THE EPISTEMIC RATIONALITY OF EMOTIONS:

A NEW DEFENCE

By Matilde Aliffi

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Philosophy, Theology and Religion

College of Arts and Law

University of Birmingham

March 2019

Arts & Humanities Research Council

UNIVERSITY OF BIRMINGHAM

Midlands4Cities Doctoral Training Partnership

THESIS ABSTRACT

Emotions and epistemic rationality have been traditionally considered to be in opposition. In the last twenty years, the role of emotions in epistemology has been increasingly acknowledged, but there is no systematic argument for the rational assessability of emotions that is compatible with both cognitivist and non-cognitivist theories of emotions and fits with the epistemic rational assessability of mental states in general. This thesis aims to fill this gap.

Using empirically informed philosophical methodology, I offer a novel account of the rational assessability of emotions that fits with the rational assessability of other mental states and that could in principle be accepted by cognitivist and some prominent non-cognitivist theories of emotions. The possibility to epistemically rationally assess emotions opens up a fresh set of questions that regards the nature of the evaluations involved in the emotions, the epistemic norms that apply to them and the extent to which we are epistemically responsible for our emotions. This thesis aims to address these questions, ultimately showing that emotions and epistemic rationality are more intertwined than we previously thought.

To my grandparents

who I remember with deep love and gratitude

ACKNOWLEDGMENTS

I will always be grateful to my lead supervisor Lisa Bortolotti for her encouragement, knowledge, trust in my ability and immense passion for philosophy. Her support has been invaluable and helped me to develop as a philosopher and beyond. I am also deeply grateful to Iain Law, my co-supervisor, for helping me to sharpen my philosophical skills. His insightful comments have been crucial for the successful production of this thesis.

I have benefitted tremendously from the support and philosophical discussions with all members of the project PERFECT in the last four years. I owe a particular debt to Sophie Stammers and Kathy Puddifoot for discussing epistemic responsibility and cognitive penetrability with me, and for their valuable advice on many issues of academic philosophy. I also owe special thanks to Valeria Motta, for her patience to discuss with me about the rationality of emotions, and for teaching me that philosophy does not always need to provide answers.

I would like to thank all members of the Philosophy Department at the University of Birmingham who have spent their precious time giving their advice on my thesis and my professional career. I am particularly grateful to Henry Taylor, Maja Spener, Herjeet Marway, Anneli Jefferson and Nick Jones for having helped me at various stages of my PhD.

I would like to acknowledge my fellow PhD students at the University of Birmingham, with whom I had the pleasure to discuss most of the ideas developed in the thesis. Helen Ryland deserves a special mention for being a friend, a source of support throughout my whole PhD journey, and also a great collaborator. Special thanks to Helen and also to Gregory Mathers, Michael Roberts, Tom Baker, Alexander Miller Tate, Will Sharp and Thomas Davies for reading previous drafts of my thesis chapters and giving useful comments.

I wish to thank Emily Walsh for teaching me how to bring kindness into philosophy and for showing me how to create new opportunities when there are none. I owe special thanks to the

Women and Non-Binary People in Philosophy group at the University of Birmingham for being a tremendous source of support at the moment when I needed it the most.

I want to acknowledge the Midlands4Cities Arts and Humanities Research Council for funding my research and for offering invaluable training opportunities throughout my PhD. This funding scheme enabled me to meet and collaborate with other fantastic PhD students at the University of Nottingham and the University of Leicester, and to attend events in the UK and overseas that contribute dramatically to improve the quality my work. I wish to thank also all staff from the University Graduate School, the Library Services and University Career Network for helping me to develop my academic skills more broadly.

I am indebted to Fabrice Teroni and Julien Deonna for their comments on whole chapters of this thesis and for their insightful discussions during my research visit of the research group *Thumos* at the Swiss Centre for Affective Science in Geneva. They challenged and enriched my ideas in a way that greatly helped to improve the quality of my work. I learnt a lot from them. I also wish to thank the other members of the research group Thumos who taught me a lot about emotions: Kris Goffin, Juliette Vazard, Magalie Schor, Michele Ombrato, Ariele Niccoli and Steven Humbert-Droz.

I wish to thank all the participants of the conferences and summer schools I attended in these four years, for their helpful comments on some of the ideas presented in this thesis. I also thank the anonymous reviewers who gave valuable feedback on previous drafts of my chapters.

I am deeply grateful to my friends who live here in Birmingham, for sharing their happiness with me and rendering my PhD journey lighter. I also wish to thank all my friends who live far away for being always there when I needed them, no matter the distance.

A special thank goes to all members of my family who always supported me during these four years. I thank my mother, for her love and her ability to look on the bright side in every circumstance. My father, for pushing me to accept the challenge of doing a PhD, and for

teaching me to look at things with creativity. To Cecilia and Giovanni, my siblings, for challenging me to do something meaningful with my study. I owe my deepest gratitude to Auro for his constant encouragement, support and for accepting all my rational and irrational emotions. I would not have been able to go through all the challenges of this PhD journey without his love.

TABLE OF CONTENTS

INTRODUCTION

I am fascinated by my emotions and their ability to contribute meaning to my life. Every day, they help me to understand what is important to me, and in which way. They direct my attention to things that are relevant to my well-being and motivate me to act in a meaningful way. They help me to understand the value of things around me. As a philosopher interested in gaining an understanding of the world, I have noticed that my intellectual musing has been imbued with emotions as well. However, the idea that a proper scientist privileges "cool" rationality over "hot" and "emotional" thinking is one that is pervasive. But this myth that we are in a better epistemic position without emotions is profoundly mistaken. In philosophy of emotions this old prejudice has been gradually dismantled, and we have come to accept that emotions play an essential part in our decision-making and our scientific practice. However, we still do not have a fully-fledged account of the epistemic rationality of emotions.

In this thesis, I will explore issues that are relevant to the current philosophical debate on the epistemic rationality of emotions. Providing a complete account of the epistemic rationality would be an enormous task that I won't be able to carry out in the limited space of this thesis. More modestly, my dissertation aims to show that rationality and emotions are more intertwined than we previously thought. The thesis is organised around the following main overarching themes:

- The evaluations that are involved in the emotions
- The epistemic rational assessability of emotions
- Epistemic rationality norms and correctness conditions of emotions
- Epistemic responsibility and control

In chapter 1, I introduce the context of my research, and I set out the fundamental problems that my thesis aims to address. I provide a brief overview of the historical context of my research and outline the methodology that I employ in this thesis.

A preliminary step for investigating the epistemic rationality of emotions is understanding the aetiology of emotion, and the evaluative information that emotions bring about. The relationship between the content of emotions and the content of causal evaluations has been overlooked, and it has often been taken for granted that the content of the emotion is identical to the content of the causal evaluation. In chapter 2, I examine the relationship between the content of the emotion and the content of the causal evaluation, and I argue that the content of emotions is not identical to the content of the causal evaluation.

Some theories of emotions claim that causal evaluations are necessary for the activation of emotions. However, emotional contagion poses a challenge to this claim. This is because the most prominent model of emotional contagion, which is the mimicry-synchrony model (Hatfield, Cacioppo, and Rapson, 1994), explains the activation of emotion without appealing to any causal evaluation. In chapter 3, I offer a solution to the challenge from emotional contagion, identifying the causal evaluations that may be involved in setting off emotions.

In chapter 4, I offer a new account of the epistemic rational assessability of emotions. I identify four key features that ground the epistemic rational assessability of mental states, and I show that most emotions meet these conditions.

The possibility to assess emotions according to rationality standards opens up new questions regarding the epistemic rationality norms of emotions. I defend a view according to which epistemic rationality norms are the norms that are conducive to correct emotions. In chapter 5, I discuss the correctness conditions of emotions. I argue that by using a relativistic account of affective evaluative properties it is possible to objectively establish whether an emotion is correct.

In chapter 6, I discuss epistemic rationality norms of emotions. In this chapter, I consider cases where emotions are epistemically irrational but deliver important epistemic benefits. I also discuss situations where emotions are epistemically rational but disrupt our epistemic

functioning. I make a case for the importance of contextual factors for offering a complete epistemic assessment of emotions.

In the remaining two chapters, I examine questions related to epistemic responsibility for our emotions. An important step for understanding the epistemic responsibility of emotions is understanding which type of control we have over our emotions. In chapter 7, I investigate whether the appraisal system that triggers the emotion is cognitively penetrable. A system is regarded as cognitively penetrable when its process can be causally affected by information stored in other cognitive systems, e.g. beliefs. Here I argue that the appraisal mechanism is diachronically penetrable but synchronically impenetrable. Understanding the extent to which cognitive states such as beliefs influence the appraisal mechanism is interesting for assessing the extent according to which we are able to regulate our emotions using rational means.

In chapter 8, I distinguish three types of control over emotional states: synchronic, diachronic and fine-tuning control. I also distinguish three different types of epistemic norms, and I present a new account of how the epistemic responsibility for our emotions relates to different types of epistemic norms and control.

CHAPTER 1: TOWARDS AN ACCOUNT OF THE EPISTEMIC RATIONAL ASSESSABILITY OF EMOTIONS

ABSTRACT

Emotions and epistemic rationality have been traditionally considered in opposition. Recently, philosophers realised that emotions may have an important role in epistemology. But what grounds the epistemic rational assessability of emotions has not been yet understood. In this chapter, I present the fundamental problems that my thesis is going to address. I will also offer a sketch of the context of my research, and I will introduce the terminology I will be using in the rest of the thesis.

INTRODUCTION

At the beginning of Shakespeare's play *King Lear*, the King asks Cordelia to express her love for him. However, his daughter turns down his request: instead of flattering him exaggerating her affection for her father, Cordelia states plainly that she loves him as her duty requires. King Lear is furious to hear this statement, and he answers as follows:

> "Let it be so; thy truth, then, be thy dower:
> For, by the sacred radiance of the sun,
> The mysteries of Hecate, and the night;
> By all the operation of the orbs
> From whom we do exist, and cease to be;
> Here I disclaim all my paternal care,
> Propinquity and property of blood,
> And as a stranger to my heart and me
> Hold thee, from this, for ever. The barbarous Scythian,
> Or he that makes his generation messes
> To gorge his appetite, shall to my bosom

Be as well neighbour'd, pitied, and relieved,
As thou my sometime daughter." (Shakespeare, 1987 [1608], 1, 1, 107-119.)

The readers of this excerpt can easily imagine that, in this scene, King Lear is experiencing an intense emotion of anger. The anger is so fierce that it motivates him to hurt his daughter disclaiming his paternal care, characterising her as "a stranger to his heart". We may want to question the rationality of his action. But we may also want to examine the epistemic status of King Lear's anger. This would be legitimate only if King Lear's anger was open to epistemic rational assessment. But is his anger open to such an assessment? And, more generally, are emotions epistemically rationally assessable? Are emotions akin to physical sensations such as itchiness and hunger, or are they on a par with belief with respect to their epistemic rational assessability status? If they are on a par with belief, we may further inquire about the epistemic status of his anger. We may try to understand how his emotion has been elicited, what were the reasons for King Lear's anger and whether he could have exerted any sort of control over his bout of anger. We could also ask whether he would be epistemically responsible for forming his emotion in this way.

Like King Lear, most of us sometimes have emotions that seem to be problematic for a variety of reasons, one of which may be tied to their epistemic status. We also seem to experience emotions that deliver us correct information about what is happening in the world and in ourselves. Fear alerts us to imminent danger: a fragile cliff, a lump in our body, or a wild boar approaching us. Anger helps us to detect what we may call a "true" slight, e.g. a sexist or racist remark. Disgust at the smell of rotten eggs may correctly inform us that the eggs may be noxious: the eggs have gone off, and they are toxic to eat.

Having emotions that inform us about the extent according to which things around us exhibit properties such as danger, offensiveness and toxicity seems to be very useful from an epistemic point of view. We may be able to form accurate judgments about the evaluative properties of

objects and events, drawing on information delivered by our emotions. But are emotions themselves open to rational assessment? Do emotions exhibit some degree of sensitivity to reasons?

In this thesis, I aim to show that emotions and epistemic rationality are more intertwined than we previously thought. I challenge the common sense view that considers emotions and rationality as in opposition. One of the key claims that will be defended in this thesis, and especially in chapter four, is that most of our emotions are open to rational assessment. However, considering emotions as being open to epistemic rational assessment opens up new complex problems: how could we generate epistemic norms? Are we epistemically responsible for our emotions? Could we control our emotions using rational reasoning? These are only some of the questions that could be asked once the epistemic rational assessability of emotion has been taken seriously. In the following chapters, I will develop an account of the epistemic rational assessability of emotions and I will address new issues arising from the possibility to rationally assess some emotions. A statement of the key problems that my research will address will be given in section 1.6 of this chapter.

In this chapter, I pave the way for my positive view of emotional evaluations and of the epistemic rationality of emotions that I will develop in the following chapters. Here I argue that we currently lack a satisfying account of the epistemic rationality of emotions that addresses the problems of cognitivist theories which coheres with the epistemic rational assessability of mental states in general.

