear on the appropriateness of our affective response, and a search for ways of dealing with the objects or events that trigger suffering. But if this is true, then there is strong support for the idea that suffering facilitates and is arguably necessary for our *understanding* of our evaluative situation. For awareness of the considerations that bear on whether, e.g., my situation really is dangerous, or my behaviour shameful, just *is* awareness of the factors or features that *make* my situation dangerous (or not), or shameful (or not). For instance, if upon waking in the dead of night I hear a noise downstairs, am consumed with fear, and search for reasons that bear on whether I am in fact in danger, then the discovery of such reasons *constitutes* my understanding of why I am in fact safe—because, for instance, I discover that the noise was just my drunk partner coming in from the pub, and so understand that the situation is harmless. Similar things apply for other cases of suffering: my guilt consumes my attention and motivates me to understand what I did wrong; my jealousy keeps my attention fixed and moves me to grasp the nature of the infidelity; my shame motivates reflection on what I did and facilitates my understanding of why my behaviour was shameful.

By the same token, the kind of reflection on our existential situation that suffering motivates is also, arguably, necessary in the development of the capacity for good decision-making that wisdom requires. For suffering, as we have seen, motivates reflection on our behavioural responses to our evaluative situation, so that we can assess how we can

best cope with the object or event that generated the experience in question. Once again, it is highly unlikely that creatures like us would develop the capacity for good decision-making—and the corresponding capacity to be good advisors—without affective engagement. For without affective experience, it is unlikely that we would regard the objects in question as mattering to us; and if they don't matter to us, why should we bother expending effort and energy in figuring out the best strategies to deal with them, both now and in the future?

It seems clear, therefore, that suffering is necessary for the kind of evaluative understanding and capacity for good decision-making that characterize wisdom. Suffering is necessary for both, because it focuses attention on our existential situation, and motivates theoretical and practical reflection about it. And without this kind of reflection, creatures like us would be ill-placed to decide correctly what to do, and to advise others that they should do, in important spheres of human experience.

3   The Importance of Reflection on Experiences of Suffering

Earlier we saw the claim that wisdom requires suitable amounts of experience, which is one reason why wisdom is associated with maturity. An important strand of recent theorizing about wisdom is that suffering and adversity—in the form of difficult or traumatic experiences—is particularly important for the cultivation of wisdom. Staudinger and Glück write: "traumatic life experiences can be conducive to the development of personal wisdom (e.g., Baltes et al. 1995), a notion prominent in concepts such as posttraumatic growth (e.g., Calhoun and Tedeschi 2006), stress-related growth (Aldwin and Levenson 2001; Park et al. 1996), or growth through adversity (e.g., Joseph and Linley 2006). After negative experiences such as accidents, life-threatening illness, or the death of a close other person, many people report self-perceived increases in aspects of personal growth such as compassion,

affect regulation, self-understanding, honesty and reliability, spirituality, and self-reported wisdom itself (cf. Park 2004, p. 232)".[10] Similar benefits are not reported from positive emotional experiences. By the same token, Jeffery Webster's 40-item *Self-Assessment Wisdom Scale* focuses on negative experiences as opposed to positive. Included in the "critical life experience" dimension of the questionnaire are claims such as "During my life I have already overcome many painful facts", "I have met a lot of the negative side of life", "I went through many difficult changes throughout life", and "I have gone through various moral dilemmas", again suggesting that (at least when it comes to self-assessment) it is negative experience that is associated with wisdom, rather than positive (Webster 2003).

One plausible answer of this discrepancy is that wisdom, as noted above, also involves *reflection on* one's experiences; and negative experiences generate much more in the way of beneficial reflection. The kind of beneficial reflection psychologists in particular focus on is a kind of *self-reflection*, in which a person tries to understand and make sense of her experiences, and fit them into a broader narrative of her life. Thus Dan McAdams writes, in a paper on the importance of narratives for our psychological well-being: "Negative events produce more cognitive activity in general and more efforts to engage in causal reasoning, compared to positive events. At the level of the life story, negative events seem to *demand an explanation*. They challenge the storyteller to make narrative sense of the bad thing that happened" (McAdams 2008, p. 254). In such cases, people try to make sense or meaning out of their experience of suffering. McAdams cites a number of further studies which show that "exploring negative life events in detail is associated with psychological maturity" (ibid.).

Further support for this line of thinking comes from recent work by Weststrate and Glück, who argue that reflection, in the form of "exploratory processing of difficult life experience", is a determinant of and

---

[10]The works cited are Baltes et al. (1995), Calhoun and Tedeschi (2006), Aldwin and Levenson (2001), Park et al. (1996), Joseph and Linley (2006), and Park (2004).

positively associated with wisdom. Their hypothesis is that "Through self-reflection, individuals reconstruct, analyse, and interpret real-life sequences of thought, emotion, and action for meaning. The life lessons and insights arrived at through self-reflection lead to an ever-deepening and more complex appreciation of life, which we might call wisdom" (Weststrate and Glück 2017, p. 802). The kind of reflection they have in mind here is not that which seeks to understand why objects and events have certain value. Instead, exploratory processing is an "investigative, analytical, and interpretive approach to self-reflection on life events, which emphasizes meaning-making (i.e., extracting lessons and insights), complexity, and growth from the past". Experimental evidence from across three studies confirmed their hypothesis, and they concluded that "wisdom was unrelated to the frequency of self-reflection, but positively related to exploratory processing of difficult life experience" (Weststrate and Glück 2017, p. 809). Importantly, they report that "these findings are limited to difficult life experiences, because pathway determinants may differ according to event valence. We should be careful not to infer that exploring the meaning of positive life events will lead to wisdom in the same way as it does with negative life events—perhaps it is best to savor, rather than scrutinize, positive life events" (Weststrate and Glück 2017, p. 810). Now Weststrate and Glück don't speculate on the reasons for why it is best not to scrutinize positive life experiences. But one suggestion is that reflecting on these might make us realize how fragile such experiences are, or how they came about merely as a result of good fortune—and reflection like this might very well undermine the positive experience entirely. Suppose we meet our beloved by happenstance—bumping into them in a crowded pub, or sitting next to them on a flight, etc. In such cases it could very easily have been the case that we didn't meet them, didn't fall in love, didn't get to enjoy the rest of our lives together. This kind of good fortune—that could very easily have passed us by—isn't something we like to reflect upon. What if I arrived at the pub 10 minutes later? What if I hadn't checked in online at just that time? Because of this, we might very well be disinclined to think about the meaning of positive experiences, on pain of their disappearing.

If the above is correct, then there is empirical support for the idea that a particular kind of reflection on experiences of suffering is important for the development of wisdom. Given the idea that experience and reflectiveness are core components of wisdom, and given the lack of evidence of the beneficial effects of reflecting on positive experiences, it is not implausible to assume that reflection on negative experience is necessary to the cultivation of wisdom, at least as this is understood both by the folk and by philosophers and psychologists. A second kind of suffering-related reflection will play a vital role, therefore, in the lives of those who are wise. This does not imply, of course, that suffering always have these effects, or, therefore, that experiencing suffering through adversity is *sufficient* for wisdom. For the kind of exploratory processing that leads to wisdom will require other factors to be in place, and other virtues. Weststrate and Glück also note that exploratory processing involves "effortful work of finding meaning in the difficult experience", and as a result, like wisdom itself, "is rare, probably because it is less pleasant than other processing modes" (Weststrate and Glück 2017, p. 810). It is plausible to assume that such processing requires forms of strength of character—such as fortitude, patience, and courage. By the same token, we should assume that we will only be in a position to grow wiser as a result of reflection on adversity and traumatic experiences if certain social conditions and structures are in place. Trauma will be difficult to even bear, let alone grow from, in the absence of family, friends, peers, and institutional structures to provide the physical and emotional support that will be needed for exploratory processing and sense-making. Still, if the above is correct, then negative experiences might not be sufficient to make us wise. But they are certainly necessary.

4   Conclusion

Suffering is necessary for wisdom, I have claimed, because (i) it captures and consumes attention, thus facilitating deliberation about emotional objects and events, and (ii) because it motivates the kind of exploratory processing that is positively correlated with growth in wisdom. If this

is correct, then the story I present here expresses an important way in which suffering has positive value for creatures like us. Moreover, if we think that wisdom is a kind of "executive virtue" that is essential for excellent choices and decision-making, then our capacity to choose and act well will depend upon our having a wide range of experiences of adversity and suffering. For the greater the range of our negative experiences, the greater the opportunities we have to reflect upon, and come to understand, our evaluative landscape and ourselves.

This has an interesting implication, however. For the idea that suffering and adversity are necessary to wisdom undermines the idea that wisdom is something that we should rationally seek or pursue—on the assumption that it would be highly irrational or imprudent to seek out adversity, trauma, and suffering in all of its myriad forms. Even if we stand to benefit from experiencing and overcoming adversity and suffering, therefore, this is not something that we should have as a rational goal or aim. This means that what is good for us—in the sense of making us more excellent or more virtuous—might diverge from what it is rational for us to pursue or to seek. We might therefore look up to and admire those who are wise, as a result of having experienced and grown from adversity and suffering. But we might not envy them, or wish we too had lived the kind of life that they had. To my mind, this raises important questions about the relation between wisdom, as the highest virtue, and practical reason, as an account of what it is rational for us to pursue. These questions must, however, be left until another occasion.

References

Aldwin, C.M., and M.R. Levenson. 2001. Stress, Coping, and Health at Mid-Life: A Developmental Perspective. In *The Handbook of Midlife Development*, ed. E.M. Lachman. New York: Wiley.

Ardelt, M. 2003. Empirical Assessment of a Three-Dimensional Wisdom Scale. *Research on Aging* 25: 275–324.

Ashvaghosha. (1959). *The Buddhist Scriptures*, trans. E. Conze. London: Penguin Books.

Baehr, J. Forthcoming. Wisdom in Perspective. *Journal of Value Inquiry.*

Baltes, P.B., U.M. Staudinger, A. Maercker, and J. Smith. 1995. People Nominated as Wise: A Comparative Study of Wisdom-Related Knowledge. *Psychology and Aging* 10: 155–166.

Bortolotti, L. 2011. Does Reflection Lead to Wise Choices? *Philosophical Explorations* 14 (3): 297–313.

Bowker, J. 1970. *Problems of Suffering in Religions of the World*. Cambridge, UK: Cambridge University Press.

Brady, M. 2013. *Emotional Insight: The Epistemic Role of Emotional Experience*. Oxford: Oxford University Press.

Brady, M. 2018. *Suffering and Virtue*. Oxford: Oxford University Press.

Calhoun, L.G., and R.G. Tedeschi. 2006. *Handbook of Posttraumatic Growth: Research and Practice*. Mahwah, NJ: Erlbaum.

Clark, L., and D. Watson. 1994. Distinguishing Functional from Dysfunctional Affective Responses. In *The Nature of Emotion*, ed. P. Ekman and R. Davidson. Oxford: Oxford University Press.

Davies, J. 2012. *The Importance of Suffering*. London: Routledge.

de Sousa, R. 1987. *The Rationality of Emotions*. Cambridge, MA: MIT Press.

Ferrari, M., and M. Weststrate. 2013. The Scientific Study of Personal Wisdom. In *The Scientific Study of Personal Wisdom*, ed. M. Ferrari and M. Weststrate. Dordrecht: Springer.

Fredrickson, B.L., and C. Branigan. 2005. Positive Emotions Broaden the Scope of Attention and Thought-Action Repertoires. *Cognition and Emotion* 19: 313–332.

Isen, A.M. 2000. Positive Affect and Decision Making. In *Handbook of Emotions*, ed. M. Lewis and J. Haviland-Jones, 417–435. New York: Guilford Press.

Johnston, M. 2001. The Authority of Affect. *Philosophy and Phenomenological Research* 63 (1): 181–214.

Joseph, S., and P.A. Linley. 2006. Growth Following Adversity: Theoretical Perspectives and Implications for Clinical Practice. *Clinical Psychology Review* 26: 1041–1053.

King, L.A., and J.A. Hick. 2006. Narrating the Self in the Past and Future: Implications for Maturity. *Research in Human Development* 3: 121–138.

Lattimore, R. 1953. *Aeschylus I: Oresteia*. Chicago: University of Chicago Press.

LeDoux, J. 1996. *The Emotional Brain*. New York: Simon & Schuster.

Leknes S., and B. Bastian. 2014. The Benefits of Pain. *Review of Philosophy and Psychology* 5: 57–70.

McAdams, D. 2008. Personal Narratives and the Life Story. In *The Handbook of Personality: Theory and Research*, vol. 3, ed. O.P. John, R.W. Robins, and L.A. Pervin. New York: Guilford Press.

Meeks, T., and D. Jeste. 2009. Neurobiology of Wisdom: A Literature Overview. *Archives of General Psychiatry* 66 (4): 355–365.

Nietzsche, F. [1886] 1992. Beyond Good and Evil. In *The Basic Writings of Nietzsche*, trans. and ed. W. Kaufman. New York: The Modern Library.

Park, C.L. 2004. The Notion of Growth Following Stressful Life Experiences: Problems and Prospects. *Psychological Inquiry* 15: 69–76.

Park, C.L., L. Cohen, and R. Murch. 1996. Assessment and Prediction of Stress-Related Growth. *Journal of Personality* 64: 71–105.

Reid, T. 1969. *Essays on the Active Powers of the Human Mind*, ed. B. Brody. Cambridge, MA: MIT Press.

Russell, D. 2009. *Practical Intelligence and the Virtues*. Oxford: Oxford University Press.

Scherer, K. 1994. Emotion Serves to Decouple Stimulus and Response. In *The Nature of Emotion*, ed. P. Ekman and R. Davidson. Oxford: Oxford University Press.

Staudinger, U., and J. Glück. 2011. Psychological Wisdom Research: Commonalities and Differences in a Growing Field. *Annual Review of Psychology* 62: 215–241.

Tiberius, V. 2016. Wisdom and Humility. *Annals of the New York Academic of Sciences* 1384: 113–116.

Tiberius, V., and J. Swartwood. 2011. Wisdom Revisited: A Case Study in Normative Theorising. *Philosophical Explorations* 14: 277–295.

Walsh, R. 2015. What Is Wisdom? Cross-Cultural and Cross-Disciplinary Syntheses. *Review of General Psychology* 19 (3): 278–293.

Webster, J.D. 2003. An Exploratory Analysis of a Self-Assessed Wisdom Scale. *Journal of Adult Development* 10: 13–22.

Wells, A., and G. Matthews. 1994. *Attention and Emotion: A Clinical Perspective*. Hillsdale, NJ: Erlbaum.

Weststrate, N., and J. Glück. 2017. Hard-Earned Wisdom: Explanatory Processing of Difficult Life Experience Is Positively Associated with Wisdom. *Developmental Psychology* 53 (4): 800–814.

10

The Grapes of Wrath and Scorn

Pascal Engel

for Patrizia Lombardo

1   Introduction

Can there be reasons for emotions, and can emotions be the basis for some kind of knowledge? I shall give a positive answer to these questions, but it will be limited in three respects. I shall deal only with two emotions, anger and contempt. I shall claim that the kind of knowledge they give us is a very specific instance of knowledge, moral knowledge. This entails that there can be such knowledge, a view which I cannot argue for, but which I shall presuppose. I shall not take my material from psychology and affective sciences, but from literature, and even more specifically, from Jonathan Swift. My justification for this strategy is that literary works can give us as much insights on emotions, through

P. Engel (✉)
École des hautes études en sciences sociales (EHESS), Paris, France
e-mail: pascal.engel@ehess.fr

ideal types, as experimental approaches.[1] The thesis which I want to defend is that the emotions of anger and contempt, as they are manifested in literary form in Swift's satires, express a certain kind of moral knowledge. Why, could one ask, should such an idiosyncratic expression of an emotion through literary form be informative? The answer lies in the specific nature of Swift's satire, which is very personal, without being autobiographical. As Claude Rawson notes, "The closeness of Swift's temperament to the things he attacked is a defining feature of his writing, and one of which he was edgily self-aware. He evokes it with a minutely inward participation" (Rawson 2014, p. 1). By focusing on Swift, I try to follow the lead of writers who take literature as a source of case studies for knowledge about mind and emotion. This knowledge, I shall argue, is not direct, but indirect, and mostly based on a certain view of virtues and vices.

2   Emotional Reasons and Justification

Whether or not one agrees that knowledge is justified true belief, knowledge at least involves having reasons for what one knows, whether or not we have access to these. Can emotions give us reasons and can they be reasons? The answer depends of course upon the theory of emotions that one holds. There must also be enough similarity between reasons for emotions and reasons for other familiar attitudes, such as belief and desire.

I shall rely on four assumptions which have been accepted by a number of writers on emotions. The first is that emotions are intentional states or episodes, which have a specific content which is most of time propositional, but which can also be directed at an object: one is afraid *that p* or *of someone* or *something*, happy that *p*, angry that *p* or *at something* or *at someone*. The second is that emotions are associated with typical bodily reactions, and have a specific phenomenology or feel, although the variations can be wide (sometimes it does not

---

[1]See e.g. Robinson (2005) and Lombardo (2014).

feel in a specific way to be afraid, and there are cold angers). A third assumption is that emotions involve a certain kind of appraisal, related to their valence: they are negative or positive. A fourth assumption is that emotions have a *formal object*, which is not the object which they have as token episodes (such as fear of this dog, anger at this person at this very moment), but the type of object at which their contents are directed. This idea is in general expressed in the following way: fear's formal object is what is *fearable* or *frightening*, anger's formal object is what is *worthy of irritation*. Each emotion has its own formal object. An emotion is correct when it fits its formal object. In this sense one can speak of emotional truth, when the emotion towards a certain content or object is appropriate to its formal object (Mulligan 2003; de Sousa 2011; Deonna and Teroni 2012). If we think of emotions as attitudes and dispositions rather than as specific episodes, we can ascribe to them properties which they share with epistemic attitudes such as beliefs, doubts or presumptions. Thus we can also say that emotions are justified when they fit their formal object. Emotions can have reasons. We say that we have reasons for being afraid or angry. They can also be reasons for epistemic attitudes. Thus my being surprised that Mary came may be a reason for my fear that Jane would not like it. Emotions can be factive: to be disgusted at something, or to be horrified at something seem to presuppose that this something exists (Gordon 1987).

   Does the fact that emotions involve these epistemic liaisons entail that they have a genuine epistemic role, such as being justifiers for epistemic attitudes? It is one thing to say that they can have cognitive role or value, and another to say that they can be justified, possibly true, or that they can yield knowledge. If emotions can have reasons or if they can be reasons, they must have the properties usually ascribed to reasons (Skorupski 2010). Moreover the reasons have to be epistemic. For this there must be a relation between an emotion as reason and a certain attitude of φ-ing. The emotion as reason has to be a relation to a fact. If the reason is epistemic, it has to be a matter of degree, and is most of the time relative to circumstances. If it can justify an attitude, it has to be sufficient for justification. Thus my reason to be angry at the rise of the price of fuel is the fact that the taxes for fuel went up. My anger will be justified if the raise of prices of fuel is a sufficient reason for being

angry. We can also say that it is correct because it is fitting to the circumstance. But what is it for an emotion to be fitting? Some angers are justified, some others are not. The price of fuel often rises up. But when does it become a reason for anger? When we are told that Achilles was angry at Agamemnon for the loss of his captive slave Briseis, we feel that the hero was angry for a bad reason, but when we are told that he was angry at Hector for the loss of his friend Patroclus, we feel that his anger was for a good reason. But where does the difference lie? Not only one can have an emotion like anger for good, bad, better or worse reasons, but it can be had for the wrong kind of reason. Suppose that a powerful demon threatens to torture me to death unless I do not become angry at someone who has been very nice and helpful to me. If I manage to obey his order, I will be angry for the wrong kind of reason. The attitude does not fit the object. In contrast being offended after having been insulted seems to be a right kind of reason for being angry. This distinction is indeed similar to the right/wrong kind of reasons distinction which has been raised for attitudes in general (see Rabinowicz and Rønnov-Rassmussen 2004; Parfit 2011: Appendix A; Skorupski 2010). So if there are to be reasons for emotions, there must be a right/wrong kind of reason problem.