In section 1.1, I briefly outline the historical context of my research, illustrating the genesis of what has been called the "emotional turn" in philosophy and the sciences (Salmela, 2014, p. 1) and the more recent "epistemic turn" in the philosophy of emotions. In section 1.2, I sketch my understanding of emotions, and in section 1.3 I distinguish them from other affective states. In section 1.4, I distinguish between three types of rationality: practical, procedural and

epistemic. In section 1.5, I discuss current theories of emotions. Instead of offering a comprehensive review of all theories of emotions that have been offered so far, I focus on three influential theories that are particularly interesting for the discussion that regards the epistemic rationality of emotions: the feeling theory, the cognitivist theory and the perceptual theory. In section 1.6, I will present the key problems that this thesis attempts to address. In section 1.7, I discuss the methodology that I employ in this thesis. Finally, section 1.8 discusses the significance of my research.

1.1. The Emotional Turn in Epistemology and Beyond

Emotions have been an object of philosophical interest since the origin of philosophy. Philosophers such as Plato, Seneca, Aristotle, Descartes, Hobbes, Spinoza, and Hume, to mention just a few, presented interesting and insightful theories of emotions. However, the intentionality of emotions and the importance of emotion for rationality has been neglected for centuries. Traditionally, emotions were considered as opposed to reason, they were regarded as idiosyncratic states, unable to provide us with objective information about the world. They were considered more akin to "irrational urges" than reliable sources of information about the world (Jaggar, 1989, p. 152). Emotions were relegated into the realm of everything that is subjective, partial and ineffable, [1] and thus were not considered suitable for serious investigation. As a result, the importance of emotions for epistemology and rationality has been largely disregarded.

Accounts that seem to suggest a different understanding of emotions, according to which emotions are an important source of information and perhaps of knowledge about the world were not sufficiently influential to bring about changes in perspective in the long term, which

[1] For a discussion about the myth that emotions are "ineffable", see Solomon (Solomon, 2007), chapter 10.

could attribute a new role to the emotions in epistemology. Baruch Spinoza's account of emotions presented in his "Ethics" (Spinoza, 2000), for instance, was very well ahead of its time. According to Spinoza, emotions such as joy and love helped the sage to acquire knowledge of the world. However, approaches with an emphasis on the role of emotion for epistemology did not catch on. Research in epistemology continued to focus on the rationality of belief and knowledge to the exclusion of emotions.

At the beginning of the twentieth century, however, recent developments in psychology and in phenomenology contributed to bringing fresh interest to the study of emotions. Towards the end of the nineteenth century, James published his account of emotions according to which emotions are nothing but feelings of bodily changes (I explain this more extensively in section 1.5.1). James' account was received with dismay and provoked an extended debate, especially in the second half of the twentieth century (Teroni, 2017). Some of the critics of James contributed to bringing forward a new way of understanding emotions, as mental states that could be rationally assessable.

Amongst the early critics of James that contributed to a new understanding of emotions there were the phenomenologists. Jean-Paul Sartre criticises the feeling theory on two main grounds: emotions, according to Sartre, have *intentionality* and *meaning*, that is, they are directed at objects in the world and they present a certain relation between the world and the mind (Sartre, 1962, p. 16). Emotions have an *organised character*, which does not seem to be taken well into account by the feeling theory, which conceives emotions as mere bodily feelings. Franz Brentano's and Edmund Husserl's works on the intentionality and the

phenomenology of our experience deeply influenced recent accounts of emotions (see, for instance, Brentano, 1995; Husserl, 1989).[2]

In the 1970s, there was renewed interest in the philosophy of emotions among the so-called "analytic philosophers" and cognitive psychologists, after the eclipse of behaviourism and logical positivism.[3] Mikko Salmela called this an "emotional turn" in philosophy and in the sciences (Salmela, 2014, p. 1). In this period, new cognitivist accounts of emotions, were developed both by philosophers and psychologists (see section 1.5.2).[4] In contrast with feeling theories, cognitivists argued that emotions were fundamentally different from feelings: while bodily feelings are a-rational, and rationality norms seem not to apply to them, emotions appear to be open to rationality standards.[5] Consider, for instance, a subject who fears a bug on his shoe. We may criticise this subject for experiencing fear of this bug. After all, this bug is innocuous, and the subject may also believe that it is actually not dangerous. And yet, the subject experiences fear. In this case, we may think that there is something irrational in the emotion. However, we are not prepared to do the same for bodily feelings. We do not think that there is anything irrational about itching sensations, fatigue or hunger. I will expand on this issue in 5.1. The rise of cognitivist theories brought about new interest in emotions in philosophy and psychology: emotions were gradually seen as a respectable subject to study and they began to be understood as relevant for rationality.

In the nineties, other sciences such as neuroscience and AI became interested in emotions. This resulted in the development of a new field of research, called "affective neuroscience"

---

[2] See Anthony Hatzimoysis for a presentation of the influence of phenomenological tradition for the study of emotions (Hatzimoysis, 2009).

[3] According to behaviourism, emotions are reducible to behaviour and bodily changes.

[4] See Teroni for an overview of the main contrasts between the feeling theory and its critics (Teroni, 2017).

[5] I examine this issue more closely in chapter 4.

(LeDoux, 2015; Panksepp, 1998) and "affective computing" (Picard, 1997). Empirical research about the role of emotion in decision making showed that emotions are crucial for practical rationality, suggesting that without a functioning emotional system, our decision-making capacity is dramatically impaired (Damasio, 2006, pp. 170-172). This claim generated a novel interest in the relationship between emotion and rationality. Research on emotions became even more interdisciplinary, and new accounts of the nature of emotions were put forward.

The importance of emotions for fulfilling our epistemic goals has been increasingly recognised: it has been argued that emotions have a fundamental role in epistemology as they make things salient to us, they could tell us how things impinge on our concerns and they foster our understanding of the world and our understanding of ourselves (Brady, 2013; de Sousa, 2008). Nowadays, few people would be bold enough to deny that having the ability to experience emotions may be crucial for rational decision-making.

Discussion in epistemology of emotion focussed also on the route from emotion to belief. It has been argued that emotions may justify our evaluative beliefs, such as "that cliff is dangerous" or "that comment is offensive": emotions involve evaluations of how the situation would impinge on our concerns, and this evaluative component of emotion may provide defeasible justification for evaluative beliefs (Echeverri, 2017; Pelser, 2014). The new interest in the role of emotion in epistemology gave rise to what could be called an "epistemic turn" in philosophy of emotion, the rise of novel research on the epistemic role of emotion. But could we consider emotions not only as conducive to epistemic rationality and possible sources of justification of our evaluative beliefs, but also as states that are open to epistemic rational assessment? In section 1.6, I explain why epistemic rational assessability poses a problem to existing theories of emotions. A solution for this problem will be offered in chapter 4. Before then, I need to clarify some basic issues that will help us to understand the arguments developed in this thesis. I will begin by explaining how I will understand emotions thorough this thesis.

1.2. WHAT IS AN EMOTION?

Emotions are affective states like joy, anger, sadness, guilt, fear, shame, and hope. Emotions lack a sensory organ, and so they have to draw on information from other mental states: beliefs, perceptions, expectations, imaginings, rememberings, etc. The mental states from which the emotion draw information are called "cognitive bases" (Deonna and Teroni, 2012). For instance, John's fear of the dog may be based on his perception of a dog, Mary's sadness for her rejected application is based on her belief that her application has been rejected, Anne's fear of the ghost is based on her imagining of a ghost.

Emotions, by and large, are mental states that (i) are set off by an evaluation of how the situation is impinging on our well-being; (ii) are intentional and directed to a particular object,[6] (iii) are short-lived,[7] (iv) come with a certain organised pattern of bodily changes, (vii) influence attention and motivation,[8] and (viii) have a certain phenomenology (see chapters 2 and 3 for a discussion about causal evaluations). There is general agreement regarding what central cases of emotions are, although whether and how it is possible to diverge from these central cases is still debated. I introduce part of the existing discussion on this issue in section 1.5, where I will present the main theories of emotions.

Consider the following passage, from Mary Shelley's book *Frankenstein.*

---

[6] There may be "objectless emotions", that is, emotions that are not directed at any particular object. For instance, Goldie asks us to imagine a person who walks in the middle of the night and is not able to pinpoint what she is scared about (Goldie, 2000, pp. 17-18, 143). I consider a similar example in chapter 2, section 2.3.2.

[7] Most emotions are short lived. However, there are many affective states that we would describe as emotions, despite their longer duration.

[8] The close connection between attention and motivation has been investigated by Brady (2013) de Sousa (1987) and Faucher and Tappolet (Faucher and Tappolet, 2007; Tappolet, 2016a).

> "I issued into the streets, pacing them with quick steps, as if I sought to avoid the wretch whom I feared every turning of the street would present to my view. I did not dare return to the apartment which I inhabited, but felt impelled to hurry on, although drenched by the rain which poured from a black and comfortless sky. I continued walking in this manner for some time, endeavouring by bodily exercise to ease the load that weighed upon my mind. I traversed the street without any clear conception of where I was or what I was doing. My heart palpitated in the sickness of fear, and I hurried on with irregular steps, not daring to look about me […]" (Shelley, 2012 [1818], p. 40).

In this scene, the scientist Victor has just created a living monster by assembling inanimate matter and dead limbs. He now appears deeply scared by this creature. His fear has many of the above-mentioned components: Victor's fear is about an object, in this case about the monster. This feature of the emotion is often called 'intentionality' (see chapter 2). The scientist seems to evaluate the monster as dangerous, although this is not explicitly said in the excerpt above. He undergoes bodily changes, e.g. his heart "palpitated", his attention is directed at avoiding the monster, and he is extremely alert to any cue that would suggest the monster's presence. The fear influences his motivation: he refuses to venture to his apartment, he wants to flee from the monster, and he is motivated to get away from danger, as he accelerates his pace. We can also imagine that Victor's fear has a phenomenology: there is some phenomenal feature that describes what is like to be scared by the monster in that way.

So, the components that could be involved in emotional experience are the following: evaluations, bodily changes (which may include emotional expression), a phenomenology and attentional and motivational tendencies (Scarantino and de Sousa, 2018). Emotional regulation is an additional feature that could be part of the emotional experience, according to some authors, such as James Gross (Gross, 2014). Victor seems also to engage in "emotional regulation" (see chapter 8 for a discussion of emotional regulation). He tries to control his emotion, reducing its intensity by doing "bodily exercise". Emotion researchers disagree on which of these components is necessary and on which component (or components) is sufficient

for triggering and/or constituting an emotion. These problems have been called respectively the "problem of parts" and the "problem of plenty" (Prinz, 2004). I present three influential theories of emotions in section 1.5.

In this thesis, I do not attempt to offer a full answer to these two problems. My ambition is to develop an account of the epistemic rational assessability of emotions that does not rely on controversial assumptions on the nature of emotion and that could potentially be adopted by different emotion theories, without needing to drastically revise the cornerstones of these theories.

1.3. Emotions and Other Affective States

Emotions, but also moods, character traits, and affective dispositions are affective states. Like emotions, all affective states are related to what matters to us: they are characterised by a "lack of indifference", as Giovanna Colombetti puts it (Colombetti, 2014, p. 1). They "bestow meaning to the circumstances of our lives" (Solomon, 1977, p. 70). The emotion of fear, for instance, presents things as dangerous to us; the mood of irritability alerts us to perceive offences; love, understood as emotional disposition,[9] for instance, disposes us to entertain different emotions, such as to be happy when the person we love achieves their goal, sad when they are not around, and scared when their well-being is put at risk.

The English language is often ambiguous and does not clearly discriminate between affective terms that refer to emotions, character traits, emotion dispositions and moods. In my thesis, I will only focus on emotions. Character traits are dispositions to entertain a certain pattern of emotions and thoughts of a certain kind (Goldie, 2000). For instance, a resentful

[9] I am not providing an account of love here, and I do acknowledge that explaining love is a controversial issue. Here I just want to consider an example of a "loving" disposition. I do not intend to claim that what we call "love" can be reduced to a disposition to experience different emotions.

person is a person who gets offended easily, and who tends to indulge in thoughts about received offences, and who will tend to be very alert to slights. In contrast, emotions are not dispositions of our character, but occurring states. The occurring aspect of emotions distinguishes them from emotional dispositions.

Consider, now, moods like gloominess, joyfulness, (non-clinical) depression, irritability and apprehension. It is still controversial what the main differences between emotions and moods are. Some scholars such as Peter Goldie (Goldie, 2000, p. 17) consider moods as differing from emotion only in degree (moods have less specific objects, last longer and are less intense), other authors think that moods and emotions belong to different psychological categories: for Solomon, for instance, moods lack particular objects and focus on the situation as a whole (Solomon, 1977, p. 71), for Carolyn Price, emotions and moods differ in their functions (Price, 2006), for Eric Lormand, moods are dispositions, whereas emotions are occurring (Lormand, 1985). In this thesis, I will mainly focus on emotions and, unless I will stipulate otherwise, the emotion terms I will be using should be taken to refer to emotions and not to other types of affective states.

1.4. "RATIONALITIES" OF EMOTIONS

The use of the word "rationality" in philosophy and in our everyday discourse varies greatly. In this thesis, I will only discuss the rationality *of emotions*, leaving open whether the rationality or irrationality of emotions has implications for the rational status of the agent that has the emotion. I believe that providing an account of the agent's epistemic rationality is a very interesting question, but I do not address this specific issue in this thesis.

In ordinary talk, we often say that our emotions are "irrational". In general, when we consider an emotion as irrational, we are implying that emotions are irrational because they are flouting a rationality norm. However, we are often unclear about what particular norms we are considering. Depending on the focus of the norm, we can distinguish between three types of

rationality: practical rationality, epistemic rationality and procedural rationality (Bortolotti, 2010). The focus of practical rationality is the mental state's relation to decision-making, whereas the focus of epistemic rationality is the mental state's relation to the evidence and the focus of procedural rationality is logical reasoning and regards how mental states are integrated together.

The norms generated by these three types of rationality may give different verdicts about the rationality of particular emotion occurrences. Thus, it is important to distinguish these three types of norm to deliver a clear assessment of the rationality of emotions. In chapter 8, for instance, I will discuss cases in which an emotion occurrence is epistemically irrational but is practically beneficial, and cases where it is epistemically rational but practically problematic. In these cases, the emotions that are epistemically irrational may not involve a failure of practical rationality. Similarly, emotions that are epistemically rational do not always appear to be also practically rational.

Let us consider how these three types of rationality norms could be applied to beliefs, and we will then see how we could apply each to emotion.

1.4.1. Rationality Norms of Beliefs

Consider the following practical rationality norm of belief.

- What one believes ought to be in accordance with one's prudential reasons.

Suppose that you feel anxious about driving long distances. However, you know that if you believe you are an excellent driver, you will feel relaxed when driving rather than anxious. Suppose that your job requires driving long distances every day. You may have practical reasons to believe that you are an exceptionally good driver. If you do not believe yourself to be an excellent driver, in spite of having the prudential reason to hold the belief, you fail to comply with the above practical rationality norm.

Epistemic rationality norms of mental states focus on the state's relation to the evidence. An example of an epistemic rationality norm is the following:

- What one believes ought to be in accordance with the available evidence.

Suppose that you have evidence that you are a terrible driver. You know that you often drive tired and above the speed limit. You have received several tickets and you have been involved in a few minor car accidents. If you believe yourself to be an excellent driver you fail to comply with the above epistemic rationality norm.

Procedural rationality norms of mental states prescribe what mental state a procedurally rational subject ought to have. The focus of rationality norms is logical reasoning. An example of a procedural rationality norm is the following:

- What one believes ought to be consistent with other beliefs.

Suppose that you believe that you are a good driver and that you also believe that you are a bad driver. If you do so, you fail to comply with the above-mentioned procedural rationality norm, as you are having contradictory beliefs.

The focus of this thesis is only on *epistemic* rationality norms. Although it is not always possible to divorce questions about procedural and epistemic rationality, it is often possible to do so. Not all epistemic rationality failures are procedural failures. Consider a subject who believes that Barack Obama was not born in the US. One day, in the newspaper, she sees a picture of Obama's official US birth certificate. Instead of revising her belief that Barack Obama was not born in the US, she revises her belief that the picture of the certificate published in the newspaper is authentic. She comes to believe that Obama was not born in the US and that his birth certificate is fake. In this scenario, the subject is complying with procedural rationality norms. The subject does not simultaneously believe that Barack Obama was born in the US and that he was not born in the US. However, the subject does not comply with epistemic rationality norms because her belief is maintained against reliable evidence.

Similarly, it is possible to find situations in which there are no failures of epistemic rationality, but there are failures of procedural rationality. Consider Ann, who believes that if it is raining outside her house, then the pavement will be wet. One day, Ann sees that the pavement outside her house is wet. From her belief and the evidence, she infers that it is raining outside her house. Her inferred belief is fallacious: Ann is affirming the consequent. Ann's pavement being wet does not entail that it is raining on Ann's pavement. Ann's pavement could be wet for many other reasons: there may be a leakage in the garden, or somebody could have washed the pavement. In this case, her belief is violating procedural rationality norms, but not epistemic norms. Her belief is fallacious independently of considerations regarding the evidence. In the following section, I explain how the distinction between the three types of rationality could be applied to emotions.

1.4.2. Rationality Norms of Emotions

Consider one possible practical norm for emotions:

- What emotion one experiences ought to be in accordance with one's prudential reasons.

Theseus wants his daughter Hermia to marry Demetrius, although she strongly opposes the marriage because she is in love with Lysander. Theseus threatens to kill her if she refuses. Ethical issues aside, we would say that Hermia, in this case, may have prudential reason to feel love for Demetrius.

An example of a procedural rationality norm may be the following:

- It is not permitted to form contradictory emotions at the same time directed at the same object.

Suppose that Juliet meets Romeo and falls in love with him. However, she discovers that Romeo belongs to a rival family and that he has just killed her cousin. Juliet loves and hates Romeo at the same time. Juliet, in this scenario, is experiencing ambivalent emotions. Ambivalent emotions are two co-occurrent contrasting emotions directed at the same object

(Pugmire, 2005). Philosophers of emotions have been discussing whether ambivalent emotions such as Juliet's involve irrationality. The discussion, in this case, is focussed on whether ambivalent emotions may involve a failure of *procedural rationality*. Whether ambivalent emotions fail to respect the above rationality norm is subject to an interesting debate, that I will not present here (Calabi, 2013; Greenspan, 1980; Kristjánsson, 2010; Price, 2015; Tappolet, 2005). I will expand on the rational status of ambivalent emotion in chapter 3. What matters here is that the philosophical discussion on ambivalent emotions focusses on whether ambivalent emotions do involve *procedural* rationality.

Consider the following epistemic norm:

- It is not permissible to form an emotion that is not formed in accordance with the evidence.

Suppose that a messenger informed Cleopatra that Antonio married Octavia. Cleopatra is furious. Although she knows that the messenger is not responsible for the marriage, she still feels terribly angry with him. In a bout of anger, she beats him. In this scenario, Cleopatra is experiencing a recalcitrant emotion: she knows that the messenger is not responsible for the bad news, nevertheless, she is angry with him. Recalcitrant emotions are emotions that persist despite our contrasting beliefs (I will discuss recalcitrant emotions more extensively in chapter 7). The contrast between emotion and belief could be seen as another case in which there is a violation of procedural rationality. However, recalcitrant emotions also seem to involve an infringement of epistemic rationality: they *persist* against the subject's better evidence. Cleopatra knows that the messenger is not responsible for the bad news, but she directs her anger against the messenger, despite her evidence that he is not responsible for it.

In this thesis, I will mainly focus on epistemic rationality: I will discuss the epistemic rational assessability of emotions in chapter 4, I will present how epistemic rationality norms could be generated in chapter 6, and I will discuss how epistemic responsibility relates to epistemic norms of emotions in chapter 8. The procedural rationality of emotions will be only

tangentially discussed as applied to ambivalent emotions in chapter 2. Contrasts between epistemic rationality and practical rationality norms will be examined in chapter 8.

1.5. THEORIES OF EMOTIONS

In the last fifty years, there have been so many accounts of emotions that it is very difficult for me to do justice to them in the space of this section. Instead of offering a survey of all existing theories, the aim of this section is to present three influential theories of emotions that shaped the existing debate on the rationality of emotions. I will organise this overview of the following issues:

- The feeling theories and the a-rationality of emotions
- Cognitivists and evaluative judgments
- Perceptual theories and the non-inferentiality of emotions

The aim of this section is to introduce the main obstacles that these theories have faced, in relation to explaining the rationality of emotions.

In chapter 7, I present and discuss another non-cognitivist theory of emotion, that is the psychoevolutionary theory. As I will argue in chapter 7, the psychoevolutionary theory considers emotions such as joy, fear, sadness, anger and disgust as a complex and coordinated pattern of bodily changes that are the product of an informationally encapsulated system. However, I will not discuss such a theory here. This is because a full explanation of such a theory will be developed in chapter 7, and because the primary aim of this section is to show that both cognitivist and non-cognitivist theories face a dilemma regarding the epistemic rational assessability of emotions. Since the problems that the psychoevolutionary theories face in explaining the epistemic rational assessability of emotions are similar to the ones that the

perceptual theory faces, I will not discuss them here. The aim of this section is ultimately to guide the reader through the key approaches to the rationality of emotions, highlighting their limitations. I will then present the problem of the epistemic rationality of emotions in the following section.

1.5.1. Feeling Theories and the A-Rationality of Emotions

When we have an emotion occurrence, we usually perceive bodily changes. For instance, we may feel a different respiratory pattern, an increase of our heart-rate, a lump in our throat etc. We may also feel a different engagement of our facial muscles, or we may adopt a different posture. Internal physiological changes and facial expressions seem to be commonly experienced as important elements of our emotional experiences, at least when experiencing emotions such as fear, anger, sadness, surprise, joy, and disgust.

However, the feeling theory goes far beyond stating the common sense claim that emotions are often accompanied by a perception of bodily changes: it *reduces* emotions to mere perceptions of bodily changes. According to the feeling theory, feelings of bodily changes are essential to emotion. Therefore, disembodied emotions[10] are simply not logically possible on this account.[11] In addition to this, for the feeling theory, feelings of bodily changes are all there is to an emotional experience. James supports this thesis with a mental experiment. He asks us to consider a situation in which we experience an emotion. According to James, if we imagine

[10] There are emotions that seem to not require bodily changes. The so-called "intellectual emotions" such as interest and wonder may not require bodily changes and yet are best accounted for as emotions.

[11] Existing debate on unconscious emotion and robots' emotions may also question the claim that feelings are a sine qua non condition for emotions. Even if it was possible to have emotions without feeling of bodily changes, the major claims of this thesis still hold.

subtracting from that emotion our perception of bodily changes, nothing remains (James, 1884, p. 193).

Another aspect where the feeling theory defies common sense regards the process that triggers emotions. Common sense tends to consider the perception of bodily changes as the effect of emotion: first comes the stimulus, for instance Victor's perception of a monstrous creature, then the emotion of fear, and finally the perception of bodily changes. On the contrary, for James' theory of emotion, Victor's perception of the monster would immediately trigger bodily changes, which would then cause the emotions.

Now, the revisionist character of James' account of emotion is not necessarily a problem for his account. We may simply accept that science is revisionist and accept these claims. One more pressing worry may regard the empirical adequacy of the account. If people with paralysis with a dramatically impaired ability to have bodily changes, were able to experience emotions, the feeling theory would be disproved. Empirical evidence seems to suggest that this is the case, although this issue is still quite controversial (Prinz, 2004, pp. 57-58). Could feelings of bodily changes be triggered without bodily changes, according to the feeling theory? Although this is a possibility that James' account cannot accommodate, neo-Jamesian accounts such as Damasio's somatic theory allow for this possibility. As in James' theory, for Damasio, emotions are essentially bodily feelings that signal the relevance of the situation for the subject's well-being. However, emotions can be triggered not only by bodily changes but also by a special mechanism, that in the absence of bodily changes activates the somatic feeling "*as if*" there were full-blown bodily changes (Damasio, 2006, p. 155).

Feeling theories[12] understand emotions as feelings. As I said in section 1.1, we are not prepared to claim that sensations such as feeling nauseous, feeling "butterflies in the stomach", or itching are rational or irrational *per se*. This is not to say that the feeling theory lacks theoretical tools to distinguish between functional and dysfunctional emotions. Employing a feeling theory framework, we could certainly distinguish between more pleasant sensations and more painful ones. Under the same framework, we could also argue that emotions are more or less conducive to rationality. Damasio, for instance, famously claimed that emotions are crucial for decision making in that they help us to restrict the number of options to consider when deliberating (Damasio, 2006). These strategies may succeed in emphasising the role of emotion in epistemology: they may show that emotions are a *sine qua non* condition for rationality. However, the feeling theory does not consider emotions *themselves* as open to rational assessment. The emotion itself would not be regarded as rational or irrational.

Is this a satisfactory account of emotions and a compelling view of the emotions' rationality? Many criticisms have been raised against the feeling theory. Here I will discuss what, in my opinion, are the two major difficulties of the account. They are that it fails to satisfactorily explain (1) the intentionality of emotion, and (2) the rationality of emotions.

When we experience emotion, such as an emotion of fear, anger or disgust, our emotion is not a mere feeling but seems to be *about* something (Solomon, 1977, p. 112). This feature is called intentionality. We say things like "Victor is scared of the monster", "Achilles is angry with Agamemnon", and "Grete is disgusted at the insect" (see also chapter 2 for a discussion of the intentionality of emotions). This feature of emotions seems to be not properly explained by an account that conceptualises emotions on a par with feelings. It is unclear whether feelings

[12] In the remaining part of this section, I will be using this term to refer both to James' feeling theory and to Damasio's somatic theory.

are *about* something. There may be a sense in which even feelings are intentional: for instance, the feeling of shivering could be directed at something, namely shivering. However, consider typical situations in which you felt angry, sad or joyful. The emotion you experienced in these situations, usually, does not seem to be directed at bodily changes but at objects in the world. In the example mentioned in section 1.3, for instance, Victor is scared of the monster, and not of some pattern of bodily changes he is experiencing.