The reason relation for emotions actually goes into both directions: What is the basis of our reasons for emotions? Let us call this basis *reasons from*. How can they be reasons for other states and attitudes? Let us call these *reasons for*. Both questions are epistemological. What is the kind of state which justifies an emotion? If we accept the view that emotions involve or can be grounds for, evaluative judgments, how can they justify these judgments? The four assumptions mentioned above suggest a parallel between the familiar problem of the justification of perceptual beliefs and the problem of the justification of emotions. Many writers have espoused a perceptual model of the justification of the value judgments associated to an emotion: just as our perceptual beliefs (say, my belief *that this is a tree*) are based on our perceptual experiences (*this looks like a tree*), our emotional experiences (say, my experiencing fear *of this dog*) are the base of our evaluative judgments (*this dog is dangerous*). The analogy with perception is all the more tempting that the cognitive base of the emotional experience is itself a perceptual experience

(my *seeing this big salivating dog*). So on this model emotional justi-
fication seems to involve a double link: an emotion is justified for an
agent just in case (i) the agent's beliefs about the object of the emotion
on which the emotion is based are (epistemically) justified, and (ii) the
emotion is an appropriate or fitting response to the situation as it is
experienced by the agent.

The trouble is that this model is much too simple. Is the relation
between the experiential base and the emotion an evidential relation? In
the case of belief, a reason is sufficient in so far as it is good enough to
justify some action or belief by itself; a sufficient reason makes the belief
permissible. But sufficient reasons do not make believing obligatory, nor
do they entail the correctness or the truth of some belief; they thus fall
short of being conclusive. So reasons can be defeasible, as the perceptual
model maintains, and still be sufficient; and reasons will be sufficient to
justify belief in conditions in which there are no defeaters. Many have
held that the reason or justifying relation is weaker than a sufficient rela-
tion, a mere *prima facie* kind of justification, or an entitlement relation.
This condition holds even more for emotions. My learning that the
price of fuel has gone up is a defeasible reason for my being angry and
for my belief that it is unjust that the price of fuel arises. And indeed it
is defeasible: there are many more things worthy of being angry about,
and in many cases the rise of the price of fuel is not worthy of anger.

As soon as one asks about the justification of emotions on this model
in analogy with the justification of perceptual beliefs, a lot of questions
arise. If emotions are supposed to be fit or correct because they are jus-
tified by the experiences on which they are based, is the justification
based on evidence for the evaluative judgment which is supposed to be
associated to it? The latter is clearly not inferred from the former. If, on
the contrary, we take the justification to be immediate, as if it sprung
directly from the emotional feel, how can it be transparent?[2] In other
words, can we move immediately from the content of the experience
(say *this is a big dog*) which produces an emotion (*fear of the dog*) to the

---

[2]Brogaard and Chudnoff (2016) and Echeverri (2017) argue against this alleged transparency of
the "dogmatist" view of emotional justification. Although they do not quote Pryor (2000) (as
Echeverri 2017 does), they clearly want to refer to an analogue of this view for emotions.

evaluative judgment (*this dog is dangerous*)? Or is the justification mediated by beliefs? Some writers have talked of emotionally laden beliefs.[3] But as soon as the cognitive basis of our emotional experiences becomes belief-laden, a dilemma looms: either it is the *beliefs* which accompany an emotion which are justified, in which case the emotion as a feeling is not in itself justified, or it is the emotion *as feeling* (as phenomenal feel) which is justified, in which case the justification is based only upon an affect, hence no justification at all. For an example of the first horn, take surprise. Surprise is an emotion mediated by beliefs—antecedent belief that not *p*, further discovery that *p*, surprise that *p*—but if only the beliefs have a justificatory power, the emotional import of the feeling of surprise does not play any epistemological role: the evaluation is purely cognitive. On the second horn, if we say that it is only the emotional part of surprise which is justified, how can the feeling itself (the startle, the eyebrows that raise, the pupil that dilate) have any epistemological import, and if there is such an import, how can it be transparent? If it is—that is if we can move from the perception to the evaluation of surprisingness (or dangerousness in the case of fear), the justificatory link between experience and belief in the emotional case is similar to the justificatory link in the perceptual case according to the so-called "dogmatist" view of perceptual justification. So, if your emotional response to a perceived object makes it seem to you that that object or event possesses some evaluative property, then you thereby have *prima facie*, immediate justification for believing that that object or event possesses that evaluative property. But if this were true, then fear, guilt and anger, absent defeating evidence, would immediately justify beliefs, rationalizing avoidance, self-condemnation, and retaliation. But we rarely take our emotions at face value, or if we do we should avoid doing so: they must also be related to certain beliefs and desires, which, together with the emotion and the value judgments, constitute our reasons to act. There are further disanalogies between the perceptual and the emotional case.[4]

---

[3]Stocker (1987) talks of "emotionally laden beliefs" which are ways the beliefs are taken. But what is justified? The beliefs, or their emotional charge?

[4]See in particular Brady (2013).

Now, from the fact that the perceptual model of emotional justification is hard to sustain literally at the level of the *reason-from*, does it follow that we should withdraw any hope of establishing a *reason-for* relation between emotional experience and the perception of value? No. Proponents of the perceptual model give examples such as being struck by the injustice of slavery through the emotion felt in reading *Uncle Tom's Cabin* or by the beauty of nature through the emotion of contemplating a sunset over the *Grand Teton National Park*, but the ways of the perception of values are most of the time more complex: reading *Huckleberry Finn* or contemplating the English countryside at daybreak on a fine summer day may give rise to more complex associations of emotional experience and value. The fact that we do not typically trust emotions as sufficient reasons to form evaluative judgments, but rather as reasons to look for non-emotional reasons to confirm our initial emotional appraisal does not show that this initial appraisal cannot justify the evaluative judgments, but just that the justificatory route is more complex. The relationship between emotional experience and evaluative beliefs need be neither direct nor foundational in the sense suggested by the simple perceptual model. It can be holistic, and such that the emotional experience and its relation to values is further confirmed by related beliefs. The fit between the perceived situation, the emotion and the issued value judgments may be more a matter of coherence than a matter of perceptual basis, and the correctness of emotions need not be based on some mysterious capacity of grasping the values within the emotional experiences. The relation between the emotion and the value need not be itself perceptual, but can be based on the idea that the correctness of the emotion is due to an appropriate response to a perceived event or action. Emotions in this sense are sources of reasons, and do not have direct epistemological impact.[5]

---

[5]De Sousa says that the role of emotions is often to attract our attention: "Paying attention to certain things is a source of reasons" (1987, p. 196). So the kind of skepticism about the perceptual model expressed by Brady (2013) need not entail the falsity of the correctness account of emotions. De Sousa (2011) suggests a more coherentist model. Pelser (2014) and Tappolet (2016) have proposed more sophisticated accounts of the perceptual model.

3   The Elusiveness of Anger and Contempt

Anger and contempt are both negative emotions which illustrate the features listed above almost paradigmatically. Aristotle defines anger as "an impulse, accompanied by pain, to a conspicuous revenge for a conspicuous slight directed without justification toward what concerns oneself or toward what concerns one's friends" (*Rhetorics*, 1378a31–1378b9). The feeling is of pain, but the formal object, what is anger-worthy,[6] is a complex relational content directed to someone for a reason, involving a judgment, to the effect that there is an injustice towards oneself or one's friends. The emotion is factive, in Gordon's (1987) sense: it is based on the knowledge that someone has done you some harm (say, insulting you), which—if something like the perceptual account is right—justifies one to judge that one has been offended by someone. The latter judgment is a moral one: some injustice has been done to you. How can the emotion be correct or fit its object? A certain standard, or norm, has been violated. The question immediately arises: how can the emotion be a reason for a judgment about an objective moral subject matter? For the feeling may be transitory and the judgments can change. Anger is most often directed not at a particular action or trait, but at the whole person. In Ariosto's epic *Orlando furioso*, the Christian knight Roland becomes furious when he discovers Angelica's love for the Saracen Medoro. But he could have be frenzy for another reason, and his judgment that an offense has been done to him might have been based on other moral standards that those of medieval chivalry. We can have distinct views of the objectivity of the standards, and one can be a non-cognitivist or a cognitivist about the nature of values, but at least the evaluative judgment to the effect than some injustice has been committed *purports* to be objective. This is enough for a contemporary reader to be able to understand Roland's wrath.

Contempt is, among the emotions, one of the least typical, for it seems to lack any characteristic feeling and bodily expression. It is a very

---

[6]There is no lexical item in English corresponding to what is the formal object of anger, in the ways the *admirable* is the formal object of admiration.

cognitive emotion, which seems to be mostly constituted by a certain kind of judgment appraising the status of its object, and the curling of the lips, joined and retracted in a derisive or sardonic smile cannot be its paradigmatic expressions. The judgment is such that it is directed at the whole person which is its object (if I have scorn for the way your dress, my scorn is directed at *you*, not at your particular clothing) and such that the person in question is taken to be inferior. Like anger, it is a "globalist" emotion (Bell 2013, p. 37) or a "characterizing attitude" (Darwall 2018). Perhaps we may call it, following Strawson (1962), a "reactive attitude", expressing a global set of emotions and feelings, which are constitutive of our status as persons and of the attitudes that we have to other persons. Strawson says that such attitudes can be either "objective"—when we distance ourselves from others and consider them as non-free, non-responsible and not liable to praise or blame—or "subjective", when we take others as free and liable to evaluative judgments. Unlike other emotions, contempt is a disposition, and rarely an episodic mental occurrence, and when it is associated to a feeling, its scope can vary, from mere amusement and derision to deep hatred. But what is it for a contempt to be fitting as an emotion? Just as for anger, the objects of contempt can be very diverse, in the sense that the kind of behaviors which might elicit contempt and the judgment of superiority which is associated to it are themselves very diverse. If contempt is linked to the perception of a social hierarchy and directed at a social status, there will as many kinds of reasons for contempt are there are kinds of social standards within a society or from one society to another: being the object of contempt in India for being an untouchable, being the object of contempt at the court of Louis XIV in Versailles because one does not have the proper degree of nobility, or being the target of the contempt of a snob in Victorian England seem to correspond to so many situations that it is hard to say that there are precise correctness conditions for this emotion. Think for instance of the way Tocqueville describes the relationship between men and women in America as contrasted with Europe:

> It has often been remarked that in Europe a certain degree of contempt lurks even in the flattery which men lavish upon women: although a

European frequently affects to be the slave of woman, it may be seen that he never sincerely thinks her his equal. In the United States men seldom compliment women, but they daily show how much they esteem them. (Tocqueville 1835, Ch. 11)[7]

In a democratic society, which aspires to equal status for individuals, contempt will be an emotion which is itself the object of a general negative moral judgment. Kant (1785) famously says that contempt violates people's basic claim of respect against others because it construes people as lacking dignity and as having no moral worth, hence as unable of improving. Kant denies that there can be any good reason for contempt, hence that the reasons for being scornful are always *wrong*, because they contradict a basic principle of morality. This is bad news for the perceptual theory of emotions as tracking values, because, on this view, the evaluative judgment associated to contempt would always be false, and the emotion of contempt would be always inappropriate. As Bell notes, the rejection of contempt is due to its globalist nature. In this sense, it seems to be not only to be rejected for moral reasons, but also for descriptive ones. For, as many psychologists have argued, there are no such global emotions: characters traits are inconsistent, transitory, unstable.[8] So contempt never fits its object, and is always based on a false ascription of character. The objection extends to the "fitting attitude" conception of value: there is nothing like being fit for such an emotion, because we can never specify its conditions of appropriateness.

This objection presupposes that the fittingness of an emotion is a descriptive property. But this is wrong. The fitting attitude or "neo-sentimentalist" analysis of emotions, as it is sometimes called (Tappolet 2016, p. 85 sq.),[9] says that evaluative and normative concepts are essentially tied to the concepts of specific responses, although it does not say that these concepts are mere projections of our emotional responses. For instance a particular action counts as admirable if the feeling of admiration is an appropriate response to this action, of fits it. This condition of

---

[7]Indeed many contemporary American feminists would balk at this.

[8]See e.g. Doris (2005). This is the line taken by "situationism" about character or virtue.

[9]Tappolet (2016, p. 85 sq.).

correctness is meant to be a conceptual analysis. But it encounters two difficulties. The first is that it seems circular: we are told that an emotion is correct when the responses are fitted to the value or norm, admiration to the admirable, anger to irascible-worthiness, contempt to the contemptible. The second is that it is formulated in normative terms, involving appropriateness and correctness: the idea is that a certain kind of feeling is *required*, or such that we *ought* to have it in order to fit the value in question. But if it is supposed to be a version of the perceptual analysis, it does not suit well a naturalist construal of the psychological state of emotion, for it seems odd to say that we perceive normative concepts. We need not, however, accept this descriptive construal of the perceptual analysis—or if it is a consequence of it, we should reject this analysis.[10] The fittingness conditions do not refer to an actual perception of value, nor to descriptive traits of character in individuals, but to an ideal of what individuals ought to be. In this sense, the Kantian interpretation of contempt is right: it refers to a judgment about what this emotion presupposes about humanity. But does that mean, as Kant implies, that this judgment is wrong? After all, some features of humanity and some kinds of actions might be contemptible, and it might be correct to point them out, and thus to withdraw the principled attitude of respect which Kant deemed to be the very foundation of morality.

4   Contempt and Moral Knowledge: A Swiftean Story

Neither the perceptual model of emotional justification nor the fittingness account entail that our perception of value on the basis of emotions have to be direct or immediate, as if we reacted to injustice in a bout of furor, or to the vileness of a character in a hiccough of scorn. Most "moral" emotions, such as shame, pity or contempt are not episodic,

---

[10]I thus would disagree with Tappolet (2016), who aims to defend such a sophisticated version of the perceptual view.

but dispositional. They can become virtues or vices, in so far as virtues and vices are based on dispositions. And they involve complex judgings and appraisals. An essential feature of most emotions is the capacity to reappraise them, and to revise them in the light of further emotions and judgments.[11] When they reach a certain level of sophistication, we learn much more from literature than from neuroscience and social psychology. The thesis which I want to put forward, but cannot argue in the space of this essay, is that moral emotions like anger and contempt do not involve judgments about values but judgments about characters who instantiate, or fail to instantiate these values. In other words, they involve judgments about virtue and vice, as dispositions stemming from emotions and involving dispositions to good or bad behavior. But anger and contempt are not only complex emotions involving judgments, they are also, as dispositions and character traits, the object of our moral appraisal. This is why, in particular, contempt has a bad reputation, as it is based on the feeling of superiority toward others. Both the tradition and common sense take it as itself despicable, as based on a wrong relation to values and reasons for emotions. I want to suggest that it is not: not only there can be reasons for contempt, and there are reasons for wrath, but these are also good reasons.

The Christian tradition has taken contempt to be in its very nature opposed to the virtue of humility. Its name is *superbia*, a vice of superiority. But the literary tradition of comedy and of satire has promoted contempt as the proper attitude toward the vices of human nature. Jonathan Swift was an heir to both traditions: he was a devoted Christian and a stubborn satirist. He has often been described as the arch-contemptor, a master of scorn. His specific feelings have been described by Samuel Johnson (1779) as a mixture of "petulance and sarcasm", of "arrogance and raillery", of a man who has "wasted his life in discontent". Thackeray (1854) says that if you had been his inferior "he would have bullied, scorned, and insulted you", and if you had addressed him, he would have been, as Yeats calls him, "intense" and "vehement".[12] Swift himself said that his *Travels* "are erected upon

---

[11]This feature has been well analyzed in the pioneering work of Livet (2002).

[12]Johnson (1779) and Thackeray (1854).

a foundation of misanthropy" and his famous epitaph at St Patrick in Dublin says that *saeva indignatio* cannot anymore "lacerate his breast". He has been ascribed all the traditional vices of superiority: Arrogance, Superbia, Hypocrisy, Apathy, Cruelty, Greed, Jealousy, Recklessness, Bitterness, Gluttony, Lust, Wrath. Such accusations are often addressed to the satirist. He displays contempt, scorn, haughtiness toward those that he satirizes, and his message is merely one of pride and superiority. Those who show such superiority not only have to be blamed because they do not respect their fellow mortals, but also because they do not deserve any respect. Swift was no exception. His morality, most biographers suggest, was questionable. He was, in the words of the critic John Middleton-Murry (1954), a "hypocrite reversed", one who turns toward others the accusation of viciousness that he himself knows to deserve. Contempt and anger are the engines of Swift's satire. "In a Jest, he said us, I spend my Rage", preferring to 'encounter Vice with Mirth' (*Epistle to a Lady*, Poems, II, p. 218). As Claude Rawson says "The angers, of course, were all too real, but Swift was temperamentally equivocal about their display. Even when we may suppose them to have been at white heat, as in *A Tale of a Tub*, the brilliant aggressive vitality is designed, for all the intensity of its sting, never to lose its cool. The contemptuous energy with which he mimicked the forms of 'modern' egocentrism and the self-promoting typographical antics of what we now like to call 'print culture', is a billowing performance of indignant impersonation in which the force and incriminating accuracy of the aggression never shows loss of authorial composure" (Rawson 2014, pp. 1–2).

Swift's angers are often characterized as a form of moral hatred. Swift's contempt was clearly on the subjective side of Strawson's reactive attitudes, even though he often seems to adopt the objective stance. On the one hand, the satirist's expression of contempt through irony presupposes that he distances himself from the characters that he represents, and often takes them as unfree, as mere puppets ruled by their passions. As F. R. Leavis says,

> Swift's ironic intensity undeniably directs itself to the defense of something that he is intensely concerned to defend, the effect is essentially negative. The positive itself appears only negatively— a kind of

skeletal presence, rigid enough, but without life or body; a necessary pre-condition, as it were, of directed negation. The intensity is purely destructive. (Leavis 1952, p. 75)

Leavis contrasts Swift's irony with that of Gibbons, which "insinuates a solidarity with the reader (the implied solidarity in Swift is itself ironical—a means to betrayal)" (ibid.). On the other hand, the satirist could not judge and disvalue these characters if he did not adopt the participant's attitude. His irony is directed at humans in general, although he knows that he is one of them.

When we talk of Swift's anger and of his contemptuous feelings, we are not talking about biographical or psychological facts. We are talking about what the reader can read into his prose, and the portrait that he implicitly draws of himself as a contemptuous and angry figure. Anger and contempt are not feelings or emotions which are represented within Swift prose, as features of his fictional characters. They are inferred by the reader through certain cues, the most salient being his irony.

Swift, however, is quite clear, and literally so, about his moralistic intentions:

I have been only a Man of Rhimes, and that upon Trifles, yet never any without a moral View. (*Correspondence* iv, p. 52)

"There are two Ends that Men propose in writing Satyr," private Satisfaction and a public Spirit, prompting Men of Genius and Virtue, to mend the World as far as they are able. (*The Intelligencer*, vol. III, 1728, in *Prose Works*, XII, p. 34)

But if my Design be to make Mankind better, then I think it is my Duty. (*Prose Works* XII, p. 34)

You see Pope, Gay, and I use all our Endeavours to make folks Merry and Wise. (*Correspondence*, iv. p. 53)

I have finished my *Travells*…they are admirable Things, and will wonderfully mend the World. (Letter to Ford, 27 August 1721, *Correspondence*, III, p. 87)

I look upon myself, in the capacity of a clergyman, to be one appointed by Providence for defending a post assigned me, and for gaining over as many enemies as I can. (*Prose Works* ix, p. 262)

I have got Materials Towards a Treatise proving the falsity of that
Definition animal rationale, and to show it should be only *rationis
capax*. Upon this great foundation of Misanthropy (though not Timon's
manner) the whole building of my *Travells* is erected (*Swift to Pope*, 29
September 1725, *Correspondence,* II, p. 607)

But these edifying and moralizing intentions are also the satirist's
main problem: he intends to denounce human vices, but his very
denunciation is itself considered as vicious. His emotion is not appro-
priate. We have here an instance of the wrong kind of reasons problem
for contempt: the satirist, by expressing his contempt for mankind,
seems to be looking for *having the attitude of contempt*, and not to have
what is contemptible as the proper object of his emotion. In Parfit's
(2011, pp. 420–432) analysis of the wrong kind of reasons problem, the
satirists reasons are *state given* reasons, directed *at the attitude* of con-
tempt, and not *object given* reasons, that is reasons directed at what is
worthy of contempt. Contempt based on state given reasons is a form
of pretense, an emotion which is not appropriate, but phony or insin-
cere. So the satirist's project seems to be faked, and hypocrite, as many
critics of Swift have argued.