The theory defies our ordinary practice of considering emotions open to rational assessment (Solomon, 1977, p. 101). As Solomon points out, in our ordinary talk, we often say that emotions are "rational", or "irrational", "reasonable" or "unreasonable". Suppose, for instance, that Victor is not scared by a giant creature that has the power to kill him, but by an innocuous earthworm. Suppose that he knows earthworms are not dangerous to him, but nevertheless he feels fear. We would call such fear irrational. And indeed we usually do regard emotions like phobias "irrational", and so does psychology according the Diagnostic and Statistical Manual (APA, 2013). We often talk about "unreasonable" love, "reasonable anger", "rational sadness". It is hard to explain this in the feeling theory framework. As I have said earlier, the feeling theory cannot easily explain why rationality standards apply to emotions, as emotions are mere feelings, and feelings are not states we would consider open to epistemic rational assessment.

1.5.2. Cognitivism and Evaluative Judgments

The central claim of cognitivists is that emotions are evaluative judgments (Nussbaum, 2001, 2004; Pitcher, 1965; Solomon, 1988, 2007; Taylor, 1975).[13] Like the feeling theory,

[13] Here I will focus on classic versions of cognitivist theories. Other versions of cognitivist theory were put forward to respond to the challenge from recalcitrant emotions, which I present towards the end of this section. These theories weakened the claim that emotions are evaluative judgments to accommodate for the possibility of

cognitivism adopts a reductive strategy. In this case, however, emotions are not reduced to feelings of bodily changes, but to evaluative judgments. Thus, genuine emotions, according to cognitivists, require entertaining an evaluative judgment featuring the same object. For instance, Victor is genuinely scared by the monster only if he judges it to be dangerous. This is because his fear of the monster can be reduced to this: the judgment that the monster is dangerous.

Cognitivists observe that emotions are states for which we may have or lack reasons. For instance, suppose that Mary is craving a chocolate bar, which she bought and left in her drawer. Suppose that she looks in the drawer, with the expectation to find the sweet. However, there she sees only the wrapping without the chocolate. Imagine, now, that she feels angry with her sister, who, in her opinion, ate her chocolate bar. But when her brother tells her that he is the one responsible for eating the bar, Mary ceases to be angry with her sister (and perhaps gets angry with her brother). According to cognitivists, cases like this seem to show that emotions respond to reasons. And this, according to cognitivists, is a feature that is typical of judgments. The alleged rationality of emotions is also easy to explain within this framework: emotions are judgments, and since judgments are open to rational assessment, emotions are also open to epistemic rational assessment.

However, cognitivist theories face serious problems. In what follows I will focus on what I take to be the main ones, which regard (i) explaining recalcitrant emotions and (ii) explaining why new-born children and non-human animals have emotions.

Recalcitrant emotions are emotions that conflict with the subject's evaluative beliefs and/or judgments. Imagine, for instance, that Sarah knows that the earthworm on her shoe is not

---

recalcitrant emotions. Patricia Greenspan, for instance, identified emotions with thoughts (Greenspan, 1988), and Robert Roberts with "evaluative construals" (Roberts, 2003).

dangerous. Nevertheless, she is scared by it. Sarah's fear is recalcitrant because it persists despite her better belief. This persistence is difficult to explain with a cognitivist framework. According to judgmentalists, emotions are evaluative judgments. Therefore, the contrast involved in experiences of recalcitrant emotions is a contrast among the subject's evaluative judgments. In Sarah's case, this would be a contrast between the belief that this earthworm is dangerous and the belief that the earthworm is not dangerous. According to judgmentalism, therefore, recalcitrant emotions involve a contradiction. However, this is controversial because they do not seem to involve a contradiction akin to the contradiction between two evaluative judgments. Sarah does not seem to hold two contradictory beliefs. Nor does it seem that the subject is endorsing a full-blown *belief* that persists despite her better evidence of the contrary. After all, she claims that the earthworm is not dangerous! Moreover, also the elicitation of the recalcitrant emotion is difficult to explain, given that the subject holds judgments that are contrary to their own emotions. And indeed, the theory may also be taken to suggest that emotional recalcitrance is not even possible (D'Arms and Jacobson, 2003). In the above-mentioned example, Sarah for instance, believes that the earthworm she sees is not dangerous. Why, then, is she afraid of it? To be afraid of the earthworm, she needs to judge it to be dangerous. But we just said that she judges the earthworm *not to be* dangerous! It seems really difficult to explain recalcitrant emotions within a cognitivist framework without attributing to the agent a contradictory belief.

Cognitivists do not only struggle to explain why recalcitrant emotions are possible, but also to explain the irrationality involved in recalcitrant emotions, since, by identifying emotion with judgments, they impute too much irrationality to the subject (Brady, 2009). In chapter 7, I will come back to this issue and I will discuss in more detail how to explain the possibility of recalcitrant emotions. For the sake of this introductory chapter, what is important is to highlight that recalcitrant emotions pose a challenge to cognitivist accounts.

A second important problem faced by cognitivists regards the possibility that non-human animals, and new-born babies have emotions. Cognitivists reduce emotions to evaluative judgments; therefore, evaluative judgments are a necessary condition for having an emotion. However, this requirement seems over demanding. Infants and animals seem to not yet have the ability to perform proper judgments: they do not have the ability to master a language. However, they do experience emotions, as they display the characteristic bodily changes typical of emotions.[14] So, cognitivists seem to make the wrong prediction: they predict that infants and animals are not able to have proper emotions, and this contrasts with our intuition that they do, as their expressions suggest.

Moreover, empirical research suggests that emotions could be set off in absence of the relevant evaluative judgment. This is the case not only in newborns and animals, but also in some instances of emotions experienced by human adults (Zajonc, 1984). Robert Zajonc used empirical evidence to show that emotions can be induced without judgments drawing on evidence from the neural structures of the brain. He observed that there is a neural pathway from the retina to the hypothalamus. For Zajonc, this data suggests that emotions can be triggered directly from sensory input, without any cognitive process involved.[15] More recently Scarantino reviews empirical data that suggests that emotions such as fear can be triggered in the absence of awareness of its eliciting stimulus (Scarantino, 2010). Cognitivists may try to respond to this evidence by adopting an "elastic strategy", which consists of "stretching the

[14] Charles Darwin collected data that suggests a continuity between human and animal's ability to entertain emotions (Darwin, 1872).

[15] It might be argued, however, that this evidence is not conclusive against the cognitive theory. Zajonic's argument presupposes that neural data give us conclusive evidence about whether cognitive activity is involved in a mental process. However, this could be disputed on the ground that there is not a token identity between mental states and brain areas (Putnam, 1967).

notion of judgments to accommodate counterexample" (Scarantino, 2010, p. 730). And indeed "elastic strategies" have been used to respond to the challenge from recalcitrant emotions. Patricia Greenspan modified the cognitivist theory arguing that emotions are thoughts, Robert argued that emotions are evaluative "construals" (Roberts, 2003) and Bennett Helm that emotions are a kind of assent (Helm, 2001). However, as Scarantino argues, these modifications of the cognitivist account seem to be methodologically problematic. This is because they render the cognitivist thesis either trivially true or difficult to falsify (Scarantino, 2010).

1.5.3. Perceptual Theories and the Recalcitrant Emotions

As we have seen in the previous subsections, both feeling theories and cognitivists face pressing problems. Perceptual theories put forward an account of emotion that is able to retain the key claim of the feeling theory, namely that emotions are necessarily embodied and that evaluative judgments are not necessary for setting off emotions, while, at the same time, accepting the cognitivist claim that emotions are evaluations of how particular objects impinge on the subject's well-being.

The perceptual theory of emotion argues that emotions are perceptions of values (see, for instance de Sousa, 1987; Döring, 2007; Prinz, 2004; Tappolet, 2012). According to perceptual accounts, emotions have a representational content: they represent particular objects as having affective evaluative properties. This theory draws on Anthony Kenny's notion of "formal object" (Kenny, 2013). However, the representational content, according to perceptual theories, is non-conceptual (Tappolet, 2016b, pp. 15-16). I do not have the space here to fully discuss the difference between conceptual and non-conceptual content of emotion. I will expand on such differences between conceptual and non-conceptual content in chapter 2.

Some perceptual theorists take seriously the identification between emotions and perceptions, claiming that emotions are *literally* perceptions (e.g. Döring, 2009; Prinz, 2004).

Other philosophers think that emotions are not literally perception, but they share important features with perceptions (e.g. de Sousa, 1987).[16] The main argument for the perceptual theory consists in highlighting that the key features that characterise perceptions are also shared by emotions. The similarities between emotions and perceptions include (i) having a phenomenology, (ii) being triggered automatically, (iii) being caused by events in the world, (iv) having correctness conditions, (v) allowing for a conflict with judgments, (vi) being informationally encapsulated,[17] (vii) being inferentially isolated, (viii) having an "analogic" content (Tappolet, 2016b). On the basis of such analogies, the perceptual theory argues that emotions are perceptions.

Critics of the perceptual theory may try to show that some of these analogies do not hold. In this section, however, I will not review existing criticisms to the perceptual theory. I will rather discuss whether this theory can account for the rational assessability of emotions.

Some perceptual accounts of emotions, e.g. (Döring, 2009; Prinz, 2004), do not consider emotions as open to rational assessment. According to this view, emotions, like perceptions, are *a-rational* mental states. This does not prevent them from arguing that emotions are relevant for rationality. Indeed, Sabrine Döring and Catherine Elgin developed an account of justification of evaluative beliefs that could be considered "dogmatist", e.g., (Pryor, 2000), according to which emotions could provide a prima facie justification for evaluative belief (Döring, 2007; Elgin, 2008). Just as my perception of a red cube in front of me provides a defeasible reason for the belief that there is a red cube in front of me, my fear of the dog provides a defeasible reason for the evaluative belief that the dog is dangerous.

---

[16] See Brady for a careful presentation of the distinction between the two kinds of perceptual models (Brady, 2013).

[17] See chapter 7 for an explanation of the notion of informational encapsulation.

Other perceptual theories have tried to accommodate the rational assessabilty of emotion. The debate regarding the rational assessability of emotions focusses mainly on recalcitrant emotions, taken to be a case of allegedly irrational emotion that calls for an explanation.

One strategy to understand the epistemic rational assessability of emotions focusses on defending the a-rationality of emotions on the grounds that emotions are non-inferential. I present this account in the following section.

1.5.4. Rational Assessability and Non-Inferentiality

According to Döring, emotions are not open to rational assessment because they are not subject to inferential relations (Döring, 2009). She argues that emotions do not stand in inferential relations with other mental states, since (i) we do not infer emotions from beliefs or other perceptions, and (ii) emotions do not need to be revised in light of other perceptions and beliefs.

Döring argues that if two properties cannot be perceived at the same time, you cannot infer the perception of one property from the other. For instance, suppose that you see the vertical line of the letter *i*. Then, you look above, and you see the dot on the top of the vertical line. According to Döring, if you see only the vertical line (and not the dot), you cannot infer the perception of the dot from the vertical line. Similarly, she says, we cannot infer an emotion from another emotion. Therefore, she concludes, emotions do not stand in inferential relations with other mental states.

Her second argument for the non-inferentiality of emotion draws on Tim Crane's work on the non-inferentiality of perception. Following Crane, Döring argues that when we are subject to a visual illusion we are not really falling into contradiction, since perception is not subject to inferential constraints (Crane, 1992). Consider, for instance, a subject who has the "waterfall illusion" (Crane, 1988). The waterfall illusion happens when we observe a waterfall for some seconds, and then we attend to a still object nearby. When we do this, we see the observed object as moving against a static background. The waterfall illusion seems to be contradictory,

since it presents an object as moving, while at the same time as being still. The waterfall illusion is also in contrast with our belief: we know that the stone is not moving, and despite this we see it as moving. However, for Crane, it is not the case that we (rationally) *ought* to revise our perception in light of the evidence. Therefore, the best explanation for the claim that it is not the case that we ought to revise our perception is that perceptions do not stand in inferential relations as beliefs and judgments do.

Similarly, Döring argues, emotions are not subject to inferential relations because it does not seem that we ought to revise them. While it is incoherent for a subject to judge a proposition *p* and to believe that not *p*, it is not incoherent to experience emotions that contrast with the subject's belief. For Döring, while it is incoherent to say "it is Monday and I believe it is not Monday", it is not incoherent for S to assert "I am not in danger, but I feel fear". Therefore, she concludes, inferential constraints do not apply to emotions. From this account, it follows that emotions are not open to rational assessment.

We may accept that emotions are not inferentially related, meaning that we are not able to infer emotions from other emotions or mental states. However, it is unclear whether this tells us something about the emotions' normative requirements. And even if this was the case, Döring's second claim seems controversial. She argues that there is an analogy between emotions and perception regarding whether we ought to revise our perception/emotion when it conflicts with our belief. However, as acknowledged by other proponents of perceptual theories, e.g. (Tappolet, 2016c), the intuition that we do not need to revise recalcitrant emotions seems quite controversial, since many have the intuition that recalcitrant emotions involve some sort of irrationality (e.g., Brady, 2009; Tappolet, 2012).

Döring's account, however, it's explicitly a revisionary account, and therefore it is to be expected that her account is controversial. We cannot reject Döring's argument only because it leads us to revise our intuitions.