There are two dilemmas of satire, which Swift exemplifies almost par-
adigmatically. The first is that the satirist can deliver his moral message
only by sharing with his readers emotions, such as anger and contempt,
which the readers find negative and so despise. This is also why sat-
ire, and the kind of negative feelings it expresses, is often self-directed.
There is a second dilemma for the satirist: his main weapons are irony
and fiction, but how can these modes of expression carry the weight
of his reprobation? If the satirist is understood too literally, the satire
is unsuccessful, and if he is too ironical and too fictional, he will not
be taken seriously, and will fail to deliver his moral message. The art of
fiction in general requires the appropriate tuning of the emotions. This
is true of satire and of the expression of contempt. The satirist is always
suspected of either pretending to be contemptuous, or if he is genu-
inely contemptuous, to be immoral because he does not have the proper
humility and respect for humanity which are these bases of morality. If,
like Swift, he is also a Christian, his case is almost desperate.

The solution to this problem is to reject the common understanding of contempt, as a bad emotion, which involves an absence of respect toward its target, and a scorn for humanity. Swift is a misanthrope only as far as he directs his scorn towards people who are themselves vicious, and at the kind of situations that they create: political injustice in Ireland (*The Drapier* letters), famine and poverty with the children of Ireland (*A Modest Proposal*), bigotry and credulousness (*A Tale of a Tub*), false learning as with the Academy of Lagado (*Gulliver*), and all the situations which are the object of his rage. So his contempt is for the right kind of reason: it is directed at real situations which exemplify moral wrongness, stupidity and vice. Swift's emotions of anger and contempt, as they are manifested in his satires, are aimed at the right targets. They are disrespectful, because in such situations, it correct, and fit, to "vex the rogues". He is clear on the fact that it does not amount to hate of mankind:

> I tell you after all that I do not hate Mankind, it is *vous autres* who hate them because you would have them reasonable Animals, and are Angry for being disappointed. (*Swift to Pope*, 26 November 1725, *Correspondence*, IIII, p. 118)

The satirist here relies on, and displays, a form of moral knowledge. Indeed, it is disputable what this kind of knowledge amounts to. Is it, as Swift seems to imply, a knowledge of morals truths, based on an acquaintance with real values, on the basis of which the emotion of contempt produces the implicit judgments of the satirist under the guise of his "mirth", as the moral realist would be tempted to say? Or does this knowledge consist in some form of understanding which falls short of being genuine knowledge (Brady 2013)? I side with the first, but indeed this claim is far from evident. A non-cognitivist about moral values and norms will claim that if there is no agreement on what the moral truths are, there cannot be moral knowledge, either in ordinary life or in its expression in satires. So the view proposed here is bound to seem question-begging. But I hope to have at least indicated how there can be *room* for a moral realistic view of the emotions of anger and contempt.

References

Bell, M. 2013. *Hard Feelings: The Moral Psychology of Contempt*. Oxford: Oxford University Press.

Brady, M. 2013. *Emotional Insight*. Oxford: Oxford University Press.

Brogaard, B., and E. Chudnoff. 2016. Against Emotional Dogmatism. *Philosophical Issues* 26: 59–76.

Darwall, S. 2018. Contempt as an Other-Characterizing, "Hierarchizing" Attitude. In *The Moral Psychology of Contempt*, ed. M. Mason, 193–216. London: Rowman & Littlefield.

Deonna, J., and F. Teroni. 2012. *The Emotions*. London: Routledge.

de Sousa, R. 1987. *The Rationality of Emotions*. Cambridge, MA: MIT Press.

———. 2011. *Emotional Truth*. New York: Oxford University Press.

Doris, J.M. 2005. *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Echeverri, S. 2017. Emotional Justification. *Philosophy and Phenomenological Research*, 1–26. https://doi.org/10.1111/phpr.12453.

Gordon, R. 1987. *The Structure of Emotions*. Cambridge: Cambridge University Press.

Johnson, S. 1779. *Lives of the Poets, Selections*. Oxford: Oxford World Classics, Oxford University Press.

Kant, E. 1785. *Groundwork for the Metaphysics of Morals*, ed. Timmermans. Cambridge: Cambridge University Press (2012).

Leavis, F.R. 1952. The Irony of Swift. In *The Common Pursuit*. London: Chatto & Windus.

Livet, P. 2002. *Emotions et rationalité morale*. Paris: Presses universitaires de France.

Lombardo, P. 2014. Literature, Emotions and the Possible: Hazlitt, Stendhal. In *Mind, Values and Metaphysics: Philosophical Essays in Honor of Kevin Mulligan*, ed. A. Reboul, 117–134. Berlin: Springer.

Middleton-Murry, J. 1954. *Jonathan Swift, a Critical Biography*. London: Jonathan Cape.

Mulligan, K. 2003. From Appropriate Emotions to Values. *Monist* 81: 61–88.

Parfit, D. 2011. *On What Matters*, vol. 1. Oxford: Oxford University Press.

Pelser, A. 2014. Emotion, Evaluative Perception and Epistemic Justification. In *Emotion and Value*, ed. S. Roeser and C. Todd, 107–123. Oxford: Oxford University Press.

Pryor, J. 2000. The Skeptic and the Dogmatist. *Noûs* 34 (4): 517–549.

Rabinowicz, W., and T. Rønnow-Rasmussen. 2004. The Strike of the Demon: On Fitting Pro-Attitudes and Value. *Ethics* 114 (3): 391–423.

Rawson, C. 2014. *Swift's Angers*. Cambridge: Cambridge University Press.

Robinson, J. 2005. *Deeper Than Reason*. Oxford: Oxford University Press.

Skorupski, J. 2010. *The Domain of Reasons*. Oxford: Oxford University Press.

Stoker, M. 1987. Emotional Thoughts. *American Philosophical Quarterly* 24: 59–69.

Strawson, P.F. 1962. *Freedom and Resentment*. London: Methuen (1974).

Swift, J. 1898. *The Prose Works of Jonathan Swift*, 12 vols, ed. Temple Scott. London: Bell & Sons.

———. 1958. *Poems of Jonathan Swift*, 2nd ed., 3 vols., ed. Harold Williams. Oxford: Clarendon Press.

———. 1999. *Correspondence of Jonathan Swift*, 4 vols, ed. D. Woolley. Bern: Peter Lang.

———. [1726] 2001. *Gulliver's Travels*, ed. C. Rawson. London: Penguin Books.

———. 2013. *Parodies, Hoaxes, and Mock Treatises*, ed V. Rumbold. Cambridge: Cambridge University Press.

Tappolet, C. 2016. *Emotions, Values and Agency*. Oxford: Oxford University Press.

Thackeray, W.F. 1854. Jonathan Swift. In *The English Humourists of the Eighteenth Century*. London.

Tocqueville, A. de. 1835. *De la démocratie en Amérique*. Paris. English trans. H. Reeves. 2003. *Democracy in America*. Clark, NJ: The Lawbook Exchange.

Part VI

The Epistemic Value of Group Level Emotions and Moods

11

Emotions In-Between: The Affective Dimension of Participatory Sense-Making

Laura Candiotto

1   Introduction: From 4E Cognition to Social Epistemology

For cognitive science, cognition is a mental activity which comprises a range of different mental processes as perception, memory, and attention. Among these processes we also find emotions: emotions are not any more understood as outside this fundamental process that brings us to acquire or retrieve knowledge, as part of our philosophical tradition has thought, but they belong to this complex cognitive orchestration. However, as Colombetti (2014) has clearly highlighted, this important conceptualisation that has been developed by cognitive science, e.g. that emotions are part of cognition, does not necessarily imply that cognition is always affective, or that bodily feelings matter for the cognitive success, for example. Instead, 4Ecognition, the approach for which cognition is embodied, embedded, enactive and sometimes extended,

L. Candiotto (✉)
Free University of Berlin, Berlin, Germany
e-mail: laura.candiotto@fu-berlin.de

has the potentialities to provide a better grasp on the function of emotions in knowledge, since it assumes a broader account of cognition, for which cognition is the activity of an embodied agent who makes sense of the environment in which she acts.[1]

In this chapter, I assume 4Ecognition as background, and I will address the question about the function of emotions in knowledge from the point of view of epistemology. In fact, not all the cognitive processes bring about knowledge as true beliefs: cognitive processes give us information about the world and ourselves, but this does not necessarily mean that the acquired information is correct. Very often we make epistemic mistakes, as unfortunately we all well know. We do them not only for biases in the processing of our brain, as in visual illusions, but also for biases in building judgements, making choices, or revising beliefs. As when we rely on an untrustworthy testimony for building the belief that we should turn left at the corner, instead of right, for reaching our destination, or when we believe that a fruitcake would be the best gift for our friend's birthday, forgetting that she has a food allergy to kiwi. And many times, emotions play a role in these mistakes, especially in practical reasoning, as when we choose to buy an expensive dress moved by the excitement of the moment, although knowing that this money is necessary for paying our flat monthly rent, or when we regret to have revised our decision to go to visit our beloved grandmother because we were not in the mood. Therefore, one of the main jobs in epistemology is to analyse the conditions that are required for the generation of true beliefs, assessing when a cognitive process is truth-conducive.

In this chapter, I analyse the conditions for ascribing to emotions a beneficial function in the social-cognitive processes that brings to knowledge as true belief.[2] Doing so, I highlight their value for

---

[1]See Ward and Stapleton (2012) for a well-informed discussion of the 4E approach. See Gallagher (2014) for the emphasis I'm putting here in the notion of "action" for understanding the conceptualisation of cognition in the 4E approach, especially regarding its extended and enactive segments.

[2]This chapter does not deal with the hot question of group minds, group mental states, or group knowledge, but it looks at the collective effort of generating knowledge in epistemic communities. It thus discusses how individuals collaborate in revising beliefs, building a defense, or interpreting a text, for example, focusing on the conditions required by the right performance of these

knowledge: emotions are not just detrimental for knowledge, as in the cases I have just mentioned, they can be very helpful too. As I have argued elsewhere (Candiotto 2017b, d), certain emotions are beneficial as (1) motivation for knowledge and (2) building blocks of intellectual virtues. Here I address this topic within the social dimension of knowledge, moved by the conviction that emotions are properties of relationships and, thus, I depict their functionality in a well-defined social epistemic activity, i.e. participatory sense-making.[3] I first ask why could emotions be beneficial for the collective processes of knowledge, especially discussing Battaly (2018) on intellectual virtues and arguing for a conceptualisation of emotions as *socially extended motivations* in virtue epistemology; then, I discuss participatory sense-making (De Jaegher and Di Paolo 2007; De Jaegher and Di Paolo 2008; Fuchs and De Jaegher 2009), arguing for a fundamental role played by emotions in *boosting epistemic cooperation* and *determining the quality of social bonds*. Since certain emotions are at the ground of the processes of social bonding, they also play a function in those ties that constitute participatory sense-making. I argue that their specific function occurs in epistemic cooperation. Epistemic cooperation is what brings about the generation of a shared meaning in participatory sense-making and thus, being emotions socially extended motivations, they boost the relationships among the agents, colouring them with an affective dimension.

---

epistemic practices. Specifically discussing participatory sense-making, I analyse the interactions between the personal and the social, thus focusing on the epistemic practices that happen *in-between*. Also, I cannot engage here with the massive debate on the meaning of knowledge, especially with the critical segment that challenges the assumption that knowledge is a true belief and argues that knowledge is more valuable than mere true belief. I thus take here the standard definition of knowledge as a true belief but, since my thesis is not dependent upon it, I think it can be efficiently employed in other accounts too.

[3]Participatory sense-making is the enactive approach to intersubjectivity. Since it is a conceptual framework, not a specific activity only, the aims of the proponents are more ambitious than the mine here, i.e. to propose an alternative to the cognitivist approach to intersubjectivity (Di Paolo and De Jaegher 2017). I take this humbler stance because my aim here is to argue for a clear-cut thesis about the value of emotions in epistemic cooperation, and not for participatory sense-making per se. However, my focus on the activity, instead of arguing for the theoretical value of the model, is not at all alien to this research project, since I infer important implications about the affective dimension of participatory sense-making from the practice itself, also highlighting why participatory sense-making seems to be one of our best options for understanding the role of affectivity in epistemic cooperation.

2   The Value of Emotions in Social Virtue Epistemology

In the attempt to provide a solution to the difficulties posed both by foundationalism and coherentism to the theory of knowledge, virtue epistemology vindicates an essential role to intellectual virtues as truth-conducive, building the conceptualisation of the conditions that are required for assuring knowledge in the abilities, or character traits, of the epistemic agent.[4] This approach is important for discussing the function of emotions in knowledge because it ascribes truth-conduciveness to certain characteristics of the epistemic agent and, thus, the conditions required for the generation of knowledge are *subject-dependent*, at least on a certain level.[5] Notably, emotions are experienced by a subject and thus we can ask if emotions play a role in this epistemic enterprise or, putting it in another way, if emotions are among the agent's properties which contribute to form beliefs with a positive epistemic status. As I have already introduced, I think not only that emotions do play a function here, but also that this function can be positive, i.e. functional to the process of knowledge building. I individuate their beneficial functions in (1) motivating towards knowledge, understood as the most important epistemic good, and (2) building intellectual virtues, as their raw material that through exercise can be regulated within patterns of rationality (Candiotto 2017b, d). My argument says that once we recognise epistemic conditions to be virtue conditions, it directly follows from virtue theory that epistemic emotions are involved

---

[4] The two main approaches available nowadays in virtue epistemology, reliabilism and responsibilism, disagree in the very identification of what is truth-conducive, being the possession of certain abilities, as reliable memory or attention, for the reliabilists and the exercise of certain character traits, as intellectual humility or perseverance, for the responsibilists. See Sosa (1980) and Greco (2002) for reliabilism and Zagzebski (1996) and Baehr (2011) for responsibilism.

[5] See Code (1993) for the emphasis on taking the subject into account for epistemology. This line of thought has been developed by most of the feminist approaches to the theory of knowledge. This view cannot be ascribed to the entire spectrum of approaches labelled under "virtue epistemology"—it depends by what we mean for "subject" as the epistemic agent, if the individual person with her own and unique characteristics or a more neutral and universal subject of the epistemic agency. However, what counts here is that this subject-dependent approach in epistemology seems to fit better than others with 4Ecognition.

in the formation of the epistemic character. And we should recognise that epistemic conditions are virtue conditions not only because the subject's virtues should be taken into account for intellectual flourishing, but also because virtue epistemology is a normative discipline and, therefore, the conditions that lead to the achievement of the excellence of the epistemic process should be spelled out, starting from the virtues of its agent, the epistemic subject.

Here I bring this discussion within the social dimension of knowledge. The reason is not to simply explore a new avenue of research, but it deals with the very meaning I assume my object of inquiry possesses: emotions are not only experienced in the social dimension of intersubjective relationships (phenomenological level), but they are the product of our affective relationship with the world (ontogenetic level). As much of the phenomenological investigation on affects has shown, *affectivity is relational*, both in its passive (reaction to stimuli) and active (interested and intentional driving force towards an object) stance.[6] Therefore, studying the function of emotions in knowledge requires us to look at the emotions' proper place, that is the interactional dimensions of *epistemic communities*.[7] I will argue in what follows that attending to the interactions of epistemic communities is the right way to understand the contribution of emotions to epistemology.

Virtue epistemology has been assumed as an internalist and individualistic approach to knowledge, since it focuses on the conditions that should be assured by the epistemic agent, for example, intellectual courage and caution as the dispositions to respond appropriately to perceived threats in our intellectual lives (Roberts and Wood 2007, p. 219). Thus, it may seem problematic to look at this account for depicting the function of emotions in social knowledge. But I do not think that this is the case. Not only because anti-individualistic approaches to virtue epistemology and accounts on open-selves and

---

[6]For affects, I mean not only emotions but also feelings, sentiments, and moods. On emotions as relational affects in phenomenology, see Slaby (2008). On emotions as active and dynamic intersubjective processes in cognitive science and evolutionary psychology, see Griffiths and Scarantino (2008).

[7]As happened for the cognitive science, I believe that a wedding between phenomenology and analytic epistemology is not only possible but very promising too.

extended characters have started to flourish in the philosophical debate (Kallestrup and Pritchard 2016; Alfano and Skorburg 2017), but also because epistemic agency, as understood especially by the virtue responsibilism as development of abilities and performance of intellectual virtues, is by definition embedded within the social dimensions of our interactions with the world. Of course, these reasons are not sufficient for turning virtue epistemology to an externalist approach about social knowledge, since the first requires an assessment of the validity of these new approaches, and the second may be assumed as the quite obvious recognition that the external worldv plays a causal role in most of our cognitive processes, as for example in providing inputs to perception. But I think that they can at least motivate to explore a bit further into this internalist and individualistic assumptions of virtue epistemology. Therefore, I first introduce the answer provided by Heather Battaly (2018) in assessing the possibility of overlap between virtue epistemology and a segment of the 4Ecognition, the extended one,[8] and then I will explain why Battaly's virtue responsibilism seems to be a good starting point for replying to the question about the function of emotions in the social epistemic agency.

3   Extending Motivation to Shared Epistemic Practices

For Battaly (2018, pp. 195–197), the development of intellectual virtues is an activity over which the agent has some control, and for which the agent is (partly) responsible. An intellectual virtue requires epistemic action, not only for its development—as an acquired property—but also because it is a disposition to epistemic actions and, thus, intellectual virtues are performed in conducting inquiries,

---

[8]Other options are available nowadays for merging extending cognition and virtue epistemology, especially within virtue reliabilism (see Pritchard 2010; Carter 2019). Here I'm looking at virtue responsibilism because it is the approach that can better explain emotion-regulation as self-mastery in epistemic agency, and at Battaly's account because it is open to going beyond the standard individualism of virtue responsibilism.

assessing judgements, or revising beliefs, for example. But for Battaly and the responsibilists, intellectual virtues are not just dispositions to action: they are those character traits that make us the kind of thinker we are. This means that they are personal qualities that disclose personal values, showing, for example, if whether we care most for the truth or for personal utility, or if whether we build our judgements employing critical thinking or being are easily manipulated by social conditioning.

In discussing the most famous case of extended cognition implied by the parity principle, the Otto case,[9] Battaly agrees with Menary (2012, p. 157) in pointing to the fact that Otto should display the epistemic virtue of diligence for being able to effectively use a notebook as an integrated extended tool. They argue that diligence, in this case, is not just a mere and accidental character trait of the agent, but what really explains how the agent manages to develop such an extended process. For Battaly this explanation perfectly fits with virtue responsibilism: it is thanks to the employment of intellectual virtues within this process of integrating an external tool that the agent can have achieved cognitive success. Most importantly for our topic, Battaly also investigates a case of extension in social interactions,[10] the one of transactive memory for which we can recall events of the past thanks to dyadic or small group dialogues.[11] Unfortunately, Battaly does not spell out the specific function of intellectual virtue in this case, but she nevertheless sketches a model of motivation which extends beyond the brain that I further develop here for building on my proposal.[12]

---

[9]The case analyses the integration of an external tool, a notebook, for the formation of a belief (the correct route to reach the MoMA) and it argues for the functional equivalence between this case and a standard internalist formation of beliefs. See Clark and Chalmers (1998).

[10]There are at least two ways for labelling this social extension in the philosophical literature, as distributed cognition (Palermos 2016) and socially extended mind (Gallagher 2013). However, as I will argue in a moment, although both these approaches point to the right direction, we need to endorse an enactivist approach—as participatory sense-making, for example—for properly grasping the function of emotions in social knowledge.

[11]I discussed the fundamental role played by dialogical interactions in distributed cognition in Candiotto (2017a, 2019), also analysing the philosophical tradition about the epistemic valence of dialogue.

[12]I stress "philosophical investigation" because I think that much of the work which has been done by social psychology in overcoming the cognitive-situative divide also regarding internal motivation should be incorporated and philosophically discussed by philosophy, especially social epistemology. See Järvelä et al. (2010).

Differently from reliabilism, responsibilism argues that internal motivations are required for epistemic virtue, since you cannot be praised for a virtuous epistemic action if it is not, at least partially, under your control. Battaly, although maintaining this core character of responsibilism, argues that motivation should not be necessarily purely internal, but that it could be attributed to a system, as when we brainstorm in a small group about the best options we have for facing a problem. And the same could be said for epistemic responsibility, especially if, taking the same example, we share the commitment to generate the best answer possible to the problem. So far, so good. But what Battaly misses in this picture is the fundamental role played by emotions in establishing the social bonds among the member of the group, especially in motivating the quest for the best answer. What happens in the room where the interlocutors are brainstorming is not just a storm of brains, but also a storm of feelings and emotions. The room is not just an aseptic surgery room, but it is the place where the living experience of inquiry occurs, being filled with moods and atmospheres animated by the feeling bodies and their fluid interactions, and modulated by specific affective arrangements (Slaby 2018). I have to stress that this kind of criticism should be addressed to many recent works in epistemology, not just to Battaly's. And I have to add that, among these works, Battaly's one provides a framework that can be developed in order to make the role of emotions clear. Thus, this criticism is functional to the development of the philosophical conceptualisation of the role of emotions in extended motivation.

Take the relevance ascribed by virtue responsibilism to internal motivation: one of the standard ways of understanding emotions in the literature of philosophy of emotion is as *motivations to action* (Goldie 2002; Döring 2003). Emotions are the driving forces of our mental life and they display a certain degree of intentionality, thus disclosing what we value most. Epistemic emotions, especially, display the value of truth, understood as the most important epistemic good (Morton 2010; Brady 2013; Candiotto 2017b). It seems fair enough to conclude that bringing emotions to virtue responsibilism is not only important but necessary for better depicting our internal motivation to knowledge. But here we can do something more: if we accept that emotions are not just internal phenomena of an inner and hidden soul, but that they are properties of our social relationships, then we can develop what Battaly said about the possibility of bringing virtue responsibilism beyond individualism and arguing that emotions are *socially extended motivations* for group knowledge.

My argument says that since emotions are one important part of the motivational component of intellectual virtues, usually conceived as purely internal,[13] they are not only what drives our interaction with the world (as affective intentionality), but also a force that is experienced by a living body who acts in the social environment of relationships. Relationships are then the vehicles through which motivation is spread out to the group. This means that emotions, as motivations to action and raw materials of intellectual virtues, are *socially extended* and, thus, they bring intellectual virtues into the intersubjective dimension of epistemic practice.

It could be objected that emotions are automatic reactions to situations out of our control and, therefore, that they do not fit well with the conscious process of motivation which belongs to intellectual virtues. However, this objection does not take in due account the process of *emotion-regulation* as acquired self-mastery, within the other processes we undertake for shaping our mental life (Debus 2016). Our emotions—like many other components of our mental life—do not magically fit the epistemic goal by nature, but as responsible thinkers we develop the ability to regulate them within patterns of rationality. And this seems to be in harmony with both doxastic voluntarism and the importance of the development of abilities ascribed by virtue responsibilism to the epistemic practice. Moreover, as Battaly remarked (2018, p. 199), different degrees of control could be assumed for assuring a sufficient authority over epistemic actions, and thus we do not need to imagine a full-fledged self-mastery over emotions for assuring control over beliefs. Moreover, we could also conceptualise an embodied self-regulation, and thus not necessarily achieved through the employment of higher cognitive processes.[14] Finally, this kind of regulation is not supposed to castigate emotions, as part of our philosophical tradition has thought, but

---

[13]For Zagzebski (1996, p. 137), intellectual virtues have two components, the motivational and the success ones. The former, leading to produce a certain desired end, which is knowledge, is assumed to rely on an internalist view of knowledge, and the latter, which establishes the reliable success in bringing about that end, is assumed to be externalist.

[14]Think for example of how our anxiety dissolves with a good run, or how our anger softens practicing loving-kindness meditation. These examples could be conceptualised within the framework of embodied intentionality and, regarding the social dimension of interaction, with enactive empathy. For enactive empathy, see Fantasia et al. (2014).

to direct them to the epistemic goal, in this case.[15] As in one of the best examples provided by our philosophical heritage about directing our emotions towards knowledge: when Socrates urged Alcibiades not to be a lover of fame and power, but of wisdom, he helped him through dialogical interactions to direct his love to a most valuable object, the truth (Plato, *Alcibiades I* in Cooper 1997; Candiotto 2017c).

In assessing the possibility of going beyond a pure internalist account of intellectual virtues and, more generally, of knowledge, Battaly referred to extended cognition. I think that this is a good move and, as I have already said, it also offers the opportunity to introduce emotions in the picture. However, I don't think that extended cognition is the best place for getting the function of emotions in social epistemology because it misses, at least in its standard configuration, the phenomenology of affective agency as ongoing processes, embodied interactions, and mutual entanglements among agents in building meanings.[16] For finding a more suitable framework, we should look at another E of the 4Eapproach to cognition, that is enactivism, and specifically at the enactivist approach of intersubjectivity as participatory sense-making, that is the topic of the next part. Within this picture, I argue, we could understand why intellectual virtues, and emotions as one of their components, are beneficial for knowledge as functions of dynamical agentive systems which generates meanings.

4   Emotions in Participatory Sense-Making: Interaction and Cooperation

Ezequiel Di Paolo, Hanne De Jaegher and Thomas Fuchs (De Jaegher and Di Paolo 2007, 2008; Fuchs and De Jaegher 2009) have introduced to the debate on social cognition a new enactivist model for which

---

[15]In a more general way, we could say that they are regulated for the agent's well-being. But this is not in contradiction with the idea that we are regulating them toward knowledge since knowledge contributes to the agent's well-being. Of course, different accounts can be provided here for explaining why and how knowledge matter for well-being, from eudaimonism and perfectionism to evolutionary psychology, but what it is important to highlight here is that virtue epistemology seems to fit well with this perspective.

[16]This does not mean that we should not develop an account of extended emotions, but that for doing so we need to integrate phenomenology to the classical literature on extended cognition.

social cognition arises from dynamical, reciprocal and mutual interactions among at least two embodied agents. This model looks to both cognitive science and phenomenology: the cognitive science is the one of dynamical agentive systems theory, and the phenomenology the one of mutual incorporation. Here I'm going to develop its epistemology, especially discussing the function played by emotions in the establishment of social bonds in the epistemic communities. I look at both the cognitive and the phenomenological level: at the cognitive level I introduce the function of epistemic emotions in *boosting epistemic cooperation* in the generation of shared meanings, and at the phenomenological level I depict the function of emotions in determining *the quality of mutual incorporation*.

For the authors, intersubjectivity is not mind-reading, i.e. a solitary decryption of the others' inner thoughts,[17] but an embodied process of generation of shared meanings, made of reciprocal and continuous interactions between at least two agents.[18] The enactivist sense-making, the core cognitive operation pursued by a living being in the process of adaptation to its environment (self-organisation, in technical terms), is here brought to the intersubjective dimension of relationships among humans, and it is conceptualised as *participatory sense-making*, the coordination of intentional action in interaction which generates shared meanings. It is important to notice that embodied interactions as coordination processes are here at the ground of the cognitive processes in a very fundamental way. The picture does not depict a solitary and

---

[17]For the criticism addressed by the authors to Theory Theory (TT) and Simulation Theory (ST) see Fuchs and De Jaegher (2009, pp. 467–469). These are the five core elements of criticism: the presupposed internalism; the disembodied and disembedded stance; the denial of the embodied dimension of interaction; the underestimation of temporal locality for diachronic processes of development; the presupposed naïve realism (the world of the other as something given). See also Gallagher and Hutto (2008) for a similar criticism to Theory Theory and Simulation Theory grounded on the evidence provided by developmental psychology.

[18]Social interactions are thus processes, and not just relationships. For the important difference between processes and relationships, see Livet and Nef (2009). As De Jaegher, Di Paolo and Fuchs, also Livet and Nef criticise the standard interactionist approach for which interactions are just functions of a structure. But the novelty of the enactivist approach on social cognition is that it does not only argue that interactions are processes, but that interactions are participatory processes, focusing on the first-person perspective and, thus, ascribing a higher value to their phenomenology.

disembodied being who, at a certain point, undertakes a joint activity with someone else. On the contrary, we find the interactions (structural coupling, in technical terms) among the living beings at the beginning, and from there we study the sense-making that arises from their embodied and coordinated interactions. But the individual does not disappear from this picture under the priority of the social. In fact, social agents participate in each other's sense-making (De Jaegher and Di Paolo 2007). *Participation* does not mean a passive and accidental sharing of the social world, but it implies the active engagement of each individual in the joint activity of sense-making. Participation is the intentional act of taking part in the common experience, building its shared meaning.

I believe that emotions can illuminate the scene exactly regarding *the meaning of participation* in the practices of epistemic communities. From a cognitive level, participatory sense-making undertakes the dynamical agentive systems point of view for which social understanding is described as the process of generating and transforming meanings in the interaction and coordination between at least two autonomous agents (Di Paolo and De Jaegher 2007; De Jaegher and Di Paolo 2008; Fuchs and De Jaegher 2009). This means that at the ground of the cognitive practices we have interactions. But what the authors mean by "interaction" is not what dynamical systems theories of cognition[19] and the standard interactionist view means (Di Paolo and De Jaegher 2017). The authors argue that the standard approach to interaction is "spectatorial", that means a third-person view that looks at interactions as properties of a structure independent by the perceiver. But, they say, interactions are what we do in the active engagement with the world and the others, therefore we need to conceptualise them from the first-person perspective as "participation".[20] The point is not to look at a pristine intersubjective dimension as a view from outside that sees

---

[19]But it is important to mention that the authors borrow a lot from dynamical system theory, especially about the conceptualisation of interactions as processes extended in time and the function played by cooperation in them. For the role of coordination and cooperation in the group epistemic practice understood as a dynamical system, see Hutchins (1995).

[20]Or, put in other terms, but I think being at the same page regarding this point, adopting the second-person perspective of the "you". See Schilbach et al. (2013) and Zahavi (2015).

relations at the ground of our social life but to be aware that the show is always subject-dependent, and that it is continuously reframed by our social actions and sense-making.[21]

However, looking at the first-person perspective of agency does not mean that their approach is individualistic. In fact, participatory sense-making explores the reciprocal, mutual and coordinated dynamics between at least two autonomous agents. It is not only *A* who built her world giving meaning to *B* (as in the I-It relation *à la* Buber[22]), but both *A* and *B* build their world together via their reciprocal engagement. As Dan Zahavi (2015) claimed, this first-person perspective to cognition needs the second-person perspective too, that is the perspective of the reciprocity between me and you, the one that gets to the "we" as the interaction between me, you, and a shared meaning. For Di Paolo and De Jaegher, participatory sense-making brings the system to gain a life of its own, and individuals co-emerge as interactors with the interaction (De Jaegher and Di Paolo 2007, p. 492). As we saw in the previous section, the point is being able to go beyond individualism without losing the first-person perspective. In the previous section we looked at this via the extension of individual motivations—and the role played by emotions in the motivational component of intellectual virtues—to ingroup epistemic activities. Here I sketch this view beyond individualism through the analysis of the function of emotions in epistemic cooperation. Grasping the real meaning of interactions as participation we can better understand why emotions matter for knowledge as socially extended motivations that belong to us. And saying "us", I mean "me" and "you", those who actively cooperate in building meanings. And this will also be another way for arguing that emotions are not only private states, but qualities of our relationship with the world.

---

[21]Taking the observer into account also means to give values to concepts as autonomy and personal choices which risk being lost in a pure functionalist approach to interactions. And, most important for our topic, to emotions, as what disclose our values, what we care most. See Vanello (2018) for a critical discussion about the conceptualisation of emotion as the perception of values.

[22]For Buber, the real relationship is the one between I and Thou, ruled by reciprocity, immediacy, and difference. On the contrary, the I-It relation is objectifying, reducing the other to an object of our world. See Buber (1937 [1923]).

The strategy is to look at interactions as we-actions performed by individuals who are not already-made, but who continuously reframe their identity, conferring new meanings to themselves and the world in encountering the others.[23]

Participatory interactions imply a process of coordination as coordination *with*, and not of coordination *to* (Fuchs and De Jaegher 2009, p. 470), in which the meaning is co-created. Coordination is "the non-accidental correlation between the behaviours of two or more systems that are in sustained coupling" (De Jaegher and Di Paolo 2007, p. 490). We can bring this important conceptualisation of coordination to epistemology looking at the *I-mode* and *We-mode cooperation* (Tuomela and Tuomela 2005). In the I-*mode* the agent regulates the cooperative action following her own goals, in the *We-mode* the regulation is others-oriented and aims at the benefit of the entire group.[24] Consider now the specific case of epistemic cooperation, i.e. when agents cooperate for epistemic goals. I do not assume coordination and cooperation as synonyms, but I take embodied and interactive coordination at the ground of cooperation as a discretionary action of individuals to work together. In fact, cooperation involves some kind of meshing of goals and intentions—e.g. some kind of awareness that what I intend to do matches what you intend to do (and, perhaps, that this matching is part of what our shared intending aims at). I cannot finegrain the relation between these two concepts in human agency here, but just assume that cooperation implies a higher level of voluntary choice that seems to be required for bringing the discourse to epistemology, especially to virtue responsibilism as a normative discipline.

In distributed cognition, the coordination between different functions in the processes undertaken for pursuing a task is what leads to the creation of group cognition (Hutchins 1995). Let assume that the

---

[23]On the co-determination of self and other within an enactivist point of view to cognition, see Thompson and Varela (2001).

[24]In this model, individual aims are not denied but are understood within a wider cooperative activity which includes both the individuals and the group. This model wants to reply to some of the dilemmas which have been highlighted by decision theory about the supposed incompatibility between personal and group gains.

same function is performed by cooperation in the epistemic contexts. Consider the cooperation between two agents in *problem-solving*—let's say a dyad composed by the defense attorney and her client, where the problem to solve is to prove the client's innocence before the jury against the accusation of an alleged offense. Building a defense is an epistemic practice and it requires the criteria that we usually use to assess knowledge, such as justification and coherence, for example. The two agents need to cooperate in reaching the goal: for example, one by providing all the required information for an alibi defense, the other by providing her expertise in preparing the alibi defense. The two functions are clearly different and require different skills, but the task is achieved only if these functions are coupled, that is only if the two agents cooperate. And this cooperation is a we-mode cooperation since both of the agents look at the same epistemic goal, that is to build a convincing alibi defense.

Let now see what happens to our case having the affective dimension under scrutiny. Imagine the scenario *A* where a client has a defense attorney assigned by the court: she didn't choose her and she is sceptical about her abilities. She thinks that someone works as a government attorney only because she is not good enough to having her own firm. Thus, she will be suspicious and guarded, maybe also a bit discouraged and insecure, and she will not provide all the information required to properly build the defence. And then imagine the scenario *B* where a client trusts the attorney because she has successfully defended her cousin a few years ago and she is confident she will do the same for her. The client will then display confidence and hope and she will provide all the information required to build the alibi defense.

Certainly these two scenarios involve strong biases that have an impact on the behaviours (as, for example, thinking that a government attorney is worse than a private attorney), but what I first want to highlight here is the transition from certain beliefs (such as "she is not a good attorney because she works for the government", and "she is a good attorney because she saved my cousin") to the emotional behaviours, and how these emotional behaviours have not only an accidental impact in the process of cooperation between the client and the defense attorney, but how they boost or reduce epistemic cooperation.

It seems fair enough to say that epistemic cooperation in scenario *A* is threatened, whereas in scenario *B* it is supported, and the criterion for assessing the difference seems to be in *how much trust* the client would accept to give to the attorney. It follows that certain emotions—as confidence and hope in *B*—boost the trusting relationships, and others—as suspiciousness and discouragement in *A*—hinder the trusting relationships. From here we can conceptualise that certain kinds of emotions are beneficial for epistemic cooperation and that if they are embedded in the relationships that regulate social epistemic processes they are beneficial for the epistemic aims. Those are the emotions that we could define as epistemic. I do not argue that only certain emotions are epistemic, but that certain emotions play an epistemic function if they boost the epistemic process. Of course, certain emotions are more easily beneficial to the epistemic practices, as confidence and hope in nurturing trust in epistemic cooperation, but I do not think we need to set rigid boundaries among those emotions that could be epistemic and others that could not. For example, epistemic emotions are not necessarily positive emotions. Negative emotions, such as feeling ashamed of mistakes or the anxiety for the unknown, are fundamental for the epistemic processes of revision of beliefs and knowledge building. Neither should epistemic emotions be found in the "epistemic sector" only. In fact, some of the emotions that are commonly labelled as moral emotions, for example gratitude or admiration, can play epistemic functions if embedded in the process of establishing cooperative bonds.

As we saw analysing case *A*, this does not mean that every kind of emotion boosts cooperation, but only the ones that nourish trustful relationships, since trustful relationships are the ones that more than others move to shared goals.[25] Interactions are very precarious activities (De Jaegher and Di Paolo 2007, p. 487), and thus also epistemic cooperation is very fragile. The balance among the agents is dynamical and susceptible to change, and new variables can always arise, reframing the system. Consider again our scenarios: the feelings and behaviours of the client will be influenced by the feelings and behaviours of

---

[25]There is plenty of evidence about it in network theory and social capital literature about social trust and shared goals. See for example Newton (1997).

the defense attorney too. These scenarios are thus more complex than how I depicted them here, since the dynamics of variabilities that can make the client revise her beliefs are always at work. For example, in scenario *A*, the initial diffidence can be softened by a reassuring behaviour of the government attorney and, depending by the quality of the interactions, the client may revise her initial sceptical belief. Or, on the contrary, in scenario *B*, the initial confidence of the client can be disregarded due to the hasty behaviour of the attorney. Thus, not only a top-down process from beliefs to emotions is operating here,[26] but also a bottom-up process that moves from the emotionally charged relationships to beliefs.[27] That's why it is important in epistemology to look at which kind of emotions can boost cooperation, in their relation to intellectual virtues. Emotions have in fact the power to modify the quality of relationships. This means that the motivation towards epistemic cooperation, that we discussed in the previous section, is mediated by the quality of relationships, that in our scenario was dependent by the trust criterion. It is the quality of relationships what makes the difference. We cannot rule out a fixed percentage of influence because it depends on the contexts and by how much the interactions are what causes the cognitive activities.[28] But what we can establish is that, since emotions *determine the quality of the interactions*, certain emotions have the power of boosting epistemic cooperation, enhancing the level of intersubjective trust and motivation towards knowledge. The aim of the next section is to analyse how they can do it from a phenomenological perspective.