In section 1.4, I distinguished between three different types of rationality: practical, epistemic and procedural. Since the main focus of this thesis is the epistemic rationality of emotions, we could wonder whether Döring's non-inferentiality argument successfully targets the thesis that emotions are epistemically rationally assessable. Döring does not specify which rationality norms are the main target of her argument. Consider again Döring's argument. Döring argues that it is incoherent to believe *p* and not *p* but it is not incoherent to fear something, say a worm, and judge it to be safe. This is because, according to Döring, beliefs are inferentially related, but emotions are not. Thus, she claims, it is not contradictory to experience emotion that conflict with belief. There is reason to think that her argument challenges the procedural rational assessability of emotions, and not the thesis of epistemic rational assessability of emotions. For the sake of the argument, let us suppose that Döring is right in claiming that emotions are not inferentially related with other mental states and therefore should not abide by logical norms such as the law of non-contradiction. If emotions are not inferentially related to other mental states, does the non-inferentiality of emotions support the *epistemic a-rationality* of emotions? Inferential relations between mental states are regulated by procedural rationality norms, and not by epistemic rationality norms. As I said in section 1.4, epistemic rationality norms focus on the mental state's relation to the evidence, whereas procedural rationality norms focus on the (logical) relations between mental states. Döring's argument seems more convincing as an argument against the procedural rational assessability of emotions, rather than an argument against the epistemic rational assessability of emotions.

This interpretation is consistent with her account of rationality developed in another paper in which she argues that the rationality she is interested in does not consist in responding correctly to reasons but in satisfying norms of coherence between mental states (Döring, 2015, p. 387). Since the former, but not the latter falls within the domain of epistemic rationality, we

can conclude that Döring's argument targets the *procedural* rational assessability of emotions, and not their epistemic rational assessability.

The non-inferentiality of emotions may be taken to support the claim that emotions are not subject to logical norms.[18] What here I want to suggest is that this argument, as it is, does not establish that emotions are not epistemically rationally assessable.

1.6. Are Emotions Epistemically Rationally Assessable?

In the previous section, I showed that the rational assessability of emotions raises a dilemma. It is difficult to offer a satisfying explanation that can account for the possibility of having recalcitrant emotions, and, at the same time explain their rational assessability. The rational assessability of emotions can be easily explained under cognitivist accounts (see section 1.5.2), according to which the irrationality of emotion is explained by appealing to the irrationality of a belief that is a necessary component of the emotional experience. But cognitivist accounts face difficulties when explaining the persistence of recalcitrant emotions, namely emotions that conflict with our evaluative judgment (D'Arms and Jacobson 2003). This raises a problem: either the subject is irrational because she has two contradictory beliefs, or recalcitrant emotions seem impossible.

[18] And indeed, the non-inferentiality of emotions, for instance, has been used to show that ambivalent emotions are not subject to procedural rationality norms also by other authors. Patricia Greenspan, for instance, discussing the rationality of ambivalent emotions argued that emotions do not follow the logic of judgments (Greenspan, 1988). Ambivalent emotions are two *conflicting co-occurring* emotions that are directed at the same object (Kristjánsson, 2010). Bennet Helm argued that emotions are not subject to procedural rationality norms such as norms of coherence (Helm, 2001).

Non-cognitivist accounts are better able to explain why recalcitrant emotions persist but struggle to explain the rational assessability of emotions (Helm 2001 p. 42; Tappolet 2016 p. 31). What grounds the epistemic rationality of emotions has been discussed especially in relation to recalcitrant emotions (see, for instance D'Arms and Jacobson 2003; Döring 2015; Brady 2009). The debate regarding the epistemic rational assessability of emotions has focussed on contrasting emotion with perception. Perception has been taken by philosophers contributing to this debate to be a type of mental state that is a-rational. The discussion thus has focussed on whether emotions are *like* perceptions, in that they are also a-rational, or whether they are different from perceptions, in that emotions, but not perceptions are epistemically rationally assessable. However, there is no systematic argument for the epistemic rational assessability of emotions that is compatible with both cognitivist and non-cognitivist theories of emotions and fits with the epistemic rational assessability of mental states in general.

In this thesis, I aim to fill this gap. I will adopt a new strategy for identifying the epistemic rational assessability of emotions. I will not investigate whether emotions, *as types*, are epistemically rationally assessable. I take the epistemic rational assessability to be a property that is not intrinsic to the psychological type, but it is intrinsic to the particular mental state's *occurrence*. In chapter 4, I identify the conditions for the epistemic rational assessability that apply to *occurrences* of mental states. In this way, the epistemic rational assessability of emotions will fit with the epistemic rational assessability of other mental states.

Considering emotions as mental states that are epistemically rationally assessable generates a new set of problems. Some of these issues regard the aetiology of emotion. Is the emotion-triggering system open to influences from beliefs? Are emotions always caused by causal evaluations, and what is their relationship with the content of emotions? Other issues may regard the epistemic norms: how are epistemic norms generated? What is the relationship

between epistemic norms and the correctness conditions of emotions? Finally, new problems may regard the extent according to which we are able to control our emotions, and what this could tell us about the epistemic responsibility we may enjoy for our emotions. This thesis attempts to shed light on these issues.

1.7. METHODOLOGY

In my research, I will use an empirically informed approach to the study of emotions. An important part of my thesis regards understanding why and how emotions are triggered (this issue will be discussed especially in chapters 3 and 4) and what type of control we have over our emotions (chapter 8). These are empirical questions, which cannot be entirely settled from the armchair. Theoretical conclusions need to be grounded in empirical research.

According to my empirically informed methodology, I will draw on empirical data and scientific studies from empirical sciences (mainly psychology) without conducting experiments myself. When possible, I will rely on robust empirical data, which is widely accepted within its own field. In my view, for this project, the value of an empirically informed approach consists mainly in two things: (i) it can constrain theories of emotions, helping us to disprove theories and (ii) it can help us to go (partly) beyond introspection[19] employing

[19] However, many fMRI and PET experiments rely on first-person data, such as verbal report of the subject's emotional or perceptual state (Gallagher et al, 2008).

imaging techniques, such as fMRI[20] and PET.[21] This does not mean that introspection or first-person data is never useful for analysing what emotions are: for instance, it could be useful for understanding the features of our emotional experience. And indeed, I will also rely on observation when discussing features of emotional experience. However, sometimes, relying on the use of our first-person data about our own emotions can be flawed, or not fine-grained enough to capture the phenomena.[22] In this thesis, I will mainly focus on recent theories of emotions, since they describe emotions with a greater level of detail and strive to be consistent with the forefront of empirical research about emotions.

The epistemic rationality of emotions cannot be established by merely looking at science. Rationality is a normative concept. As David Hume notoriously remarked, a set of purely descriptive premises does not entail any normative conclusion (Hume, 2000 [1738]). I will, therefore, complement this approach with a more traditional a-priori methodology. In chapter 2, when I will be discussing the evaluative content of emotions, I will largely use conceptual analysis and intuition-pump arguments. In chapter 4, when I will be identifying the features that matter for the epistemic rational assessability of the emotions, I will use a reflective equilibrium methodology to identify a theory of epistemic rational assessability that accounts for our intuitions. This method was developed by John Rawls and Nelson Goodman and was

[20] fMRI or functional Magnetic Resonance Imaging is a visualisation technique that maps neuronal activation by using the magnetic induction caused by the change in the blood flow. For an introduction of fMRI see Ugurbil (Ugurbil, 1999).

[21] PET or Positron Emission tomography is an imaging technique that visualises data about the changes in the blood flow of the brain. This technique is able to provide data about brain area activations, given that a change in brain activity correlates with a change in blood flow. See Raichle for an introduction about PET (Raichle, 1999).

[22] First person data may still have an important role for studying the emotions. See Elizabeth Irvine (2014) for a discussion about the use of personal data and of problems of an empirically informed philosophy of mind.

employed to define respectively rules of inference and the concept of justice (Goodman, 1954; Rawls, 1999). It consists of five stages: (i) generating intuitions about paradigmatic cases that display the property that we want to define, (ii) formulating a theory that accounts for the intuitions, (iii) should intuitions and theory conflict, revising either the theory or the judgments, (iv) checking the coherence with other background beliefs, and (v) terminating the enquiry when an equilibrium between intuition and theory is achieved.

In this thesis, I will focus on central types of emotions, such as the emotions of fear, anger, sadness, disgust, joy, guilt, hope, jealousy, worry, pride and admiration. My arguments have been formulated with these central types of emotions in mind. I will not directly discuss so-called "intellectual" emotions such as curiosity and love for research (see, for instance, Morton, 2010) intellectual interest and contemplation (see, for instance, Stocker, 2013), and the so-called "epistemic emotions" such as surprise, curiosity and trust (Arango-Muñoz and Michaelian, 2014; Meylan, 2014).

1.8. SIGNIFICANCE

Most discussions in epistemology of emotions have focussed on the epistemic role of emotional experiences. My research investigates issues arising from the idea that emotions *themselves* could be open to epistemic rational assessment. This thesis explains the rational assessability of emotions and expands the scope of epistemic rationality to mental states that have been considered beyond rational assessment.

My account of epistemic rational assessability could in principle be accepted both by cognitivists and by some non-cognitivist theories of emotions such as the perceptual theory and the psychoevolutionary theory, offering a novel solution to the problem of reconciling empirical adequacy theories of emotions and intuitions about the epistemic rational assessability of emotional states. The account that I will be putting forward explains the epistemic rational assessment in a way that fits with the epistemic rational assessability of

mental states in general. This goes towards the aim of delivering an account of the epistemic rationality that coheres with the epistemic rationality of other mental states, according to which each emotion may be epistemically rationally assessable in virtues of features that are not unique to the emotions, but that are shared with other epistemically rationally assessable mental state occurrences. This account may also be used to assess the rational assessability status of other mental states.

In this thesis, I also identify new questions arising from the claim that emotion occurrences are open to rational assessment. I will discuss the significance of the other claims in detail in the following chapters. What I think is important to highlight here is that, once we assess emotions as I suggest we should do in chapter 6, we end up noticing that irrational emotions are often experienced not only by people that suffer from mental health illnesses such as people with a phobia or bipolar disorders but also by everyone. This contributes to show a continuity between emotions experienced by people who have emotional disorders and emotions of people that do not suffer from emotional disorders. This, by itself, is a step forward in the direction of reducing the stigma associated with emotional disorders.

Irrational emotions are widespread, but, as I will argue in chapters 7 and 8, we often have the possibility to use the rationality norms as an ideal to tend to, regulating our emotions and our emotion forming dispositions over time. This is very interesting not only because it opens up new questions about the extent to which subjects are responsible for their emotions, from an epistemic point of view, but also because it delivers to us clear guiding principles that can help us to improve the epistemic status of our emotions. If the bad news is that irrational emotions are ubiquitous, the good news is that they can be trained and regulated over time. For the wider public, the ultimate aim of this thesis is to empower people to take up (epistemic) responsibility for how they form and shape their emotions and emotional disposition in the long term.

1.9. Summary and Conclusion

In this chapter, I have set the ground for the arguments that will be developing in this thesis. I provided the reader with a sketch of the historical context of my research, identifying an "epistemic turn" in philosophy of emotions. New questions regarding the epistemic rational assessability of emotions have been raised. I argued that the epistemic rational assessability of emotions poses a dilemma for existing theories of emotions, and that we need a serious investigation of the emotions' rational assessability that coheres with other mental states.

I also anticipated that the solution that I will develop in chapter 4 can be found investigating not the rational assessability of the emotion *type*, but the rational assessability of emotion occurrences. The possibility to consider at least some emotions as being open to rational assessment raises controversies about the epistemic rationality status of emotions, our epistemic responsibility for having emotions and about the degree to which we are able to control them. I will come back to these questions in the following chapters of my thesis.

CHAPTER 2: THE CONTENT OF THE EMOTION AND THE CONTENT OF THE CAUSAL EVALUATION

ABSTRACT

A central claim of existing theories of emotions is that emotions have contents. Another important claim is that emotions are triggered by causal evaluations, i.e. evaluations of the impact of the situation for the subject's well-being. Surprisingly, the relationship between the content of the emotion and the content of the causal evaluation has been overlooked in the existing literature. In this chapter, I investigate the relationship between the content of the emotion and the content of the causal evaluation. I will argue that the content of the emotion is not identical to the content of the evaluation that triggers the emotion.

INTRODUCTION

Before giving a detailed explanation of the epistemic rational assessability of emotions, that I will provide in chapter 4, I need to discuss one central claim that is usually endorsed by many theories of emotions: the claim that emotions involve evaluations.[23] Evaluations involved in the emotional experience can be of two different types: causal and constitutive. On the one hand, *causal* evaluations are causal antecedents of emotions: they are responsible for setting off emotions. On the other hand, *constitutive* evaluations are not causal antecedents of the emotion, but they are components of the emotions. Causal evaluations also have an epistemic function: they explain why emotions are activated. Constitutive evaluations, instead, describe

[23] Most theories of emotions endorse this assumption e.g. (Griffiths, 1997; Prinz 2006a; Ellsworth and Scherer, 2001; Scherer, Schorr, and Johnstone, 1984; Nussbaum, 2001). However, feeling theories of emotions (James, 1884), and the Schacter and Singer's theory of emotions do not endorse it (Schachter and Singer, 1962). In this chapter, I assume that emotions involve evaluations.

the ways in which events are presented to us in our emotional experience. Constitutive evaluations are part of the emotional experience, and they determine, at least in part, the content of the emotion.