---

[26]And also, broadening the spectrum of analysis, from normativity to personal beliefs. This is important to highlight because, in our case, epistemic cooperation could be assumed as a normative structure which rules out the dynamics of social interactions. On the role of normativity in collective actions, see Brink et al. (2017).

[27]On the function played by emotions in the revision of beliefs, see Livet (2016).

[28]In this regard, De Jaegher et al. (2010), have differentiated among enabling condition (when the interaction facilitate/hinder the cognitive phenomenon), contextual factor (when interaction simply influences the cognitive phenomenon), and constitutive element (when interaction is part of the cognitive phenomenon). I add to this list the beneficial function, namely, a beneficial outcome is something that facilitates the cognitive phenomenon to occur, not only supporting the process but also enhancing it. The beneficial function confers the causality in place with a very positive value and I argue that epistemic emotions serve this function in the epistemic practices. See Candiotto (2017a).

5   Emotions in Participatory Sense-Making: Inter-Affection

In the previous section, I stated that emotions can illuminate the meaning of participation in the practices of epistemic communities and I argued that their function is to boost the processes of cooperation, determining the quality of the interactions. Here I look at the phenomenology of participation, and I analyse its affective dimension to explain the sense in which emotions determine the quality of interactions.

From a phenomenological point of view, Fuchs and De Jaegher (2009) analyse three different components of the processes of interaction: bodily resonance, affective attunement, and coordination of gestures and expressions (both in verbal and non-verbal communication), not only because these embodied interactions are at the ground of every action we undertake, epistemic ones included, but especially because regulated patterns of emotions can drive these interactions to participatory sense-making. As I have already said, emotions are motivations for action, and since mutual incorporation is achieved only through action, specific emotions bond the epistemic agents to their epistemic community, making them wish to collaborate, or to intentionally take part in a joint epistemic activity. In these cases, epistemic emotions as motivations towards knowledge are socially extended. But being so, and this is the point of this section, they confer a specific *affective colour* to the relationship they boost, and it is the specificity of emotion what matters for the intensity of boosting.

Consider a classroom as an example of epistemic community. Students meet on a daily basis, attending the same lectures and interacting in group activities. The goal of going to school is—or at least it should be!—learning, that means not only acquiring justified true beliefs about the world, and the methods to achieve them, but also improving understanding and thus being able to use what has been learned at school in daily life. Emotions play a fundamental role in motivating students to learn, not only in making the process juicier, but also improving the students' active participation in the process. Just think how important it is to engage in a "warm and welcoming" environment, made of reciprocal appreciation, serenity and gratitude, for pursuing group activities well, that means for being motivated to do

our best for achieving the task of our group. The "warm and welcoming" environment is not made of emotions as cognitive states only, but of bodily expressions of affection and respect, as when your classmates' faces light up when they see you, inner feelings of confidence and security, and acts of care, as smiling and sharing physical touch, for example.[29] Affects are in fact felt from within and perceived in others in an embodied manner.[30] Also, think about the curious attitude towards the unknown. Curiosity is not an emotion per se,[31] but wonder for the discoveries (and the thrilled excitement that run throughout your body) and even an aporetic feeling of uncertainty (maybe experienced with a drop in the stomach or with excessive sweating) are emotions related to a curious attitude towards the world, a disposition that has been recognised as crucial for actively participating in the processes of learning as an active learner (Brady 2009; Watson 2016). Therefore, the affective dimension, which I depicted here as experienced in the social epistemic dimension of the classroom, confers on the environment a specific flavour, qualifying it, in this case, as "warm and welcoming". It is the affective dimension, made of specific emotions, moods, and feelings, what makes the environment "warm and welcoming", or, saying it with other words, specific emotions, moods, and feelings characterise the interactions among the agents as "warm and welcoming".

Looking at the processes of learning in a classroom from the phenomenological perspective of participatory sense-making also means to

---

[29]A multi-component theory of emotion (Scherer 1984) seems the one that best represents the wider spectrum of affective experience, from physiological reactions to cognitive appraisals, passing through motor expressions, behavioural tendencies, and subjective feelings. However, as Slaby (2008) has highlighted, this approach often forgets that the embodied dimension does not only deal with the private experience of self-affect, but it possesses a precise intentionality for which we could detect bodily driving forces. Therefore Slaby (2008, pp. 434–440) argues for a model of affective intentionality from the perspective of the feeling body, describing its five central features: diffuse localization, world-directed intentionality, hedonic valence, self-consciousness, and motivational force. I do not need to argue for this model here, although I think it is one of the best available nowadays, because it is enough for my argument that the reader would accept that if we look at the phenomenology of emotional experience in participatory sense-making, we cannot take emotions just as mental states, but we need to embrace a wider perspective which includes bodily feelings, expressions, etc.

[30]De Jaegher (2015) argues that not only self-affection but also inter-affection is embodied.

[31]It seems correct to conceptualise curiosity as a metacognitive feeling (Litman 2009), but this does not exclude that some emotions are strictly related to a curious attitude toward the world.

see how intellectual virtues are embodied in the very same practice of learning, and how they are spread to the epistemic community of the classroom. This means that being motivated by our enthusiasm and interest for French poetry, for example, makes us ready to diligently translate into the best possible English our favourite poems, and also to let our classmates appreciate the beauty of the poems. But this also means that, if we are a teacher, we can engage the students into a participatory process of interpretation of the poems, asking them to interrogate themselves about the meaning of a poem for their life. These examples show not only how much emotions and intellectual virtues have to do with the process of participatory sense-making, but also how a participatory sense-making/4E framework can naturally accommodate what seem like plain phenomenological facts about learning, and that this is a point in its favour. Building together a meaning, giving sense together to something that we didn't understand before, is an epistemic activity that requires effort, strength, and motivation. And emotions and intellectual virtues can do a lot in supporting it. But they also qualify mutual incorporation of a very specific atmosphere. And this "qualification" has a strong impact on the practice which is performed within the environment. Let's take again the example of the warm and welcoming environment for the group's activities. What will happen within a hostile and judgemental environment, filled with envy and anger? I think that it is easy to imagine that this affective environment will be quite disruptive for ingroup and intergroup cooperation. Presumably, many of the members would prefer to do their job alone or would ask the teacher to change group. This means that the quality of the interactions is coloured by feelings, emotions, and moods and, therefore, that affectivity constitutes the quality of mutual incorporation in the epistemic enterprise of participatory sense-making. Finally, considering the different effects on participatory sense-making produced by the "warm and welcoming" and the "hostile and judgmental" environments, we can conclude that a responsible emotion-regulation of the motivational component is necessary for getting to the success component of intellectual virtues, as I argued in Parts II and III. But we should also be aware that emotion-regulation is not just a magical tool of an inner mind, but that it is triggered by the active engagement with the world: we can

influence our feelings, we can influence the others' feelings, we can be influenced by the others' feelings because we take part (we *participate*) in the social world. Although this puts us under threat of manipulation, this is good news, since it implies that we have the power to responsibly behave as a trustful epistemic agent, spreading our motivation towards cooperative activities. And this is also the answer to a criticism that can be addressed against my thesis about the value of emotions in epistemic cooperation. It could be highlighted that groups like mafia or nationalists cooperate, trusting each other most of the time and having emotions among their main motivations. In the same way, the defense attorney and her client in my example can cooperate in order to deceive the jury. But these objections underestimate the role I ascribe to intellectual virtues in emotion-regulation withing the normative dimension of epistemology. Of course, there are plenty of examples of manipulation of emotions in cooperation, but this does not deny that, under certain criteria, they really are beneficial for knowledge. This means that in order to get to know through social interactions, emotions in cooperation should be regulated by intellectual virtues and responsible epistemic agency.

Bringing participatory sense-making to epistemology does not imply that the meanings we build together would always be justified, or that the interpretations that we are going to provide could not be wrong; but this does not deny the epistemic valence of the activity, it just highlights that certain conditions are required for becoming *skilful sense-makers* or for making participatory sense-making a *reliable* epistemic practice. As I hope is clear enough now, I think that these conditions should be found in the virtues of the epistemic agents and of their interactions, and this also means that they should be found in the affective roots of intellectual virtues and in their embodied motivations, the emotions.

6   Conclusion: Emotions In-Between

Overcoming the individualistic stance in epistemology does not mean to deny the role of subjects in building knowledge. On the contrary, in this chapter, I assumed one of the fundamental principles of virtue

epistemology, which states that we should look at certain abilities of the subject for assessing the epistemic practices. I argued that we should look at certain emotions among the abilities of the subject, specifically those emotions that boost the motivation towards knowledge. At the same time, discussing participatory sense-making, I argued that we should look at the relationships among agents for understanding, both conceptually and phenomenologically, the role of emotions in knowledge. Emotions, in fact, are what determine the quality of relationships. Analysing epistemic cooperation I discovered that some of them empower the process of knowledge building since they are at the ground of trustful relationships. Emotions are thus *in-between*: they are among agents, in their reciprocal relationships; they are felt by our bodies, not just in a solipsistic way, but in the intercorporeality of our joint actions; they disseminate their motivational force in the socially extended dimension of cooperation. It follows that in this chapter I did not undertake the traditional interpretation for which emotions are private states of the soul, neither the more extreme position for which emotions are outside ourselves only, but arguing that they are *in-between* I focused on the relational and agentive dimension of participation which highlights the embodied interactions among individuals. Relational structures and agents are thus "equiprimordial" (Di Paolo and De Jaegher 2017), at least if we analyse the epistemic dimension of intersubjective generation of meanings. Looking at one of the sides only—the personal or the social—jeopardises the *action*, that for our topic means to not being able to grasp the dynamics that underline epistemic actions, among them the affective ones. Therefore, for understanding the role of emotions in knowledge we should look at the deep entanglement between the personal and the social, focusing on the *in-between* of participatory sense-making.

References

Alfano, M., and J.A. Skorburg. 2017. The Embedded and Extended Character Hypotheses. In *The Routledge Handbook of Philosophy of the Social Mind*, ed. J. Kiverstein, 465–478. London: Routledge.

Baehr, J. 2011. *The Inquiring Mind*. Oxford: Oxford University Press.

Battaly, H. 2018. Extending Epistemic Virtue: Extended Cognition Meets Virtue-Responsibilism. In *Extended Epistemology*, ed. J.A. Carter, A. Clark, J. Kallestrup, S.O. Palermos, and D. Pritchard. Oxford: Oxford University Press.

Brady, M. 2009. Curiosity and the Value of Truth. In *Epistemic Value*, ed. A. Haddock, A. Millar, and D. Pritchard, 265–283. Oxford: Oxford University Press.

———. 2013. *Emotional Insight: The Epistemic Role of Emotional Experience*. Oxford: Oxford University Press.

Brinck, I., V. Reddy, and D. Zahavi. 2017. The Primacy of the 'We'? In *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*, ed. C. Durt, T. Fuchs, and C. Tewes, 131–147. Cambridge, MA: MIT Press.

Buber, M. 1937 [1923]. *I and Thou*, trans. R.G. Smith. Edinburgh: T. & T. Clark.

Candiotto, L. 2017a. Boosting Cooperation: The Beneficial Function of Positive Emotions in Dialogical Inquiry. *Humana.Mente: Journal of Philosophical Studies* 33: 59–82. Special Issue: The Learning Brain and the Classroom, ed. A. Tillas and B. Kaldis.

———. 2017b. Epistemic Emotions: The Building Blocks of Intellectual Virtues. *Studi di estetica* XLV, IV Serie (7): 7–25. Special Issue: Epistemic Emotions, ed. L. Candiotto.

———. 2017c. On the Epistemic Value of Eros: The Relationship Between Socrates and Alcibiades. *Peitho, Examina Antiqua* 1 (8): 225–236.

———. 2017d. The Route of Goodness: Epistemic Emotions, Self-Realization, and Perfection. *Thaumàzein* 4: 243–258. Special Issue: Philosophy of Birth: Emotions and the Formation of Person, ed. G. Cusinato.

———. 2019. Plato's Dialogically Extended Cognition: Cognitive Transformation as Elenctic Catharsis. In *The Routledge Handbook of Classics and Cognitive Theory*, ed. P. Meineck, W.M. Short, and J. Devereux, 202–215. London: Routledge.

Carter, A. 2019. Virtue Epistemology and Extended Cognition. In *The Routledge Handbook of Virtue Epistemology*, ed. H. Battaly, 420–432. London: Routledge.

Clark, A., and D. Chalmers. 1998. The Extended Mind. *Analysis* 58 (1): 7–19.

Code, L. 1993. Taking Subjectivity into Account. In *Feminist Epistemologies*, ed. L. Alcoff and E. Potter, 15–48. New York and London: Routledge.

Colombetti, G. 2014. *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge, MA: MIT Press.

Cooper, J.M. (ed.). 1997. *Plato, Complete Works.* Indianapolis and Cambridge: Hackett Publishing.

Debus, D. 2016. Shaping Our Mental Lives: On the Possibility of Mental Self-Regulation. *Proceedings of the Aristotelian Society* cxvi, Part 3: 341–365.

De Jaegher, H. 2015. How We Affect Each Other: Michel Henry's 'Pathos-With' and the Enactive Approach to Intersubjectivity. *Journal of Consciousness Studies* 22 (1–2): 112–132.

De Jaegher, H., and E. Di Paolo. 2007. Participatory Sense-Making: An Enactive Approach to Social Cognition. *Phenomenology and the Cognitive Sciences* 6: 485–507.

———. 2008. Making Sense in Participation: An Enactive Approach to Social Cognition. In *Enacting Intersubjectivity: A Cognitive and Social Perspective to the Study of Interactions*, ed. F. Morganti, A. Carassa, and G. Riva, 33–47. Amsterdam: IOS Press.

De Jaegher, H., E. Di Paolo, and S. Gallagher. 2010. Can Social Interaction Constitute Social Cognition? *Trends in Cognitive Sciences* 14: 441–447.

Di Paolo, E., and H. De Jaegher. 2017. Neither Individualistic Nor Interactionist. In *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*, ed. C. Durt, T. Fuchs, and C. Tewes, 87–106. Cambridge, MA: MIT Press.

Döring, S. 2003. Explaining Action by Emotion. *Philosophical Quarterly* 211: 14–30.

Fantasia, V., H. De Jaegher, A. Fasulo. 2014. We Can Work It Out: An Enactive Look at Cooperation. *Frontiers in Psychology* 5, Article 874.

Fuchs, T., and H. De Jaegher. 2009. Enactive Intersubjectivity: Participatory Sense-Making and Mutual Incorporation. *Phenomenology and Cognitive Science* 8: 465–486.

Gallagher, S. 2013. The Socially Extended Mind. *Cognitive System Research* 25–26: 4–12.

———. 2014. Pragmatic Interventions into Enactive and Extended Conceptions of Cognition. *Philosophical Issues* 24 (1): 110–126.

Gallagher, S., and D. Hutto. 2008. Understanding Others Through Primary Interaction and Narrative Practice. In *The Shared Mind: Perspectives on Intersubjectivity*, ed. J. Zlatev, T.P. Racine, C. Sinha, and E. Itkonen, 17–38. Amsterdam: John Benjamins Publishing.

Griffiths, P., and A. Scarantino. 2008. Emotions in the Wild. In *The Cambridge Handbook of Situated Cognition*, ed. P. Robbins and M. Aydede, 437–453. Cambridge: Cambridge University Press.

Goldie, P. 2002. Emotions, Feelings and Intentionality. *Phenomenology and the Cognitive Sciences* 1: 235–254.

Greco, J. 2002. Virtues in Epistemology. In *Oxford Handbook of Epistemology*, ed. P. Moser, 287–315. New York: Oxford University Press.

Hutchins, E. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.

Järvelä, S., S. Volet, and H. Järvenoja. 2010. Research on Motivation in Collaborative Learning: Moving Beyond the Cognitive-Situative Divide and Combining Individual and Social Processes. *Educational Psychologist* 45 (1): 15–27.

Kallestrup, J., and D. Pritchard. 2016. From Epistemic Anti-individualism to Intellectual Humility. *Res Philosophica* 93 (3): 533–552.

Kimmel, M. 2009. Intersubjectivity at Close Quarters: How Dancers of Tango Argentino Use Imagery for Interaction and Improvisation. *Cognitive Semiotics* IV (1): 76–124.

Litman, J.A. 2009. Curiosity and Metacognition. In *Metacognition: New Research Developments*, ed. C.B. Larson, 105–116. New York: Nova Science Publishers.

Livet, P. 2016. Emotions, Beliefs, and Revisions. *Emotion Review* 8 (3): 240–249.

Livet, P., and F. Nef. 2009. *Les Êtres sociaux. Processus et virtualité*. Paris: Éditions Hermann.

Menary, R. 2012. Cognitive Practices and Cognitive Character. *Philosophical Explorations* 15 (2): 147–164.

Morton, A. 2010. Epistemic Emotions. In *The Oxford Handbook of Philosophy of Emotion*, ed. P. Goldie, 385–399. Oxford: Oxford University Press.

Newton, K. 1997. Social Capital and Democracy. *American Behavioral Scientist* 40: 575–586.

Palermos, O. 2016. The Dynamics of Group Cognition. *Minds and Machines* 26: 409–440.

Pritchard, D. 2010. Cognitive Ability and the Extended Cognition Thesis. *Synthese* 175 (1): 133–151.

Roberts, R.C., and W.J. Wood. 2007. *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Oxford University Press.

Scherer, K.R. 1984. On the Nature and Function of Emotion: A Component Process Approach. In *Approaches to Emotion*, ed. K.R. Scherer and P. Ekman, 293–317. Hillsdale: Erlbaum.

Schilbach, L., B. Timmermans, V. Reddy, A. Costall, G. Bente, T. Schlicht, and K. Vogeley. 2013. Toward a Second-Person Neuroscience. *Behavioral and Brain Sciences* 36 (4): 393–414.

Slaby, J. 2008. Affective Intentionality and the Feeling Body. *Phenomenology and Cognitive Science* 7: 429–444.

———. 2018. Affective Arrangements and Disclosive Postures. Towards a Post-Phenomenology of Situated Affectivity. *Phänomenologische Forschungen*, ed. T. Breyer, J. Jansen, and I. Römer, 197–216. Hamburg: Felix Meiner Verlag.

Sosa, E. 1980. The Raft and the Pyramid: Coherence Versus Foundations in the Theory of Knowledge. *Midwest Studies in Philosophy* 5 (1): 3–26.

Thompson, E., and F.J. Varela. 2001. Radical Embodiment: Neural Dynamics and Consciousness. *Trends in Cognitive Sciences* 5: 418–425.

Tuomela, R., and M. Tuomela. 2005. Cooperation and Trust in Group Context. *Mind & Society* IV (1): 49–84.

Vanello, D. 2018. Affect, Perceptual Experience, and Disclosure. *Philosophical Studies* 175: 2125–2144.

Ward, D., and M. Stapleton. 2012. Es Are Good: Cognition as Enacted, Embodied, Embedded, Affective and Extended. In *Consciousness in Interaction: The Role of the Natural and Social Context in Shaping Consciousness*, ed. F. Paglieri, 89–104. Amsterdam: John Benjamins Publishing.

Watson, L. 2016. Why Should We Educate for Inquisitiveness. In *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, ed. J. Baehr, 38–53. New York: Routledge.

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

Zahavi, D. 2015. You, Me, and We: The Sharing of Emotional Experience. *Journal of Consciousness Studies* 22 (1–2): 84–101.

12

Group Emotions and Group Epistemology

Anja Berninger

1   Introduction

In everyday conversation we often talk about groups as being both epistemic and emotional subjects. We talk of a jury having concluded that the defendant is guilty. We say that a search committee sees candidate A as being the best fit for the academic position advertised. We speak of a research group having come up with a brilliant new experiment. We also say that an expert panel was angry when their suggestions were ignored, that a group of football fans rejoiced when their team won the finals, and that a company threatened with a hostile take-over was dominated by fear.

In this paper, I would like to examine the relation in which such shared emotions stand to a group's epistemic states and activities. In what sense do shared emotions help or hinder our epistemic enterprises? How do they shape the way that groups engage in these

A. Berninger (✉)
University of Stuttgart, Stuttgart, Germany
e-mail: anja.berninger@philo.uni-stuttgart.de

epistemic undertakings? Are they a negative force (as we often seem to suggest in everyday discourse) or can they actually be of benefit and, if so, how?

Within the philosophical debate, the term "shared emotion" as well as the related terms "group emotion" and "collective emotion" have been used in a number of different ways. In this paper, I will use the terms "shared emotion" and "group emotion" synonymously to describe cases where two individuals or more experience the same (or a very similar) emotion. Furthermore, I will also discuss an additional condition sometimes put forward in the debate, namely the idea that the individuals in the group should also be aware of the fact that they are in the same sort of emotional state (Brady 2016; Salmela and Nagatsu 2016).

In focusing on these relatively minimalist definitions, I aim to avoid any discussion on whether and how emotions might be shared in a deeper sort of way as is often discussed in the phenomenologically inspired literature on the issue (see e.g. Thonhauser 2018). I will also by and large circumvent issues which concern the question of whether there might be group level emotions (such as e.g. group level fear) without any individual member of a group being in the emotional state in question (i.e. any member of the group experiencing fear) (see e.g. Huebner 2011). The reason I omit these interesting and important debates is that my main goal in this paper is to get clear on some aspects of the *epistemology* of group emotions, while these discussions focus on *ontological questions*.

Engaging with these epistemological questions forces us to bring together two philosophical debates: the debate on how we should best describe the nature and the structure of emotions and the debate on how we should best understand groups and their various activities. A very natural idea on how to bring together these two fields is to take up a dominant approach in the philosophy of emotions and then see how it plays out once we use it to think about groups instead of individuals. Following this natural-seeming approach, I will start by discussing what I take to be a (if not *the*) dominant approach in the philosophy of emotions. Namely, I will focus on the frequently made suggestion that we should understand emotions as perception-like

representations of value properties (Prinz 2004; Tappolet 2000; Roberts 2003; Döring 2007).

Thus, in the first section of the paper, I will describe what this view would mean in terms of group emotions and the role these play for group epistemology.[1] Overall, I will suggest that taking this approach to shared emotions and group epistemology does not lead to plausible results. This is partly due to what I take to be some of the inherent problems regarding this position on emotions. Most importantly, these perceptual positions do not take into account the far-reaching changes in cognition associated with emotional states.[2] Ignoring these features leaves us with an impoverished view on emotions, but it also stops us from fully comprehending the importance shared emotional states have for collective epistemic undertakings.

I then go on to sketch a position which builds on the cognitive changes in question. I suggest that this view will let us understand emotions within a group context not so much as information-bearing states, but rather as ways of thinking that can facilitate cooperation and create a joint epistemic outlook. Overall (and very roughly put), my suggestion will thus be that shared emotions (correctly understood) should be seen as important for epistemic undertakings because they turn groups into *unified epistemic subjects*. In making this claim, I am following positions originally brought into the debate on the role of shared emotions for group actions (Michael 2011; Salmela and Nagatsu 2016). Thus, we can say that the approach I advocate here tries to bring suggestions from the philosophy of action and group epistemology closer together so as to foster better understanding of the role emotions play for the epistemic enterprises of groups.[3]

---

[1]It should be noted that I do not wish to claim that this position is actually a dominant approach within the study on group emotions. Nevertheless, if one assumes that reflections on group emotions take their starting point in the study of individual emotions, this would be a natural place to begin the investigation.

[2]Here, I pick up on suggestions I have made in past publications, see Berninger (2016, 2017).

[3]For a recent study that analyses the role positive emotions have for a group's cooperation with respect to certain epistemic projects, see Candiotto (2017).

2   Shared Emotions as Representations of Value Properties?

In current philosophical literature, emotions are frequently suggested to be representations of value properties. There are several ways in which this general suggestion can be spelled out, the most convincing of which does not see emotions as akin to beliefs, but rather to more basic representational states such as perceptions. Thus, some authors suggest that in the representational content of an emotion, a value property is ascribed to a particular object. For instance, when I fear a tiger, according to this view, I represent the tiger as being dangerous (Tappolet 2000; Roberts 2003; Döring 2007). Other views suggest that the representation of the particular object is not part of the emotion proper. According to these views, my fear of the tiger amounts to a state with the representational content "danger" (Prinz 2004).

With this admittedly broad-brush characterization in place, I would now like to turn to the question of how we should describe the epistemic role of group emotions if we take these perceptual positions as a starting point. In so-doing, we first need to make the transition from speaking about individuals and their emotions to speaking about shared emotional states. There is some debate about what conditions need to be fulfilled so that we can truly speak of such shared emotions. For the purposes of this paper, I would like to largely circumvent this debate. Rather, for now, I would like to settle on what I take to be a relatively minimalist conception of shared emotion that avoids far-reaching ontological commitments. According to this conception (which I take from Brady 2016), we can speak of a shared emotion when the members of a group are in the same emotional state and each one of them has some awareness as to the fact that the other members are in this state as well. Thus, to *collectively* fear the tiger it is not sufficient that each one of us experiences fear. If this were the case, we would not be jointly, but individually feeling an emotion. Only once

we are aware of the emotion of others do we come to experience the emotion as a *shared emotion*.[4]

With this, let us now turn back to the representational view. Applying this to the case of groups, we might then say that group members share an emotion when they are in the same (or at least very similar) representational states and they have some form of awareness of the fact that the other group members are in such a representational state as well. Based on this, we may then ask how this influences the epistemic position of the group as a whole. One initially plausible answer runs like this: according to the representational view, emotions are representations of value properties. We might thus say that (in so far as they are correct) they give *information* on the value properties of a situation or object (e.g. the value properties of the tiger) to the individual who is experiencing the emotion in question. They thus function much like perceptions do for non-evaluative qualities. If we assume there is such a far-reaching analogy here, then the individual in the emotional state would (in many cases at least) be justified in forming a belief about the value property of the situation based on the emotion.

Building on these reflections, we can then develop a similar analysis for the group case. Here, if an emotion is widely shared in the group, the group members would first individually become aware of the fact that the situation or object represented has certain value properties, as well as the fact that other group members perceive them in a similar way. Based on these states, they could then individually form a belief about the presence of these value properties. When many or all

---

[4]Brady mentions two additional conditions to explain this claim somewhat further. He suggests that (frequently) shared emotions are the result of contagion (in the sense that one member of the groups catches the emotion from another). In other cases, emotions arise individually, but nevertheless there will be a connection between these individual emotional states in the sense that group members mutually endorse each other's emotional state (Brady 2016). In other words, there needs to be some level of emotional conformity (over and above group members just happening to be in the same emotional state) for us to be able to speak of shared emotions. In settling for this relatively open position, I leave much of the current debate on these issues aside. For an instructive overview of some recent developments in this field, see Michael (2016).

individuals in a group have such a belief (and perhaps again some sort of awareness as to the fact that other group members also have this belief), one could then ascribe the belief in question not only to these individuals, but also to the group as a whole.[5] To go back to the initial example: if we, as a group, come to fear a tiger, we each represent the tiger as dangerous (and are aware of the fact that the others are in a similar sort of representational state). This makes it more likely that we will each form the belief that the tiger is dangerous. And, given that each one of us has this belief (and we are aware of the fact that others have this belief as well), we can ascribe this very belief to the group.[6]

One might wonder whether the fact that we do not just individually fear the tiger, but that we do so with mutual awareness of the others' fear has any interesting *epistemic role* to play in this example. Indeed, it is at least *prima facie* plausible that it does. Thus, we might think that the fact that others share our emotional reactions (and seem to have beliefs based on that emotion) makes it more likely that we ourselves will take our emotional reaction to be appropriate (and thus in turn base our own beliefs on it).

As I already suggested, I do not want to claim that what I have sketched here is the most frequently voiced position on group emotions. Yet, it is a position that can plausibly be developed on the basis of popular positions in emotion theory. In my view, there is a kernel of truth involved in this way of explaining matters, but I still take the position to be misleading in several ways. Yet, I also think that this failure to convince is instructive for developing a more plausible position on the epistemic relevance of group emotions.[7]

---

[5]I am assuming that we can do so (in part) by using a simple aggregation function. For a related (but more complex) position on group belief, see List and Pettit (2013).

[6]There are different ways of spelling this out, of course, depending on one's overall epistemic position and on the question of how one analyses the epistemic role of perception and the formation of beliefs in groups.

[7]Of course, a full-blown analysis of the epistemic role of shared emotions would also need to take other positions into account. In this respect, Brady's account would be extremely instructive (Brady 2016). However, due to limitations in space, I cannot offer a thorough analysis of his view in the course of this paper.

Let me elaborate on some of the concerns I have: first of all, as I have already highlighted elsewhere, the perceptual position does not make good use of research indicating that emotions are not just representational states, but rather correlated with far-reaching changes in cognition. I will come back to this in the course of the paper, but to highlight just a few results of relevance: there is evidence that positive emotions enhance our creative problem-solving capacities, improve our cognitive flexibility, but also make us more distractible (Isen et al. 1987; Dreisbach and Goschke 2004). Anger, in turn, leads to a decrease in the rational weighing of action outcomes and patterns otherwise typical of hasty decision making (Leith and Baumeister 1996). In my view, there is good reason to think that these cognitive changes are part of the emotion proper and thus that this is an aspect that should be incorporated into a fully fledged theory of emotions and their structure (Berninger 2016, 2017).

A second (connected problem) that I would like to highlight here is that this position leaves us with a much too rosy picture of the effects emotions generally have on group epistemic processes. In everyday discourse, emotions often receive negative coverage. It is frequently suggested that they cloud judgement and thus turn us into less-well-functioning epistemic subjects. Interestingly, outside philosophy many tacitly assume that this can be the case even when the emotion in question fits the situation well. Thus, even if the tiger is dangerous (and fear might thus be considered as an appropriate reaction), emotions are still seen as potentially hampering to our ability to react to this threat. Representational theories tend to focus on explaining how (in contrast to this negative assessment) emotions might be epistemically and practically relevant (and rightly so). While I take much of what has been said in this debate to be important, I still think that we need to be able to explain why it is that emotions which fit the situation perfectly well can nevertheless have disastrous consequences for our thinking both on the group and the individual level.

A third problem I see here is not directly connected to the emotion theory involved. Rather, it has something to do with the way that groups *as epistemic subjects* are portrayed in the sketch given above. The position I outlined is very much focused on groups as entities that can

form beliefs based on the beliefs of their individual members. Prima facie, there is nothing wrong with this, but it still leaves much that is interesting about groups as epistemic entities out of the picture. There has been some discussion lately on whether groups might have certain epistemic traits such as e.g. being conscientious, creative and so on (Mathiesen 2011). Furthermore (but connected to this), it seems that a full-blown view of how emotions are involved in group epistemology should see groups as being the sort of entities that *engage* in epistemic undertakings. In other words, they are not just entities which collectively hold some belief or another, but rather groups that collectively weigh evidence, reason, reflect, search for answers, look for additional pieces of information and so on. They engage in certain activities to collectively further their epistemic goals.[8] If we take emotions to involve far-reaching cognitive changes of the kinds sketched above, then a theory connecting positions on group emotions and group epistemology should have something to say on how emotions feature in these various activities. In what follows, I would like to sketch such a view.

3   Emotions and Ways of Thinking

Below are some ideas of how the cognitive changes associated with emotions are relevant on the individual level.[9] Once the outlines of this position are clear, I will then turn to the group level and the epistemic activities of groups.

---

[8]In stressing the importance of these activities, I also take up ideas originally voiced by Seumas Miller. In a recent paper, he develops the idea of "joint epistemic action". According to this view, an epistemic action is an action with an epistemic end (such as knowledge, but also understanding, etc.). Examples of groups engaged in such joint epistemic actions according to Miller are a group of detectives trying to establish the identity of Jack the Ripper or a group of scientists trying to find a cure for cancer (Miller 2014). In this paper, I use the somewhat vaguer notion of an epistemic activity or an epistemic undertaking. This is primarily because there is some debate as to whether the talk of action is sensibly applicable to the realm of the mental (Strawson 2003). More recently, the importance of looking at epistemic activities in group epistemology has also been stressed by Proust, see Proust (2018).

[9]This section of the paper is largely based on my discussion in: Berninger (2016, 2017). Note, however, that in these reflections I do not take group emotions into account.

I will start off with highlighting what I take to be the kernel of truth ingrained in the representational theories of emotions. Thus, some perceptual theories put strong emphasis on the importance of certain patterns of salience for emotions (Roberts 2003; Döring 2007). This idea has also been taken up in recent non-perceptual theories of emotions. For example, Michael Brady has also suggested that emotions are intimately bound up with changes in what we attend to. He suggests: "One important aspect of this relationship is that emotions can make things *salient* for us, or can alert us to potentially important objects or events. For example, my anxiety over my new white carpet makes salient the fact that you're drinking red wine; […] my delight at being reunited with my beloved makes salient all of the wonderful things about her; and so on for many other cases of emotion" (Brady 2016).

Although, in my understanding, the empirical evidence on these issues is less clear than is frequently assumed, I do think that the suggestion has a strong initial plausibility, even if some of the details remain yet to be spelled out. Thus, in line with these suggestions I will assume that emotions really are bound up with certain shifts in attention.[10]

Let me now turn to the empirical results on the *cognitive changes* associated with emotions. I take these results to indicate that we should understand emotions, at least in part, as *manners of thinking* (Berninger 2016). The idea here (roughly) is to say that different emotion types encompass different characteristic changes in cognition. To further illustrate this with two examples: we might say that fear encompasses a fast-paced form of thinking, with a narrow scope of attention and a low variability of mental contents (Pronin and Jacobs 2008; Derryberry and Tucker 1994). Joy, in contrast, would be associated with a fast-paced form of thinking, with a broad scope of attention, a big-picture style of cognitive processing and a high variability of mental contents (Pronin and Jacobs 2008; Fredrickson and Branigan 2005; Johnson et al. 2010).

---

[10]The central issue here is that many of the experiments conducted are concerned with *trait* and not *state* emotions. See Berninger (2016) for a more detailed assessment of this issue. For more on the role that Brady assigns to attention within his overall theory of emotion, see Brady (2016).

Once we focus on the relation between emotions and these changes in cognition and attention, our understanding of the function of emotions also shifts. Thus, from this perspective emotions are not only (and not primarily) providers of information. Rather, we should understand them as solving certain mental processing problems we face.

An idea leading in this direction was brought into the discussion by Ronald de Sousa, namely the claim that reference to emotions can help us solve the "philosopher's frame problem" (de Sousa 1987, p. 193). The assumption behind this problem is that human organisms store large quantities of information, which they are in principle able to retrieve. In spite of this huge quantity, humans somehow manage to (mostly) only retrieve information which is relevant for the situation at hand. The question is how we are able to sift the relevant from the irrelevant without going through every piece of information accessible to us (which of course would take far too long and be too demanding). De Sousa's answer is, in short, that emotions limit the range of information available to the organism. They do so because (as I already suggested above) they are intimately bound up with specific ways of allocating our attention.

There is a second (philosophically less belaboured) problem of cognitive processing for which emotions play a decisive role. As Thomas Goschke has pointed out, flexible cognitive systems, like us humans, constantly face antagonistic demands on cognitive processing (Goschke 2013). Goschke suggests that there is a whole range of such competing demands, which he refers to as "control dilemmas".[11] I will only take up three examples which I consider to be especially relevant with respect to the relation between emotions and epistemic endeavours.

The first dilemma Goschke mentions is the so-called *shielding-shifting dilemma*. The central question here is, how easily bodily action or cognitive goals should be shifted. Thus, depending on the situation, it might be more or less beneficial to shield one's current action goals

---

[11]I take up this formulation here, though it should of course be stressed that these are not dilemmas in the strict philosophical sense of the term.

from competing demands or to react with a higher degree of flexibility (Goschke 2013).

The second dilemma Goschke mentions is the *selection-monitoring dilemma*. His suggestion is that there is always a question as to how much information an organism should take into account at any given time. More specifically, the question here is in how far aspects of a situation not directly relevant to the task at hand should be monitored. To take an example Goschke himself brings up: When working on a research paper, it is beneficial for the individual to concentrate on the task at hand, thus ignoring distractions (such as loud music emanating from a neighbour's apartment) as much as possible. Nevertheless, the individual still needs to monitor the environment at least to some degree. Thus, even the most dedicated researcher should not ignore the noise of the fire alarm going off (Goschke 2013).

A last example for the types of dilemma situations Goschke has in mind is the *exploitation-exploration dilemma*. The suggestion here is that choices need to be made between trying out new action options and drawing on previously learned action options. While new action options are more risky at first, trying out such new options enlarges the individual's action repertoire in important ways in the long run (Goschke 2013).

As these examples already indicate, Goschke is focused primarily on the link between emotions and *bodily* actions though I think that much of what I have discussed here can also be applied to purely *mental* activities such as weighing competing evidence or reflecting on what the best solution for a given problem would be. Here, we may well face constraints very similar to the ones highlighted by Goschke.

In addition to this, and based on the discussion thus far, Goschke's account can also be extended to include trade-offs which clearly apply to mental activities. Thus, one might claim that there are certain speed-accuracy-trade-offs. Usually, we only have a limited time span to reflect on a given problem so that, generally speaking, it might seem beneficial to process information at relatively high speeds. However, working at high speeds will potentially make us less accurate in our reflective processes. A second issue that quickly comes into view is a big-picture-detail-trade-off. Often, deliberation and reflection will be

enhanced through taking the details of a situation into view, but this is not always the case. Thus, in some cases, we must abstract away from the details to see the "bigger picture" and base our reflection on its central features (Berninger 2017).

There is no *one* solution to any of these antagonistic demands. Rather, we need to adapt our way of processing information and decision making to the situation at hand. This will always mean that not all demands can be met equally, but rather, trade-offs need to be made. For example, it seems beneficial to pursue risky, new action options in situations that allow for this sort of experimenting, while in other situations it may be much more beneficial for the individual to "play it safe".

If we think that emotions are (partly) defined as different manners of thinking, then it seems that they can play a decisive role in balancing these demands without us needing to make conscious decisions about how to process information (Goschke 2013). Because emotions are directly linked to certain manners of thinking, and those manners of thinking include aspects such as degrees of cognitive flexibility, thinking speed, detail drivenness, etc., they are well-suited for making the necessary trade-offs.

4 Shared Emotions and Collective Ways of Thinking

Let us now turn to the group level and discuss how this position on emotions might be further developed. Overall, the claim I wish to argue for here is that shared emotions work towards turning the group in question into a *unified epistemic subject*.

To see how this is the case, we briefly need to reflect back on what sharing an emotion means, given the position I have sketched thus far. A central aspect I highlighted on the individual level was that emotions involve certain manners of thinking as well as certain shifts in attention. If an emotion is shared, this means that several subjects also share a specific way of thinking and the respective patterns in attention and that they have some awareness of the fact that other group members are in the thus described emotional state as well.

There are two different ways in which such a shared way of thinking and attention allocation may be of relevance from the perspective of the group. First of all, we can stress that some of the mental parameters involved in an emotional state are relatively basic in nature. Thus, in the section before, I drew on empirical results to highlight the fact that different emotional states seem to involve different thinking speeds. It seems that changes in these basic parameters that are shared across different agents might be important facilitators of joint epistemic activities (such as jointly thinking through a problem). Thus, it seems plausible that interaction and coordination (such as communication about the problem) on the group level will run much more smoothly when all participants in the joint epistemic enterprise share some of the very basic parameters of their thinking. In this sense, purely cognitive endeavours might actually be similar to bodily action, where some authors have drawn attention to the importance of basic synchronization in bodily movements (Michael 2011).