In this chapter, I investigate the relationship between causal and constitutive evaluations. Gaining a better understanding of the relationship between emotions and causal evaluations will help us to acquire a better understanding of the content of emotions. This will be useful for individuating the epistemic rationality norms of emotions (chapter 6), as well as for understanding the correctness conditions of emotions, which I will discuss in chapter 5.

Given the close relationship between causal evaluations and emotions, it seems *prima facie* plausible to think that the information carried by emotions is closely related to the information embedded in causal evaluations. However, Robert Solomon showed that the cause of the emotion might differ from the intentional object of emotion (Solomon, 1984). This may suggest that, on some occasions, not only the cause and the object of the emotion differ, but that the content of the causal evaluation and the content of the emotion are different. However, Solomon does not develop this point further. More recently, Larry Herzberg presented an intuition-pump argument that challenges the identification between the intentional object of the emotion and the mental representation that, according to him, triggers the emotion (Herzberg, 2009). However, his argument does not properly target the thesis that the content of the emotion is identical to the content of the causal evaluation.[24] The aim of this chapter is to fill this gap.

The lack of consensus between researchers on the nature of emotions renders the enterprise of explaining the relationship between the content of the emotion and the content of the causal evaluation more complex: different theories of emotions endorse different accounts of causal

[24] See section 2.3.1 for a discussion of this point.

and constitutive evaluations. For instance, cognitivists think that constitutive evaluations are very similar to evaluative beliefs (Nussbaum, 2004; Solomon, 1988), whereas non-cognitivists think that constitutive evaluations are not propositionally structured, and that they are mere associations between an object and an evaluative term, such as danger, loss, unexpected (Griffiths, 1997).[25]

The chapter proposes a central thesis: the rejection of the identity between the content of the emotion and the content of the causal evaluation. It is beyond the scope of this chapter to argue for the widely accepted thesis that emotions usually involve causal evaluations, or that emotions have contents. Although these two assumptions are popular in the literature, they are not endorsed by all theories of emotions. In chapter 1, I outlined the feeling theory of emotions, which claims that emotions are nothing over and above the feeling of bodily changes. The feeling theory would not endorse the claim that emotions involve causal and constitutive evaluations (James, 1884). Therefore, the argument offered in this chapter is a conditional argument: if causal and constitutive evaluations exist and if emotions and causal evaluations have contents, then the relationship between the content of the emotion and the content of the causal evaluation is as I propose here.

A better understanding of the relationship between the content of the emotion and the content of the causal evaluation would improve our understanding of the contribution of the information carried by causal evaluations to the content of the emotions. More precisely, it has the potential to show that the accounts of the content of the emotions that assume an identification between the content of the emotion and the content of the causal evaluation are not compelling. This opens up the possibility that the content of the evaluation that constitutes the emotion brings about an element of *originality*, which cannot be inherited from the

[25] See also Moors (2010) for a defence of causal evaluations that are non-propositionally structured.

emotion's cognitive bases, nor from the emotion's causal evaluation. This argument is important because understanding the relationship between causal evaluations and the content of the emotions is crucial for assessing the correctness of the emotion, which I will discuss in chapter 5, and its epistemic rationality, which I discuss in chapter 6. This is because the assessment of the epistemic rationality of the emotion may also rely on the assessment of causal evaluations. Understanding whether the evaluative content of the emotion is not inherited from the causal evaluation of the emotion, but it has an element of originality may also affect the epistemic status of the emotion. Moreover, my account has also interesting applications regarding ambivalent emotions. I will discuss these applications in section 2.4.

In section 2.1, I use psychologists' accounts of causal appraisals to understand the notion of causal evaluations. In section 2.2, I argue that there is a lack of clarity concerning the relationship between the content of the emotion and the content of the causal evaluation. Then, I explore the relationship between the content of the causal evaluation and the content of the emotion, and I argue that the content of the causal evaluation cannot be identified with the content of the emotion. In section 2.3, I defend this thesis by using an intuition-pump. In section 2.4, I show that acknowledging that the content of the causal evaluation could be different from the content of the emotions has interesting applications for understanding the rationality of ambivalent emotions.

2.1. Explaining Evaluations as Appraisals

Imagine coming back to your house and seeing that it is particularly difficult to turn the key in the lock. You are surprised because you have never encountered this difficulty. When you finally succeed, you enter your living room and see all your possessions scattered everywhere. The pillows are on the floor, the sheets you left on the table are disseminated on the ground, and the small desk lamp is smashed. As a result of this sight, you immediately think that something bad has happened in your home and you feel fear. You hypothesise that burglars

entered your apartment. You know that you do not have any precious objects at home other than your grandmother's earrings. You fear that the earrings were stolen, and you run towards the drawer next to your bed, desperately looking for them. Your heart beats vigorously, and your hands tremble when you open it. The earrings are missing, and you start to cry, sad that your grandmother's precious gift was lost. And you feel anger at the person who stole them.

As this example shows, depending on the perceptual stimuli and the beliefs you have, you feel a different emotion. Consciously or not, you evaluate what is happening to you, and thus a given emotion arises. You do not expect the lock to be difficult to open and therefore you are surprised. When you see the turmoil in your living room, you feel scared. You feel hurt by the burglar's actions, and this makes you angry at them. In this imagined scenario, your evaluations of the situation as unexpected, dangerous, offensive, and loss-involving are the causal evaluations that activate your emotions.

We continuously look for things that are relevant to our well-being (Lazarus 1991). When a particular evaluation has taken place, an emotion is triggered. In this way, causal evaluations explain under what condition a given emotion activates. For instance, fear activates as a result of an appraisal of something as dangerous, whereas anger activates as a result of an evaluation of something as offensive. Philosophical accounts of emotions accept that emotions are usually triggered by evaluations (see, for instance, Nussbaum 2001; Ben-Ze'ev 2000; Price, 2015, Deonna and Scherer 2010). Psychologists also agree on the thesis that emotions are triggered by a *causal* evaluation, which they call causal appraisals (Moors 2010; Scherer et al. 1984). Psychologists such as Richard Lazarus and Klaus Scherer individuated the relationships between the types of evaluations that trigger the emotions and emotions' types (Lazarus, 1991; Scherer et al., 1984). The set of evaluations, or "appraisals" in their terminology, that are associated with anger, for instance, is composed of elements such as the evaluation of the high goal-relevance of the situation, the frustration of a goal, the threat to personal and social esteem,

etc. Such components of each emotion type can be summarised in more general evaluations, called by psychologists "core relational themes" (see subsection 2.1.2), and "formal objects" by philosophers (Teroni, 2007). For instance, in case of anger, the core relational theme is "a demeaning offence against me and mine" (Lazarus, 1991), whereas, in case of fright, the appraisal is "facing an immediate, concrete and overwhelming physical danger" (Lazarus, 1991).

The evaluation embedded in the emotion is relative to the subject's perspective; more precisely, evaluations are relative to the subject's concerns.[26] In this chapter, when I will use the expression "evaluative properties" I will only refer to *thick affective* properties. Thick properties are properties, such as being dangerous, being a loss, and being offensive, which have both an evaluative and a descriptive component (Williams, 2011). They are often contrasted with *thin* evaluative properties, namely properties that are wholly evaluative, such as the property "being good". Thick *affective* properties are thick properties that are conceptually related to emotions (Tappolet, 2004).

Emotions present their object as having evaluative properties that are dependent on the subject's concerns. Consider, for instance, King Duncan's death in Shakespeare's play *Macbeth* (Shakespeare, 2015 [1623]). Lady Macbeth rejoices at the success of her plot to kill King Duncan. However, Donalbain and Malcolm, King Duncan's sons, are sad and terrified by King Duncan's death. Lady Macbeth wants to become the Queen of Scotland and believes that committing this regicide is the essential step to achieve this aim. She plots King Duncan's killing and she strongly cares about the success of her murderous plan. Conversely, Donalbain and Malcolm care about their father's life and about their own safety. Their father's killing not only deprives them of their beloved father but also puts their lives at risk. This explains why

[26] See chapter 5, section 5.5 for a relativistic account of concerns.

different people who face the same situation, in this case, the death of King Duncan, react with a different emotion.

Cognitivists and perceptual theories do not explicitly distinguish causal from constitutive evaluations. But they can easily accommodate this distinction. Consider, for instance, Martha Nussbaum's cognitivist account. The evaluative judgments that constitute the emotion, in her account, would be a constitutive evaluation, whereas the judgment that causes the emotion to activate would be its causal evaluation. For instance, when Nussbaum describes the grief she experienced when her mother died, she claims that the judgment that her mother was very important to her and also cut off from her did not only *cause why* she feels the emotion, but also *constituted* it. As, for cognitivists, evaluative judgments cause the emotional experience and are also necessary constituents of it, they can be read as endorsing the thesis that evaluative judgments that are constitutive of emotions are identical with the judgments that cause them (see Nussbaum 2001, pp. 42-43).

Perceptual theories could accommodate the distinction between causal and constitutive evaluations. Constitutive evaluations would be the representations that are part of the emotional experience. Consider, for instance, Jesse Prinz's account (Prinz, 2004). He would describe Nussbaum's grief for her mother as representing her death as of being an irrevocable loss, and he would claim that this evaluation would constitute her grief. What could be the causal evaluation, in Prinz's account? The causal evaluation of her grief would be the evaluative representation that is involved in the triggering of emotional experiences. In certain cases, this would be an evaluative judgment. For instance, a perceptual theorist would accept that, in some cases, believing that some medical procedure is dangerous may cause the activation of fear, thereby acting as a causal evaluation. However, in other cases, the emotion may not be triggered by a causal evaluation, but by a mental representation that is part of what Prinz calls a 'calibration file'. A calibration file is a mental mechanism which contains mental

representations that track a certain evaluative property (Prinz, 2004, p. 100). For instance, calibration files for fear may include representations of guns and representations of health hazards. When a mental representation that is part of a calibration file is triggered, we could understand the association between the mental representation and the evaluative property associated to it as a causal evaluation.

The following three subsections will clarify the notion of causal evaluation used in the chapter, drawing on existing accounts of evaluations as appraisals in the philosophy and psychology literature.

2.1.1. Appraisals as Occurrent Evaluations

In the psychology literature, causal evaluations are called "appraisals". This term is used to refer both to the evaluative process of the situation and to the occurrent causal evaluation. The appraisal theory, for instance, defines appraisal as the process that is responsible for continuously evaluating the situation (Moors, 2010). However, it also defines "appraisal" as the single constituent that is part of the appraisal process (Ellsworth and Scherer, 2001). To avoid confusion, in this chapter by "appraisal" I will refer only to occurrent mental states that are the result of an appraisal process and not to the process itself. This use of the notion of evaluation to refer to an occurrent state and not to a process is consistent with philosophers' use of the notion of evaluation. Cognitivist theories about emotions, for instance, endorse the claim that emotions are triggered by an evaluative judgement of the situation. According to them, the evaluation is an occurrent mental state which is an evaluative judgement (Nussbaum, 2001). Imagine, for instance, finding a snake in your path and feeling fear. Cognitivists would explain your emotion of fear appealing to your conscious or unconscious belief that the snake is dangerous. The causal evaluation here is your occurrent belief. These evaluations can be either the cognitive state (e.g. a belief) that triggers the emotion or the output of the appraisal

process that triggers the emotions. Whether we have a proper appraisal mechanism is not discussed here.

2.1.2. Molar and Molecular Appraisals

Richard Lazarus, a pioneer in the research about appraisals of the emotions, examined the different kinds of person-environment relationships in different types of emotions (Lazarus, 1991). He identifies the main variables, or dimensions that figure in the content of the appraisals and establishes an identification between a given set of appraisals and a given set of emotions. These variables characterise the features of the person-environment encounter relations that are in place when we experience a specific emotion. These features are presented here: (1) *goal relevance*, i.e. the relevance of the situation for the subject's well-being; (2) *goal congruence or incongruence*, i.e. whether the event is congruent with the subject's goals; and (3) *type of ego involvement*, which stands for the kind of bearing the situation has for the subject and the other persons, values or ideas the subject identifies with. These three features result in three distinct evaluations that are called "primary appraisals" because they are evaluations concerning the impact of the situation on the subject's well-being.

In addition to these features, Lazarus presents other evaluations that he calls "secondary appraisals" (Lazarus, 1991). Secondary appraisals characterise the subject's ability to cope with different situations, according to the following parameters: (4) *blame or credit*, (5) *coping potential* and (6) *future expectations* (Lazarus, 1991).[27]

These variables are the molecular appraisals' components that identify each emotion. For instance, the emotion of sadness for the stolen earrings is characterised by an appraisal of the

[27] Recently, psychologists refined and added new variables, adding, for instance, the variables of novelty, legitimacy (Roseman, 1991) or motivational relevance, e.g. (Smith, 2009). For a review of appraisal theories, see Moors and Scherer (2013).

relevance of the situation for the personal well-being (the ring was a symbol of my affection to my grandmother); the incongruity of the situation with the subject goals (the loss of the earrings is incongruent with my goal of remembering and caring about a present my grandmother gave to me); it might include the loss of some type of ego-involvement, such as personal esteem (I was not able to prevent this) or meaning and ideas (loss of connection with my grandmother); it includes no blame and not favourable coping potential.