This is not the only way in which shared emotions might be relevant to epistemic undertakings. As I have suggested above, emotions do not just encompass changes in such *basic* parameters. Rather, different types of emotion involve manners of thinking that differ along a whole set of aspects such as content variability, detail orientation as well as specific directions of attention. Thus, we can say that when subjects share emotions, they do not just show similarities in the sense of having similar representational contents or sharing a specific evaluation of a given situation. Rather, their whole way of thinking, reflecting and engaging in problem-solving should be similar. The same is true for the cognitive limitations the individuals face (such as the important aspects of a situation they may miss because of their current cognitive set-up). What this suggests is that when we share emotions, we actually also share a cognitive outlook on the world. These outlooks involve similar ways of problem-solving even in cases where individuals do not directly communicate about the issue at hand.

Expanding this somewhat further we can suggest that when groups share emotions of a specific type, they will be more likely to come up with typical kinds of ideas and solutions. Groups dominated by fear (and thus by fear-specific ways of thinking) are unlikely to come up

with particularly creative ideas or problem solutions. Rather, they are more likely to engage in fast-paced thinking and favour tried and tested action options. In this last aspect, they differ from groups in which joy is the dominant emotional state, since these groups are likely to be more creative in their problem solution.[12] It is on the basis of the different ways of thinking that we can ascribe groups with certain features such as being creative in an approach to a problem. We can thus see here that often (though not necessarily always) these ascriptions will be connected to specific emotions the group exhibits.

As in the case of the perceptual theory, one may wonder whether the fact that the group exhibits a shared emotion makes any *epistemic difference*. And indeed, one aspect of the answer is prima facie less clear than in the case of the perceptual theories. Thus, it seems that in the theory sketched here it is important that we are in the same emotional state. But, whether we know others are in the same emotional state as we are does not make any difference.

This is no criticism as such of the theory proposed, but it does highlight an aspect that might be perceived as a weakness of the approach. It may seem that really we are not talking about unified agency here in any interesting way. To illustrate this point, imagine a group of company employees who are supposed to come up with a new management strategy and all of which experience fear. We can thus say that the group in question will proceed in their collective mental processes in a fearful way. But, at the same time, the fear they experience can be directed at very different objects. For example, one employee might be afraid to lose her job, while another worries about the quality of the products offered by the company, while yet another is generally struck by anxiety when forced to engage in collective enterprises with her colleagues. One might suggest that the theory advocated seems to allow for these cases,

---

[12]In my view these differences are overlooked by Salmela and Nagatsu who already point towards a position similar to the one I am advocating here. In a recent paper, they suggest that "collective emotions lubricate creative processes, speeding imagination and recombination of ideas, thus enabling *group flow*" (Salmela and Nagatsu 2016). I think, here we need to be more careful in the sense that not all shared emotions will lead to these results, while some emotions (such as shared joy) might indeed do so.

but that really it makes no sense to speak of a *unified epistemic agent* in this case.

As a reply, it is important, first of all, to state that unification here is a matter of *degree*. A group can be more or less unified in this respect. In the case described above, we can say that the group will display unity in some respects of its thinking, but not in others. Thus, it is at least plausible to think that the group as a whole will exhibit a relatively fast-paced form of thinking. However, while all of the individuals are likely to have their attention focused on some threatening aspect of the situation, their attentional focus is likely to differ. Thus, one of the employees is likely to be especially focused on job security, another on the threatening aspects of the social interaction taking place and so on.

In my view, the fact that the position allows for these different degrees of unification is actually an advantage, not a disadvantage as may seem at first. Thus, it allows us to see how emotions may have some sort of unifying potential even in cases where that which is actually shared is quite limited. It makes us see how groups dominated by fear may be engaged in certain typical ways of thinking and decision making, even if different individual members fear quite different things (and for rather different reasons).

5   Conclusion

In this paper, I have suggested that shared emotions fulfil an important function within groups. They synchronize cognitive processes and initiate shared ways of thinking thus establishing a collective epistemic outlook. Understood in this way, emotions smooth cooperations within that group and ensure joint ways of problem-solving, thus turning the group into a (more) unified epistemic subject.

In my view, this position lets us see some of the benefits, but also some of the issues surrounding group emotions more clearly. Thus, we can see how group emotions might be beneficial for groups involved in epistemic undertakings. They guarantee that the individuals involved exhibit similar patterns of thinking. Therewith, they make the interaction between these individuals run more smoothly and they guarantee

that the group as a whole exhibits what one might call a joint epistemic outlook.

On a less positive note, it also helps us recognize why our everyday assessment of emotions is often negative in nature. Shared emotions, while being central to the formation of a unified epistemic subject, limit the outlook of group agents in important ways. Thus, the emotional states of the group may lead to some pieces of evidence being overlooked or (depending on the nature of the emotion in question) opinions to be hastily formed and so on. They can even do so in cases, where the type of emotional reaction (such as e.g. fear) seems adequate (in the sense that the situation in question really *is* dangerous), i.e. the way of thinking associated with fear can still be counterproductive in the situation at hand.[13]

In the course of this paper, I have focused on a minimalist position concerning the question of how we should define shared emotions. It should be noted that there are other suggestions in the field which assume that sharing emotions is not the only thing of interest. Thus, Bryce Huebner has developed the notion of *truly collective* emotions, which is much more demanding (ontologically speaking) than the one in play in this paper. According to his view, we should understand the group as functioning as a single system, with different aspects of the emotion being distributed across different individual components of this system (Huebner 2011). This makes it possible to think that a group can be in an emotional state (such as fear), while none of its individual members are in this state. I think that much speaks in favour of this approach and that it is indeed compatible with the theory of emotions that I have sketched here. We could imagine, for instance, that different aspects of a typical fearful way of thinking might be realized by different sections of a company, thus leading to a sort of distributed emotion. However, it is also clear that more work would need to be invested to show in greater detail how these approaches fit together.

---

[13]For fear this might e.g. be the case where the threat scenario we are facing is very complex in nature. Here, any solution might for instance demand high degrees of creative thinking, while fear fosters ways of thinking that do not exhibit the traits conducive to creative problem-solving.

References

Berninger, Anja. 2016. Thinking Sadly: In Favor of an Adverbial Theory of Emotions. *Philosophical Psychology* 29 (6): 799–812. https://doi.org/10.1080/09515089.2016.1159294.

———. 2017. *Gefühle und Gedanken. Entwurf einer adverbialen Emotionstheorie.* Münster: Mentis.

Brady, Michael S. 2016. Group Emotions and Group Understanding. In *The Epistemic Life of Groups*, ed. Michael S. Brady and Miranda Fricker, 95–110. Oxford: Oxford University Press.

Candiotto, Laura. 2017. Boosting Cooperation: The Beneficial Function of Positive Emotions in Dialogical Inquiry. *Humana. Mente Journal of Philosophical Studies* 33: 59–82.

Derryberry, Douglas, and Don M. Tucker. 1994. Motivating the Focus of Attention. In *The Heart's Eye: Emotional Influence in Perception and Attention*, ed. Paula M. Niedenthal and Shinobu Kitayama, 167–196. San Diego: Academic Press.

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

Döring, Sabine A. 2007. Seeing What to Do: Affective Perception and Rational Motivation. *Dialectica* 6 (3): 363–394. https://doi.org/10.1111/j.1746-8361.2007.01105.x.

Dreisbach, Gesine, and Thomas Goschke. 2004. How Positive Affect Modulates Cognitive Control: Reduced Perseveration at the Cost of Increased Distractibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30 (2): 343–353. https://doi.org/10.1037/0278-7393.30.2.343.

Fredrickson, Barbara L., and Christine Branigan. 2005. Positive Emotions Broaden the Scope of Attention and Thought-Action Repertoires. *Cognition and Emotion* 19 (3): 313–332. https://doi.org/10.1080/02699930441000238.

Goschke, Thomas. 2013. Volition in Action: Intentions, Control Dilemmas, and the Dynamic Regulation of Cognitive Control. In *Action Science*, ed. Wolfgang Prinz, Miriam Beisert, and Arvid Herwig, 408–434. Cambridge: The MIT Press.

Huebner, Bryce. 2011. Genuinely Collective Emotions. *European Journal for the Philosophy of Science* 1 (1): 89–118. https://doi.org/10.1007/s13194-010-0006-2.

Isen, Alice M., Kimberly A. Daubman, and Gary P. Nowicki. 1987. Positive Affect Facilitates Creative Problem Solving. *Journal of Personality and Social Psychology* 52 (6): 1122–1131.

Johnson, Kareem J., Christian E. Waugh, and Barbara L. Fredrickson. 2010. Smile to See the Forest: Facially Expressed Positive Emotions Broaden Cognition. *Cognition and Emotion* 24 (2): 299–321. https://doi.org/10.1080/02699930903384667.

Leith, K.P., and R.F. Baumeister. 1996. Why Do Bad Moods Increase Self-Defeating Behavior? Emotion, Risk Taking, and Self-Regulation. *Journal of Personality and Social Psychology* 71 (6): 1250–1267.

List, Christian, and Philip Pettit. 2013. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.

Mathiesen, Kay. 2011. Can Groups Be Epistemic Agents? In *Collective Epistemology*, ed. Hans B. Schmid, Daniel Sirtes, and Marcel Weber, 23–44. Frankfurt: Ontos Verlag.

Michael, John. 2011. Shared Emotions and Joint Action. *Review of Philosophy and Psychology* 2 (2): 355–373. https://doi.org/10.1007/s13164-011-0055-2.

———. 2016. What Are Shared Emotions (for)? *Frontiers in Psychology* 7: 412. https://doi.org/10.3389/fpsyg.2016.00412.

Miller, Seumas. 2014. Joint Epistemic Action and Collective Moral Responsibility. *Social Epistemology* 29 (3): 280–302. https://doi.org/10.1080/02691728.2014.971908.

Prinz, Jesse. 2004. *Gut Reactions: A Perceptual Theory of Emotions*. Oxford: Oxford University Press.

Pronin, Emily, and Elana Jacobs. 2008. Thought Speed, Mood and the Experience of Mental Motion. *Perspectives on Psychological Science* 3 (6): 461–485.

Proust, Joëlle. 2018. Consensus as an Epistemic Norm of Group Acceptance. In *Socially Extended Epistemology*, ed. J. Adam Carter, Andy Clark, Jesper Kallestrup, S. Orestis Palermos, and Duncan Pritchard, 132–154. Oxford: Oxford University Press.

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

Salmela, Mikko, and Michiru Nagatsu. 2016. Collective Emotions and Joint Action. *Journal of Social Ontology* 2 (1): 33–57. https://doi.org/10.1515/jso-2015-0020.

Strawson, Galen. 2003. Mental Ballistics or the Involuntariness of Spontaneity. *Proceedings of the Aristotelian Society* 103 (1): 227–256. https://doi.org/10.1111/j.0066-7372.2003.00071.x.

Tappolet, Christine. 2000. *Emotions et Valeurs*. Paris: Presses Universitaires de France.

Thonhauser, Gerhard. 2018. Shared Emotions: A Steinian Proposal. *Phenomenology and the Cognitive Sciences* 17 (5): 997–1015. https://doi.org/10.1007/s11097-018-9561-3.

13

In Search for the Rationality of Moods

Anthony Hatzimoysis

1   Introduction

A sad mood can affect our way of thinking by slowing down our reasoning process, making us pay more attention to detail (Wong 2016; Clore et al. 1994; Mineka and Nugent 1995; Clark and Teasdale 1982; Forgas 1992; Burke and Mathews 1992; Forgas 1995; Salovey and Birnbaum 1989; Forgas and Bower 1987; Isen et al. 1987). Elation appears to have the opposite effect, allowing an abundance of information to flow through our senses, painting a multi coloured picture of our social environment with a broad brush. Moods, therefore, are affects which themselves affect how we reason (Clore et al. 1994; Bless et al. 1992; Bodenhausen et al. 1994; Murray et al. 1990; Forgas and Fiedler 1996; Wegener et al. 1995; Isen and Daubman 1984; Wong 2016). However, that is not the issue I will explore in this essay.

A. Hatzimoysis (✉)
Department of History and Philosophy of Science,
National and Kapodistrian University of Athens, Athens, Greece
e-mail: ahatzimoysis@phs.uoa.gr

Moods also set the stage for a variety of human activities, enabling and facilitating, or obstructing and undermining our self-centred or other-regarding projects. Moods have, thus, a direct effect on the realisation of our intentions, setting limits on the potency of practical reason (Griffiths 1989). However, that is neither the issue I will address here.

What I am interested in is the rather more basic question as to whether moods themselves can be rational. In particular, I would like to consider why the question about the rationality of moods is one that is rarely, if ever, posed. Books and articles on the rationality, or the fittingness, or the appropriateness of emotions are currently an important part of the philosophical literature (De Sousa 1987; Frank 1988; Ben-Ze'ev 2000; Greenspan 1988; Goldie 2000). Why is there no such corresponding interest in the rationality of moods?

Answering that question is a complex matter, whose unfolding might have to take into account not only purely theoretical, but also practical or historical considerations, pertaining to the formation of the philosophy of emotion as a distinct discipline in the past couple of decades, its inheriting certain topics from other disciplines, primarily from moral psychology and from philosophy of mind, as well as the understandable tendency of junior academics to get things quickly published on issues which are already well-established—and the rationality of moods is not one of them.

Nevertheless, my own concern is not with the historical but with the conceptual issue of what it is about mood, as a specific type of affect, that makes it not easily amenable to standard models of rationality. We may think of that issue through an analogy with other mental types: doxastic states are or should be grounded on epistemic reasons, desiderative states are or should be grounded on normative—prudential or moral—reasons; are mood states in their turn grounded on some kind of epistemic, or practical, or even *sui generis* moody reasons?

The fact that this kind of question is not often discussed, makes me think that the default answer to it, is: 'No—moods are not grounded on reasons'. And my hypothesis as to why that counts as the default answer is that most philosophers of emotion, despite their many and deep disagreements, share a basic line of reasoning. The rationality of an affective

state is somehow depended upon how that state is related to what the state is about, its so-called intentional object; but, given that moods do not seem to bear an intentional relation to an object, it is hard to see how they can be in the offing for rational assessment.

The first part of the paper I shall look at the premises that informs the position that moods do not seem to be intentional states. I shall explore that issue in the context of the current debate over the representational content of affectivity. I will outline three ways of attributing intentionality to moods, raising for each one of them a series of problems. Although none of the problems on its own appears insurmountable, they jointly appear to undermine the plausibility of making sense of the rationality of mood states by giving prominence to their alleged intentional dimension.

Secondly, I look at an account that is encountered in the literature on the psychology and physiology of moods; the account sets moods as mechanisms whose function is to monitor the balance between the demands raised by our natural and social environment, and the physical or psychological resources we may expend in meeting those demands. Thus, moods might after all be subject to criteria of assessment with respect to how well they represent how one fares in the situation in which one finds oneself. That is a promising way to proceed in our exploration of mood states; it faces though some a formidable challenge when it comes to the phenomenology of mood experience, as I try to show in the second part of the paper.

Attention to the phenomenology of affectivity is not an optional means for enriching the conceptual analysis of mood; according to several philosophers of mind, phenomenology provides the right methodological tools for making sense of the apparently diffuse and all-enveloping character of mood experience. Instead of treating moods as a surface colouring of a evaluatively neutral environment, that approach sets moods as the inescapable background of our perceptual, cognitive, and desiderative engagement with reality. In the third part of the paper I address the approach to moods as background feelings and raise some doubts about the ability of that approach to provide standards of assessment of a mood state, standards that would permit appraising the mood itself as rational.

2   Moods as Intentional States

A state may count as rational in a theoretical sense if it represents correctly its intentional object: that is a general statement of the notion of cognitive rationality that seems to be currently in play in many discussions of affective phenomena (de Sousa 2011). Disappointment counts as rational if it concerns an actual failure, and fear is deemed rational if it is directed at a real and imminent threat. It should be noted that different notions have been applied in the literature for conveying what is distinctive of the cognitive rationality of affective states, including reasonableness, fittingness, appropriateness, and proportionality. For the purposes of this paper I treat all those notions—which in certain contexts are significantly different—as falling under the heading of rationality.

Thanks to its generality, that account of theoretical rationality for affective states is spacious enough to accommodate alternative approaches, which see affective phenomena under the heading of evaluative cognition, of appraisal, of judgment, of construal, or of perception (cf. Morag 2016, for an informative yet heterodox overview of the domain).

Despite their substantial differences, those approaches share a commitment to the role of affectivity in referring to something beyond oneself, and thus constituting a means (either direct and simple, or mediated and complex) of acquiring knowledge about the salient features of one's situation. However, the claim that rationality is somehow linked to the alleged representational function of affective states, marks that model of rationality as distinct from theories which focus on the practical, behaviourally adaptive, or strategic function of affective phenomena.

If we are to build a viable model for the rationality of moods by drawing on their ability to represent correctly their intentional target, it is imperative that we supply an adequate theory of the intentionality of moods. The modest aim of the present section is to show that such a theory is hard to come by; and thus, that an attempt to move from intentionality to rationality appears short-circuited from the start.

The standard way to present the intentionality of a mental state is by citing the objects with which it is correlated. In the case of moods, such a correlation looks difficult to sustain. I shall focus here on three scenarios encountered in the literature.

The first candidate for an intentional object that is attributed to moods is the whole world (Solomon 1976; Lyons 1980; Baier 1990). That proposal appears to do justice to the overwhelming character of many moods, the fact that moods pervade our experience, that they suffuse all aspects of our encounter with reality. However, taken literary, the suggestion that the world is the intentional object of our moods is, in my view, problematic for three reasons. First, the proposal employs a notion that is not easy to determine. It is not clear whether we are invited to think of the world as a maximally inclusive situation encompassing all others, or perhaps as an object which has in it everything, or as the totality of phenomena linked by a complex network of references to each other. It can be retorted that the proposal requires nothing more than a loose understanding of the term, as employed in ordinary contexts. However, that retort does not really answer our query; it rather shows that the appeal of the proposal trades on the ambiguity of the basic term it employs. Additionally, the proposal makes excessive demands on the representational capacities of ordinary subjects. An affective state that is intentionally correlated to the whole world would entail an ability to form representations that moves well beyond the perspectival, partial, and limited access to one's immediate environment. Finally, even if we manage to sort out the above issues, the suggestion that moods are intentionally directed at the world founders on the problem of distinguishability between kinds of affective states. For instance, to be outraged with the whole world is not a mood: it is an intense and global emotion.

The second proposal treats the generality of the intentional object in a distributive manner. Instead of setting one object (the whole world) as the intentional correlate of mood, it takes as object the members of a disjunctively defined set composed of anything that comes our way; for any object encountered, we have, during a mood experience, a corresponding intentional relation (Solomon 1976; Goldie 2000; de Sousa 2010;

Sizer 2000). That proposal has the advantage of offering a way of distinguishing between emotions (which are directed at a single object), and moods (which take an indeterminate plurality of objects). However, the proposal appears to me to fail on the issue of intelligibility. A major task for an account of intentionality is to help us make sense of the character of the relevant experience. By correlating the experience with its intentional target, the account should contribute to our understanding why the person is in the mood that she is. However, treating every different object in one's environment as the intentional object of mood, may render inexplicable why one is in the particular mood that he is. A mood is typically a state of long duration: it may last for hours, days, or more. During that period, there is large number of objects which 'come one's way', and which differ in their evaluative shape. In a state of gloom, for instance, a person might happen to listen to a jolly tune. While it would be incorrect to think that just by listening to the tune the person will snap out of her sad mood, it would also be wrong to claim that the person is unable to notice the life-affirming air of that tune. However, decreeing that the jolly tune is the intentional object of the gloomy mood does not render the mood intelligible. Given the wide variety in the evaluative profile of the objects in one's environment, the postulated intentional connection between mood and any object that happens to come one's way, fails to discharge the task of rationalizing explanation.