Each emotion is also associated with a "core relational theme", a theme that characterises and summarises the molecular appraisals. The core relational themes individuate and distinguish each emotion. Anger, for instance, was associated with the core relational theme "a demeaning offence against me and mine" and happiness as "making reasonable progress toward the realisation of a goal" (Lazarus, 1991, p. 122). Core relational themes intervene in what Lazarus calls "molar appraisals", i.e. appraisals that summarise the molecular appraisals components. In the case of sadness, for instance, the molecular appraisals' components are summarised by the molar appraisal of having experienced an irrevocable loss.

In what follows, for the sake of simplicity of the exposition, I will be referring only to the molar and not to molecular appraisals. Therefore, I will just discuss the relationship between the content of the emotion and the content of causal evaluations understood as causal *molar* appraisals.

2.2. The Content of Emotions and the Content of Causal Evaluations

One of the striking properties of mental states is their ability to carry information about objects. My perception of a computer in front of me, among other things, is about the computer in front of me, and my belief that Rome is the capital of Italy is about Rome being the capital of Italy. These mental states are carrying information about something. This information is called the

"content" of the mental state.[28] The content of a mental state describes a certain state of affairs, whose representation can be correct or veridical. The content of the mental state spells out the conditions under which the mental state is correct; for instance, the content of my belief that today in Birmingham is sunny spells out the conditions under which my belief is correct: my belief is correct if and only if today in Birmingham is sunny.

To understand whether the content of the causal evaluation and the content of the emotion are identical, we need to understand how contents of causal evaluations are individuated. In what follows, I identify the content of the emotions and the content of causal evaluations by spelling out their correctness-conditions. The perception of my computer situated on my desk presents me with some information, namely that my computer and my desk have a certain colour *c*, a certain location *l* and a certain shape *s*. This information is the content of my perceptual experience. My perception is accurate when my computer and my desk have the colour *c*, the location *l* and the shape *s*. If they do, the information that my perception is offering to me is accurate.

Similarly, the content of the emotion comprises, at least in part, the information emotions give to us. My fear of a snake presents me with the information that the snake is dangerous to a certain extent. This information constitutes the content of the emotion because it spells out the correctness condition of the emotion: the emotion of fear is correct if and only if the snake is dangerous to the extent indicated by the intensity of the fear.[29] The contents of causal evaluations are individuated in the same way: causal evaluations carry information about the cause of the emotion.

---

[28] For the sake of simplicity, in this chapter, I assume that carrying information about something is identical to having content. I am aware that this is a controversial issue.

[29] I discuss the correctness conditions of emotions in more detail in chapter 5.

In this chapter, I assume that causal evaluations are mental states that have content, that is, they carry information about something. Consider your evaluation that the burglar harmed you, or that the lack of the earring represents a loss for you. These two evaluations are about something, respectively the burglar and the loss of the earring. The contents of these two evaluations can thus be expressed with the propositions "the burglar harmed me" and "missing the earring is a loss for me".[30] The first observation about this issue is that causal evaluations that trigger the emotions are multifarious: even if the evaluations can be conscious, like when I appraise the lack of the earrings as involving loss, in many cases (perhaps most of them), we are not conscious of causal evaluations that intervene in the activation of the emotions. However, at least in some cases, we seem to be able to bring them to consciousness. After reflection, we might be able to realise that our angry reaction directed at our best friend was caused by an evaluation of one of his remarks as offensive. However, sometimes not only our causal evaluations are unconscious, but they are the result of automatic associative processes. Consider someone who is scared by an earthworm regardless of the fact that she does not regard it as dangerous. Can we legitimately attribute content to all these three kinds of evaluations, and in particular to subpersonal associative evaluations? Whether subpersonal mental states that cannot be brought to consciousness are contentful states is a controversial issue that has been the subject of extended debate. Cognitive science generally defends the existence of sub-personal contentful states in different contexts, e.g. language perception (see, for instance,

[30] This is an approximation. In this chapter, for the sake of simplicity, I do not consider what the intensity of emotion represents. In chapter 5, when I spell out the correctness conditions of emotion, I suggest that the evaluative content of the emotion depends not only on the type of emotions that is experienced, but also on the intensity of emotion. More precisely, I claim that the intensity of the emotion modifies the extent according to which an object has a certain evaluative property. This omission will not affect the plausibility of my argument for the thesis that the content of the causal evaluation and the content of the emotion may differ.

Bermúdez, 1995, 2007; Stich, 2011). This is because sub-personal mental states are carrying information and have veridicality conditions. In what follows, I endorse this view.

Much ink has been spilled on trying to qualify the content of mental states, and the subject is still a matter of controversy (Blackman, 2013; Deonna and Scherer, 2010; Döring, 2009). The literature distinguishes between conceptual and non-conceptual, propositional and non-propositional contents.[31] A subject is able to have a conceptual content if it fulfils Gareth Evans' Generality Constraint (Evans, 1982). This constraint states that to be able to have a conceptual content such as "the dog is dangerous", a subject should be able to entertain the thought that there are other things, other than a dog, that are dangerous. If a child cannot conceive that DOG is *b*, or that *A* is *dangerous*, where *b* is a variable that ranges over properties and *A* is a variable that ranges over objects, the child cannot have a conceptual content "the dog is dangerous". This distinction is useful to characterise the contents of causal evaluations that seem to be triggered in an automatic fashion or that are experienced by creatures that are not able to master concepts or propositional thoughts. Under the assumption that a child cannot fulfil Evan's Generality Constraint for the concepts DOG and the property *dangerous*, it follows that the causal evaluation involved in the child's fear of the dog is not conceptual because the child is not yet able to conceive that the DOG is *b* or that something else is dangerous.

However, when the child acquires linguistic abilities, she will be able to master the concepts of DOG and DANGEROUS. As Tye puts it: "to say that mental content is non-conceptual is to say that its subject need not possess any of the concepts that we, as theorists, exercise when

[31] For a discussion between conceptual and non-conceptual contents in philosophy of perception see Josefa Toribio (2007). For the sake of my argument, I do not need to rely on a specific view of the content of perception, nor on a specific view of the content of the emotions.

we state the correctness condition for that content" (Tye, 2000, p. 62). Does this imply that all the causal evaluations that have the content "the dog is dangerous" are conceptual? At least regarding causal evaluations and emotions, this does not follow.

This is because the content can be non-conceptual not only when the subject does not possess the concept used in the correctness condition of the relevant mental state, but also when the subject masters that concept, e.g. when the emotion is triggered by an association. The content of causal evaluations is sometimes inferentially isolated from other belief-like mental states. This may happen in cases of recalcitrant emotions, namely emotions that conflict with the subject's evaluative beliefs (Benbaji, 2013; Brady, 2009; Döring, 2015). As recalcitrant emotions show, sometimes the appraisal system triggers an emotion regardless of the subject's explicit evaluations. For instance, someone may feel scared at an earthworm regardless of her belief that the worm is not dangerous (Griffiths, 1997). I discuss the triggering process of recalcitrant emotions in more detail in chapter 7. Depending on whether the notion of inferentiality is constitutive of the notion of conceptual content, we might think that these inferentially isolated causal evaluations are non-conceptual. Note that inferentially isolated causal evaluations are not necessarily non-propositional. This is because causal appraisals can still be inferentially isolated even if they can be expressed via a proposition.

The main idea of non-conceptual content is that even if a subject lacks a concept *C*, she might still be able to have discriminative and attentive abilities that ground the ability to spell out the correctness condition of the mental state in question. In doing so, we can identify the content of the mental state without necessarily employing concepts. This argument has been extensively discussed in the philosophy of perception (Bermúdez, 2007). The suggestion here is that the subject has discriminative abilities regarding the relevance of the situation for its well-being. As we can distinguish between the lines that compose the hair of my friend

(Peacocke, 1992), without having a concept describing them, we can discriminate the kind of bearings a situation has on our well-being without having concepts for discriminating them.

2.3. Is the Emotional Content Identical to Causal Evaluations?

In the previous sections, I illustrated three central assumptions of this chapter: (1) emotions have contents, (2) emotions often require causal evaluations, and (3) causal evaluations have contents. If emotions are caused by causal evaluations, what is the relationship between the content of causal evaluations and the content of the emotions? Some philosophers seem to think that the content of the emotion is identical to the content of the causal evaluation. According to cognitivists, for instance, the content of the emotion is inherited from the content of the evaluation that causes it (e.g., Nussbaum, 2004). The content of the evaluation also explains why emotion is activated. This seems to imply the content of the evaluation that constitutes the emotion and the content of the causal evaluation are identical. For instance, describing the emotion of grief at the death of her mother, Martha Nussbaum identifies the content of the emotion with the content of the evaluative judgments, which explains why her sadness sets off. In other words, she seems to suggest that the content of the emotion and the content of the causal evaluation are identical (Nussbaum, 2004).

Other philosophers identify the content of the emotion at least in part with the content of the causal evaluation. Non-cognitivists, such as Jesse Prinz, also seem to assume an identification between the target of the emotion and its cause, suggesting an identity between the correctness conditions of the causal evaluation, and the correctness condition of the emotions. He claims that "when I am sad about the death of a child, I have one representation of the child's death and I have sadness attached to that representation […]. The sadness doesn't represent the death. Saying that my sadness is *about* the death does not mean that my sadness represents the death; rather it means that the death is what has caused me to become sad" (Prinz 2004, p. 62). Although Prinz does not explicitly argue for an identification between the content of the

emotion with the content of the causal evaluation, he seems to suggest, as Herzberg points out, a "semantic defence" of the identity between the intentional object of the emotion and the object of the causal evaluation, in that the meaning of being "about" something is, according to Prinz, "being caused by" it (Herzberg, 2009, p. 172).

In this section, I argue that the content of the emotion cannot be identified with the content of the causal evaluation, and also that the evaluative component that constitutes the content of the emotion cannot be always inherited from the content of the causal evaluation. I first discuss the argument presented by Herzberg (2009) against the thesis that the content of the emotion is identical to the content of the causal evaluation (subsection 2.3.1) and I then present my intuition-pump argument from individual cases (section 2.3.2).

2.3.1. Cause and Cognitive Bases

Herzberg challenges the thesis that the intentional object of the emotion is the same as the causal evaluation presenting cases where, according to him, the content of the causal evaluation is different from the content of the emotion (Herzberg, 2009). Interestingly, some of the situations that he presents do not seem to properly target the thesis that the intentional object of the emotion is the same as the emotion's *causal evaluation*.

Herzberg considers a situation where a subject plans a picnic, but when the day of the picnic arrives, he worries that it is going to rain, even if he lacks a belief that it is going to rain. He claims that the causal evaluation, in this situation, is the belief that the picnic is today. And this is different from the intentional object of the emotion, that would be "that today is going to rain". As he puts it:

> "Suppose that I have been planning a picnic with friends for a month, and very much hope it will occur. The day has arrived, and I am quite anxious (in the sense of worried) *that it will rain today*. […] My worry is triggered only by a belief that *the picnic is today*, given three other causal factors: a predisposition to be anxious about the success of my social plans, a hope that the picnic will occur, and a strong desire that it will not rain today" (Herzberg, p. 175).

Although the belief that the picnic is today may play an important causal role for the activation of the emotion, it does not qualify a causal evaluation. This is because belief "the picnic is today" does not have the evaluative structure described in section 2.2. The belief "the picnic is today" is merely one of the emotion's cognitive bases. The cognitive bases of an emotion are the mental states that ground the emotional experience providing information to the emotional system (Deonna and Teroni, 2012). For instance, my fear of a snake is based on the perception of a snake, Ann's embarrassment for a blunder is based on the belief that she made a blunder. What I want to suggest here is that Herzberg's argument seems to conflate the notion of cognitive basis with the notion of causal evaluations. This argument, for instance, seems to better target the thesis that the intentional object of the *emotion's cognitive basis* is identical to the intentional object of the emotion, rather than the thesis that the intentional object of the emotion's causal evaluation is identical to the intentional object of the emotion. Although this is surely an interesting thesis to discuss,[32] this Herzberg's argument does not successfully show that the content of the causal evaluation may be different from the content of the emotion.