The third attempt to articulate an account of the intentionality of moods is presented not by reference to some object, either singular or plural, but in terms of their representational content. We are invited to think that what is represented during a mood occurrence is a property, as such: in an anxious mood, it is threatingness, in an irritable mood, it is offensiveness, in a mood of contentment, it is delightfulness, etc. The property represented by a mood is not attached to anything in particular: it is unbound (Mendelovici 2013, 2014). The proposal admits that moods appear to lack intentional object; but it claims that this lack does not deprive them of intentionality, since moods are states with a content which represents an unbound affective property. That approach initially appears promising, since it sets a criterion for distinguishing emotions (which are directed at an intentional object), from moods

(which are not thus directed), while acknowledging that each mood has a different character precisely because it involves a different mental content, due to the unbound property that it represents (scariness, offensiveness, delightfulness, etc.). However, in my opinion, the proposal is problematic on conceptual grounds. More precisely, it is an account of intentionality that combines some conceptual confusions regarding the idea of an unbound affective property. Here I shall mention only one of those confusions, which concerns a fundamental metaphysical issue. The intentionalist claims that moods are intentional because they represent unbound properties. But what exactly is an 'unbound property'? Unless the proponent comes clear on this issue, her proposal is hard to understand, and even harder to assess. At a minimum, we require a disambiguation of the meanings that may be involved in that notion. First interpretation: unbound is a property that is not bound to one particular substance. That is a claim easy to understand, because it is trivial. No property is ever bound to one particular substance, each property is in principle instantiable at more than one place at once, and that is why we often call objects particulars, and properties universals. A second interpretation could be: unbound is a property that is not instantiated. If that is the meaning of the notion of 'unbound property', then it is hard to comprehend how someone can be in a particular affective state that represents a non-instantiated property. Notice that the intentionalist does not claim that someone is in a particular affective state because that state represents that a certain property is not instantiated: that would be the representation of a fact (or of a state of affairs, or of a proposition), and our representationalist denies that this is the meaning she intends. A third interpretation might venture to focus not on what type of property an unbound property is but on how we might think of it: 'unbound' is a property thought of not as instantiated by an object, but thought of merely as a property. Again, that claim offers little help in making sense of what is distinctive about the representational content of mood, or indeed, of any affective state. Considering a property merely as a property appears to me to be a task for logical, conceptual, or metaphysical analysis; it is simply not clear why the representation of the properties—not as being (or soon to be) instantiated, neither as owned (or soon to be owned) by some object, nor as exemplified

(or soon to be exemplified) by some situation, but—merely as properties, should have any affective significance at all.

The moral to be drawn from the above discussion is that we still lack a satisfactory account of the intentionality of moods; such a lack might, for some, indicate that moods enjoy a rather complex intentional relation to the world, whose structure is not easy to articulate; while for some others, it might corroborate their suspicion that no such intentional relation exists. In either case, appealing to intentionality may provide little joy to anyone who would attempt to ground the rationality of moods on their ability to correctly represent their intentional object. A theory of rationality that appeals on the state's alleged intentional relation to an object would be hard to sustain in the absence of a viable account of how an intentional relation between moods and objects might be possible in the first place.

3   Moods as Second-Order States

Scepticism about the prospects of the intentionalist accounts, have led to the articulation of alternative models that approach moods as second-order states, which bear no direct correlation to the world, but which may activate first-order cognitive, conative, or affective states that are intentional in their nature. The literature on the second-order approach to moods is voluminous, but here we shall limit ourselves to three theories that seems to be directly relevant to the question at hand.

The first theory treats moods as non-intentional states that select, out of the vast number of intentional states a person may have, which are presently active, and which remain latent (Lormand 1985; for critical discussion see Griffiths 1989; Wong 2016; Tappolet 2017). Hence, among our numerous beliefs and desires, only some of them enter into the explanatory, inferential, and justificatory processes which account for our, presently occurring, emotionally expressive behaviour, and our mental or physical actions. Moods are not to be identified with the first-order cognitive or conative states whose intentional contents are interconnected so as to give rise to certain pieces of behaviour or the performance of particular actions. Rather, moods form a higher-order

mechanism of selecting and, in that sense, activating those intentional states which form our direct engagement with reality. That theory has both some theoretical virtues; regarding, though, the question of the rationality of moods, it seems to me rather unhelpful. The problem it faces is simple, but rather hard to resolve: let us assume that a mood explains and—given certain assumptions about the rationalizing role of psychological explanation—renders intelligible, and, in that sense, justifies the first-order states it activates; however, what accounts for the occurrence, the explanation, or the justification of that mood itself? What could be the criterion for assessing, in the context of the present theory, whether the second-order state itself is rational? Unless we are offered a satisfactory answer to that question, the activation theory could not help in grounding the rationality of moods.

As we shall see, that basic problem reoccurs in models which, although more sophisticated, treat moods as something like 'the unmoved mover' of our psychological life. Take for instance the quite popular dispositional theory, according to which moods are temporarily heightened dispositions to make certain kinds of judgement, to form certain evaluative beliefs, or to proceed with certain sorts of appraisal (for discussion see Wong 2016; Tappolet 2017). Each mood is a second-order, relatively short-term disposition (as opposed to a subject's overall character, or emotional temperament), which marks that subject's susceptibility to a specific range of emotions. However, even if we grant that moods do operate as higher-order factors that determine which lower-order emotive states will occur, under certain situations, that would do nothing to assuage the worry that moods themselves are rationally groundless.

Consider finally, the functionalist approach which is currently gaining momentum in affective science. Moods are treated neither as representational states, nor as merely dispositional states, but as a mechanism which effects changes at a deep level of our cognitive organization. More precisely, that approach sets moods as mechanisms whose function is to monitor the balance between demands and resources: the demands raised by our natural and social environment, and the resources (physical or psychological) we may expend in meeting those demands. When the demands exceed the resources, the balance is

negative, and we feel 'low'; when the resources surpass the demands, the balance is positive, and the we feel 'high'. By monitoring one's current level of mental and physical energy in comparison to the demands generated by one's situation, mood serves the important function of setting up the agent to engage in the right task using the right amount of energy (Sizer 2000; Wong 2015, 2016).

That is a promising way to proceed in our exploration of mood states; it also provides a theoretical context for addressing the question of rationality of moods. If what makes something a mood is its discharging a certain monitoring role, moods might after all be subject to criteria of rational assessment with respect to how well—promptly, comprehensively, or accurately—they represent how one fares in the situation in which one finds oneself.

I believe though, that the theory faces a serious challenge at the level of the phenomenology of affective experience. Recall that the theory offers a criterium of telling the hedonic valence of a mood by means of checking how our energy repositories fare *vis a vis* the exigencies or requirements which we ourselves perceive as arising in the situation in which we find ourselves. Accordingly, we are invited to think that when the energy available exceeds the perceived demands of the situation, we are affectively 'plus'—what is roughly referred to as being in a 'positive' mood—and when the situation calls upon us to expend energy which our current psychological and physical state cannot supply, we are affectively 'minus'. That line of reasoning appears to me to have the following implication.

If the situation, *as perceived by us*, contains nothing inviting; if there is nothing to attract our interest in way that would trigger a desire to perform any cognitive or practical task; if, in a nutshell, the world around us involves nothing worth pursuing; then our energy level, however little it might be in absolute terms, it is evidently more than sufficient for meeting the energy requirements of a situation which we experience as raising no demands on us. Hence, according to the functionalist theory, that would be a case of being in a truly high mood. However, that claim is highly counter-intuitive, if not outright absurd.

Experiencing the world as devoid of any significance—as a field where nothing could spontaneously call for our attention, and where

values exercise no pull on our affection—is anything but 'being in a positive mood': on the contrary, it is characteristic of negative mood states, ranging from passing boredom, to persistent ennui, all the way to deep melancholy, and to major, clinical depression.

Recall that the theory under examination purports to account for the valence of moods in terms of their complex representational function; it could therefore be thought that the theory offers a ground for rationally assessing moods themselves in light of how well they discharge that function. However, looking at the particulars of that process, shows that the theory leads to absurd conclusions. At a minimum, it is a theory that rings false to the phenomenology of affective experience. Therefore, that theory, as it stands, cannot provide the right epistemological context for approaching the question of the rationality of moods.

4   Moods as Background Feelings

Phenomenology might be employed to the benefit of an account that does justice to the intimate relation between the way one apprehends the world, and the mood in which one is. Accordingly, the question of the rationality of moods might be better approached by paying close attention to the felt background of our sense of reality. Perhaps, contrary to traditional accounts of affectivity, the connection between mood and rationality runs deep, yet it is not often noticed precisely because it is ubiquitous and indirect. It is ubiquitous, since any engagement with world takes off from some mood state, whose very presence renders things around oneself salient as appealing or appalling, welcoming or annoying, familiar or uncanny. It is also indirect, since the mood is a pre-intentional state, constituting the background in the context of which intentionally directed emotions target their objects.

We may note that there is a variety of felt experiences listed under the heading of 'mood'. The present approach dwells on affective experiences which are variants of a non-localized, felt sense of reality and belonging, constituting a sense of how one finds oneself in the world as a whole (Ratcliffe 2008, 2015, 2017, 2019).

Accordingly, moods can be thought of as existential feelings, which form an inextricable constituent of our experience. We may distinguish between levels of existential feelings, differing in degrees of specificity and conceptual articulation, ranging from a pure feeling of being alive, through to feelings of homeliness or general insecurity, all the way to more sharply defined experiences, including anxiety, *ennui*, or elation (Slaby and Stephan 2008; Stephan 2012a, b). Alternatively, we may think of the felt rootedness of oneself in the world (what Heideggerians would attempt to convey with the notion of *Befindlichkeit*) as being ever present, yet subject to a variety of more specific configurations which mark the distinctive character of different world attunements (*Stimmungen*) (Hatzimoysis 2010; cf. Fernandez 2014).

The importance of such an account for our understanding of moods is hard to overemphasize. I am in sympathy with many of its methodological strictures, and consider its combination of first-person narrative analysis, and analysis of data from the neuropsychology of affectivity, as highly illuminating of mood phenomena. It is worth asking therefore whether that account can aid us in our search for the rationality of moods.

Generally speaking, we may assess the theoretical rationality of a state by checking its two ends: where it comes from (epistemically), and where it is heading at (cognitively). According to the account under consideration, moods are not headed towards anything: they are *non*-intentional states. That claim is not meant to imply that they are disconnected from the world; after all each mood constitutes one's sense of one's being in the world. Neither does it entail that moods are unrelated to our perception, cognition, or volition, since moods are taken as forming an anticipatory structure of experiencing the world, a structure which makes intentional, mental and bodily, acts possible. Hence, the term often employed by proponents of that account for characterizing moods is not 'non-intentional' but 'pre-intentional' (Ratcliffe 2019). However, that notion should not blind us to the fact that, according to that account, moods are not themselves intentionally directed towards anything. Hence, it is not possible to build a theory of the rationality of moods in terms of how well or badly they fit, or match, or represent, their intentional object.

What about the other end of the mood state: what we may call its epistemic source (or its cognitive basis, or its rational ground)? The answer to that question I think may be given through a further question: *what* source (or basis, or ground)? The latter question is intended as rhetorical, for it appears to me that a core claim of the present account is that mood states are not *grounded* but *grounding*. That is indeed the very reason why moods are characterized as *background* feelings, or as possibility *structures*, or as styles of *anticipation* of experience (Ratcliffe 2019; Slaby 2008; McLaughlin 2009). They do not follow upon intentional activity; rather, moods open up or foreclose certain ranges of possible experience, enabling or disabling different ways of engaging with things, allowing one to be attuned to the world and to one's own self, in some, usually unthematized, manner. If that is the case, then looking for an epistemic basis of moods will be in vain. A mood cannot be simultaneously what grounds and directs all epistemic activity, *and* what epistemic activity may independently assess, since any assessment would be already conditioned by what the corresponding mood deems as appropriate, or fitting, or correct—in one word, as rational.

5   Conclusion

Our short journey through the sea of contemporary theories of mood has steered toward a sceptical destination. We saw that some of the most prominent views about mood states fail to offer a secure ground for a viable account of the rationality of this type of affective states. Part of the explanation for that failure lies, in my opinion, with an implicit assumption made by most of those views, to wit that moods may somehow operate either independently of, or clearly prior to emotions. Perhaps, that, currently prevalent, hierarchy needs to be rethought. Alternatives that can be put on the table may range from a simple reversal of priorities-treating mood as derivative from a type of long-term or low-intensity (Delancey 2006) emotion, to a more radical re-categorisation of affective states, by thinking of both mood and emotion as varied exemplifications of feeling consciousness (cf. Hatzimoysis

2017), or passions (cf. Charland 2010). Addressing the merits and limitations of each one of those approaches requires a separate, and much longer analysis. However, what I hope is made apparent from our preceding discussion is that the search for a rationality of moods raises a host of philosophical puzzles that deserve our attention.

References

Baier, A. 1990. What Emotions Are About. *Philosophical Perspectives* 4: 1–29.

Ben-Ze'ev, A. 2000. *The Subtlety of Emotions*. Cambridge: MIT Press.

Bless, H., D.M. Mackie, and N. Schwarz. 1992. Mood Effects on Encoding and Judgmental Processes in Persuasion. *Journal of Personality and Social Psychology* 63: 585–595.

Bodenhausen, G.V., G.P. Kramer, and K. Susser. 1994. Happiness and Stereotypic Thinking in Social Judgment. *Journal of Personality and Social Psychology* 66: 621–632.

Burke, M., and A.M. Mathews. 1992. Autobiographical Memory and Clinical Anxiety. *Cognition and Emotion* 6: 23–35.

Charland, L. 2010. Reinstating the Passions: Arguments from the History of Psychopathology. In *The Oxford Handbook of Philosophy of Emotion*, ed. P. Goldie, 237–259. Oxford: Oxford University Press.

Clark, D.M., and J.D. Teasdale. 1982. Diurnal Variation in Clinical Depression and Accessibility of Memories of Positive and Negative Experience. *Journal of Abnormal Psychology* 91: 87–95.

Clore, G.L., N. Schwarz, and M. Conway. 1994. Cognitive Causes and Consequences of Emotion. In *Handbook of Social Cognition*, ed. R.S. Wyer and T.K. Srull, 323–417. Hillsdale, NJ: Lawrence Erlbaum Associates.

DeLancey, J. 2006. Basic Moods. *Philosophical Psychology* 19: 527–538.

de Sousa, Ronald. 2010. Emotion. In *Stanford Encyclopaedia of Philosophy*, ed. E. Zalta. http://platostanford.edu/archives/spr2013/entries/emotion/.

———. 2011. *Emotional Truth*. Oxford: Oxford University Press.

Fernandez, A.V. 2014. Depression as Existential Feeling of De-Situatedness? Distinguishing Structure from Mode in Psychopathology. *Phenomenology and the Cognitive Sciences* 13: 595–612.

Forgas, J.P. 1992. On Bad Mood and Peculiar People: Affect and Person Typicality in Impression Formation. *Journal of Personality and Social Psychology* 62: 863–875.

———. 1995. Strange Couples: Mood Effects on Judgments and Memory About Prototypical and Atypical Targets. *Personality and Social Psychology Bulletin* 21: 747–765.

Forgas, J.P., and G.H. Bower. 1987. Mood Effects on Person Perception Judgments. *Journal of Personality and Social Psychology* 53: 53–60.

Forgas, J.P., and K. Fiedler. 1996. Us and Them: Mood Effects on Intergroup Discrimination. *Journal of Personality and Social Psychology* 70: 36–52.

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

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

Greenspan, P. 1988. *Emotions and Reasons: An Inquiry into Emotional Justification*. New York: Routledge and Chapman and Hall.

Griffiths, P.E. 1989. Folk, Functional and Neurochemical Aspects of Mood. *Philosophical Psychology* 2 (1): 17–30.

Hatzimoysis, A. 2010. Emotions in Heidegger and Sartre. In *The Oxford Handbook of Philosophy of Emotion*, ed. P. Goldie, 215–236. Oxford: Oxford University Press.

———. 2017. Sartre on Affectivity. In *Thinking About the Emotions: A Philosophical History*, ed. A. Cohen and R. Stern, 272–290. Oxford: Oxford University Press.

Isen, A.M., and K.A. Daubman. 1984. The Influence of Affect on Categorization. *Journal of Personality and Social Psychology* 47 (6): 1206–1217.

Isen, A.M., K.A. Daubman, and G.P. Nowicki. 1987. Positive Affect Facilitates Creative Problem Solving. *Journal of Personality and Social Psychology* 52 (6): 1122–1131.

Lormand, E. 1985. Towards a Theory of Moods. *Philosophical Studies* 47: 385–407.

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

McLaughlin, B. 2009. Monothematic Delusions and Existential Feelings. In *Delusion and Self-Deception*, ed. T. Bayne and J. Fernández, 139–164. London: Psychology Press.

Mendelovici, A. 2013. Intentionalism About Moods. *Thought* 2: 126–136.

———. 2014. Pure Intentionalism About Moods and Emotions. In *Current Controversies in Philosophy of Mind*, ed. U. Kriegel, 113–134. London: Routledge.

Mineka, S., and K. Nugent. 1995. Mood-Congruent Memory Biases in Anxiety and Depression. In *Memory Distortion: How Minds, Brain and Societies Reconstruct the Past*, ed. D. Schacter, 173–193. Cambridge, MA: Harvard University Press.

Morag, T. 2016. *Emotion, Imagination, and the Limits of Reason*. London: Routledge.

Murray, N., H. Sujan, E.R. Hirt, and M. Sujan. 1990. The Influence of Mood on Categorization: A Cognitive Flexibility Interpretation. *Journal of Personality and Social Psychology* 59: 411–425.

Ratcliffe, M. 2008. *Feelings of Being: Phenomenology, Psychiatry and the Sense of Reality*. Oxford: Oxford University Press.

———. 2015. *Experiences of Depression: A Study in Phenomenology*. Oxford: Oxford University Press.

———. 2017. *Real Hallucinations: Psychiatric Illness, Intentionality, and the Interpersonal World*. Cambridge, MA: MIT Press.

———. 2019. Existential Feelings. In *The Routledge Handbook of the Phenomenology of Emotions*, ed. T. Szanto and H. Landweer. London: Routledge.

Salovey, P., and D. Birnbaum. 1989. Influence of Mood on Health-Related Cognitions. *Journal of Personality and Social Psychology* 57: 539–551.

Sizer, L. 2000. Towards a Computational Theory of Mood. *British Journal for the Philosophy of Science* 51: 743–769.

Slaby, J. 2008. Affective Intentionality and the Feeling Body. *Phenomenology and the Cognitive Sciences* 7: 429–444.

Slaby, J., and A. Stephan. 2008. Affective Intentionality and Self-Consciousness. *Consciousness and Cognition* 17: 506–513.

Stephan, A. 2012a. Existentielle Gefühle und Emotionen: Intentionalität und Regulierbarkeit. In *Feelings of Being Alive*, ed. J. Fingerhut and S. Marienberg, 101–121. Berlin: De Gruyter.

———. 2012b. Emotions, Existential Feelings, and Their Regulation. *Emotion Review* 4: 157–162.

Solomon, R. 1976. *The Passions*. New York: Anchor Press.

Tappolet, C. 2017. The Metaphysics of Moods. In *The Ontology of Emotions*, ed. H. Naar and F. Teroni, 169–186. Cambridge: Cambridge University Press.

Wegener, D.T., R.E. Petty, and S.S. Smith. 1995. Positive Mood Can Increase or Decrease Message Scrutiny: The Hedonic Contingency View of Mood and Message Processing. *Journal of Personality and Social Psychology* 69: 5–15.

Wong, M.Y. 2015. Towards a Theory of Mood Function. *Philosophical Psychology* 29: 179–197.

———. 2016. The Mood-Emotion Loop. *Philosophical Studies* 173: 3061–3080.

Index