Another of Herzberg's situations seem to show that there is a distinction between the content of the causal evaluation and the content of the emotion. This is the cases of emotional *mis-direction*, which are situations in which the subject directs her emotion at an object that is different from the triggering cause (Herzberg 2009, p.175). According to Herzberg, the function of emotional direction is to guide the emotion's motivational properties. Herzberg considers mis-directed emotions as "dysfunctional" because the motivational properties of the emotion are targeting an object that is different from the object that caused the emotion, and

[32] It has been argued that the object of the emotion is the same object of the cognitive bases of the emotion (Deonna and Teroni, 2012, p. 5). Herzberg's argument may challenge such a view.

this may be problematic for the subject (Herzberg, p. 175). An example of a situation with "dysfunctional" direction in Herzberg's sense is one in which a subject becomes angered because of an inappropriate remark of their boss, but instead of directing their anger at their boss, they direct it at someone else, e.g. their hapless colleague. In this case, the anger is caused by an evaluation that is different from the evaluative component of the content of the emotion. The content of the causal evaluation may be "this remark is offensive", but the anger, instead of being directed at the boss, is directed at somebody else, in this case their colleague. I take these to be interesting cases, which may have the potential to show a genuine difference between the content of the causal evaluation and the content of the emotion. However, it is unclear to what extent these are convincing cases. In describing these cases, Herzberg seems to assume that a single emotional episode may remain "undirected" for a certain time, as Herzberg acknowledges, and we need empirical evidence that this is possible. Moreover, he also assumes that when the emotion is finally directed to an object, there is not any subsequent causal evaluation directed at the same object that is responsible for the triggering of the emotion, and this seems controversial.[33]

In what follows, I want to focus on different cases. I attempt to show that divergence between the content of the causal evaluation and the content of the emotion is not only a prerogative of emotions that are dysfunctional in Herzberg's sense, but are a wider phenomenon. In particular, I will highlight the possibility that the content of the emotion and the content of the causal evaluation may differ in the *scope* of the intentional object.

[33] Perhaps these cases are better explained as influencing the subject's mood, that then alters the subject's sensibility to certain objects or events.

2.3.2. Intuition-Pump from Individual Cases

In this section, I present situations, which show that the content of the emotion and the content of the causal evaluation diverge. The aim of this section is to suggest that it is possible to have cases in which there is such a difference. I will also suggest that the situations here presented are not rare cases. Firstly, I present two situations in which it seems that the content of the causal evaluation is identical to the evaluative component of the content of the emotion. I will then move on to consider situations in which it seems that the content of the emotion diverges from the content of the causal evaluation. Consider the following situations:

> PRIZE WINNING: Amy receives a letter, which says that she has received a prestigious prize. Because of this news, she is very happy.
>
> CAR CUTTING: Mary is driving and suddenly a car cuts in front of her. As a result of this, she feels an intense fear.

Assume that in the PRIZE WINNING situation, the evaluation of the prize awarded sets off Amy's happiness. In this situation, assume that Amy's causal evaluation is the belief: "winning the prize is a great achievement". In this situation, it seems very plausible that Amy's causal evaluation coincides with the content of the emotion. This identification is initially appealing because it supports the idea that emotions present us with information about the objects in the environment, and that these objects are causes of the emotions.

Similarly, in the CAR CUTTING example, Mary's fear is triggered by the evaluation of the car as dangerous. In this case, Mary's fear is activated because of the causal evaluation of the car's movement as dangerous. Also, in this situation, it seems that the content of Mary's evaluation is identical to the content of Mary's fear. That is because this situation is such that Mary's fear is directed towards the unexpected car, and Mary's fear seems to be correctly instantiated only if that car was dangerous for her. Again, this solution seems elegant: it neatly

explains the relationship between causal evaluations and the content of emotions. But can the content of an emotion be considered always identical to the content of its causal evaluation?

In what follows, I defend a negative answer to this question. One possible reason for this claim is that there may be other components of the content of the emotions that are different from the evaluative content. Following for instance Peter Goldie (Goldie, 2000), we could argue that emotions can have phenomenal content and *also* an evaluative content. And if this is the case, then the content of the causal evaluation will be trivially not identical to the content of the emotion. This is because the content of the emotion will be more complex than the content of a causal evaluation because it will include both phenomenal *and* evaluative content. However, my argument against the identification of the emotional content with the causal evaluation content aims to show that even the content of the causal evaluation is not identical to the evaluative component of the content of the emotion.

I will present some situations where the cause of emotion does not coincide with the object of the emotion. And when this is the case the content of the causal evaluations, and the content of the constitutive evaluation diverge. In what follows, for the sake of simplicity of exposition, I will refer to the content of the constitutive evaluation with the expression "the content of the emotion". This will not imply that the content of emotion can be reduced only to the content of its constitutive causal evaluation.

Consider the following situation:

SCARY NOISE: Peter is walking on a mountain path. He hears a sharp scream, and he feels fear.

In SCARY NOISE, Peter's perceptual experience is evaluated by his emotional system, which immediately triggers fear in him. Assume Peter's emotional system previously stored an association between screaming and danger. Because of this association, when he registers the

presence of a scream, the representation of danger activates.[34] At this point, Peter feels fear. Assume that the kind of causal evaluation involved in this situation is triggered by an associative process: the presence of the scream is associated with the presence of danger so that when the scream is heard, the representation of danger is activated, and fear occurs. Here, the content of the causal evaluation is "there is danger". Now, assume that the content of the emotions is identical to the causal evaluation. Under this assumption, Peter's fear has the content "there is danger".

There are reasons for questioning this identification. A standard view in philosophy is that emotions are necessarily directed at particular objects, and occurrent affective states whose particular objects are missing are normally considered to be moods, rather than emotions (de Sousa, 1987; Lyons, 1980). Peter's affect seems to qualify as an emotion because it has some of the other characteristics that normally are associated with emotions: it has a short duration, a high intensity, it is caused by an event that is localised in time, and it tells us about something that regards the external world and not the subject's overall personal life (Parkinson, Totterdell, Briner, and Reynolds, 1996, pp. 6-11).

So, if you think that Peter's affective state is an emotion and not a mood, then there should be an object that is the target of Peter's emotion of fear. This entity could be, for instance, the entity that caused the scream, or the entity that is screaming. If this is the case, then the emotion's content will be different from the causal evaluation content: while the emotion's content is "the object that causes the scream is dangerous", the causal evaluation is more simply "there is danger".

However, cognitivists would not consider Peter's emotion to be a full-fledged emotion because by assumption Peter's emotion is triggered by an associative evaluation that is the

[34] This does not imply that Peter believes that the scream is dangerous.

output of an associative process and not by a causal evaluation with a propositional structure. More needs to be done for persuading cognitivists that the content of the causal evaluation and the content of the emotion may differ. My strategy will be to find other situations where the content of the emotion and the content of the evaluation diverge. Below, I will present situations that involve a causal evaluation that could plausibly have a propositional structure. Consider the following scenario:

OFFENSIVE REMARK: Sarah overhears her friend Alison, who says "Sarah is dull and unattractive". Sarah thinks that Alison's remark is offensive and feels anger at Alison.

In this situation, Sarah evaluates Alison's remark. The content of Sarah's evaluation is that Alison's remark is offensive. If Sarah is asked why she is angry at Alison, she will probably say that it is because of Alison's remark. Alison's remark is the cause of the emotion, and her causal evaluation is about Alison's remark. However, at least on some occasions, the content of Sarah's emotion is different from the content of Sarah's causal evaluations. This is because when Sarah feels anger, the target of her emotion can be Alison as a person, not only her action of pronouncing an offensive remark. This is not to say that always, in analogous cases, we are angry at the person that is saying the remark, rather than at the action of telling an offensive remark. It seems like a platitude that we can be angry at the person that is uttering an offensive remark. When this happens, then there will be a divergence between the content of the emotion and the content of the causal evaluation. In particular, the divergence is in the focus of the evaluation: the emotion has a *global* scope, whereas the causal evaluation has a *particular* focus.[35]

[35] See Abramson (2010) for a discussion of global and particular scope relative to the emotions of contempt, shame and disdain.

Consider the following two interesting cases in which the emotion's and the causal evaluation's content differ:

JEALOUSY: Karl and Mary have a relationship. One day Karl learns that, unbeknownst to him, Peter invited Mary to have a dinner by candlelight. He feels jealous of Peter.

FEAR of the DOG: John does not tend to be scared by dogs, however, this time he sees this dog's big teeth and feels fear.

In these situations, again, we have a divergence between the contents of the emotions and the contents of the causal evaluations. In JEALOUSY, Karl's emotion seems to be directed at Peter, as a person, and not at the event of Peter's having dinner by candlelight with Mary. Thus, the content of the emotion seems to be the proposition "Peter is threatening my affection", whereas the content of the causal evaluation is the following: "*Peter's having dinner by candlelight with Mary* is threatening my affection".

Similarly, also in FEAR of the DOG, the content of the emotion differs from the content of the causal evaluation. This is because it could be that the cause of the emotion is the dog's big teeth, and therefore the content of the causal evaluation is "these big teeth are dangerous", whereas the content of the emotion is not that the big teeth are dangerous, but that the whole dog is dangerous. Aaron Ben-Ze'ev's argued that emotions are often – although not always – directed to agents and living creatures, and not to non-agential objects (Ben-Ze'ev, 2000, p. 29). Once the emotion of fear has been triggered, the intentional object of the emotion is not the same as the object of the causal evaluation: while the object of the emotion is an agent, the object of the causal evaluation is a property of the teeth.[36] There are many cases of anger where

[36] It might be possible, using Ben-Ze'ev's suggestion to offer a proposal on how to generate situations that support a distinction between the content of the emotion and the content of the causal evaluation. For instance, it could be

the object of the causal evaluation coincides with the actual cause of the emotion, and the object of the anger is the agent that is perceived to be responsible for the emotionally salient situation. This agent could be, for instance, the subject who utters the offensive sentence, the person that "owns" the offensive mental state, or the person who entertains an offensive action.

To conclude, these examples suggest that the cause of the emotion that features in the content of the causal evaluation does not always match the content of the emotion. Therefore, the content of the causal evaluation cannot be considered identical to the content of the emotion.

2.4. Emotional Ambivalence and Content Divergence

In this section, I am going to provide a useful application of thesis that the content of the emotion is not identical to the content of the causal evaluation. I argue that the divergence between the content of the emotion and the content of the causal evaluation is useful for understanding why ambivalent emotions are triggered, and for assessing their *procedural* rationality. Ambivalent emotions are emotions with opposing polarities *directed at the same*

---

argued that the intentional object of the emotion is the nearest agent that is regarded by the subject to be the cause of the emotion. And that the intentional object of the causal evaluation is the nearest event, proposition or entity such that if it did not happen, the emotion would not have been activated. When the cause of an emotion is not a proper agent, but another entity (as FEAR of the DOG) or an event (as in JEALOUSY) or something caused by the agent (as the remark in OFFENSIVE REMARK and the noise in SCARY NOISE), the content of causal evaluation and the content of the emotion may diverge. Although this may be an interesting proposal to explore, this falls beyond the scope of this chapter and I do not want to argue for this generalisation here. Arguing for such a generalisation would require also explaining whether the existence of emotions directed at events, e.g. anxiety about global warming, is a counterexample of the generalisation. Here I merely want to suggest that the object of the emotion and the object of the causal evaluation diverge not only in exceptional circumstances, but also in more ordinary situations.

object (Greenspan, 1988; Kristjánsson, 2010; Tappolet, 2005). Consider, the following situation:[37]

> Bethany and Alice are applying for a post-doctoral position at the University of Oxford. They are best friends and are both good candidates to gain the job. Unfortunately, there is only one position open, and given that Alice obtains the job, all the other candidates, including Bethany had their application rejected. Bethany is happy for Alice because she is one of her best friends and she thinks that she deserved to have a place there. However, she is also sad: she craved that opportunity and she lost it.

As this example shows, Bethany has conflicting emotions: Bethany is happy and sad at the same time. It is difficult to account for this contrast while maintaining an identity between the content of the emotion and the content of the causal evaluation. This is because a good explanation of this phenomenon should explain both the reason why conflicting emotions activate, as well as accounting for the existence of contrast in the emotions. If the content of the causal evaluation were the same as the content of the emotion, we would have the two alternatives presented below.

Account One:

The content of the causal evaluation$_1$: My friend's success is good.
The content of the causal evaluation$_2$: My failure is a loss.
The content of the emotion$_1$: My friend's success is good.
The content of the emotion$_2$: My failure is a loss.

Account Two:

The content of the causal evaluation$_1$: Alice's success is good.
The content of the causal evaluation$_2$: Alice's success is a loss.
The content of the emotion$_1$: Alice's success is good.
The content of the emotion$_2$: Alice's success is a loss.

[37] This situation is a variation from Greenspan's example (Greenspan, 1980).

In account One, the contents of happiness and sadness have different objects: the object of happiness and the object of its causal evaluation is the success of Bethany's friend; the object of sadness and the object of its causal evaluation is Bethany's failure. It seems plausible that account One successfully explains *some* cases of two co-occurring, contrasting emotions.

It seems also possible that Bethany experiences two co-occurring emotions directed at the same object, namely Alice's success.[38] One might want to deny this possibility. However, there is empirical evidence that shows that it is possible to experience two co-occurring contrasting emotions, although the empirical evidence does not fully establish that two co-occurring contrasting emotions *directed at the same object* are possible (Barrett, Robins, Wildschut, Sedikides, et al., 2010; Ersner-Hershfield, Mikels, Sullivan, and Carstensen, 2008; Larsen, McGraw, and Cacioppo, 2001).[39] Even if we lack a decisive reason for thinking that cases like this are possible, without a strong reason to believe that it is *impossible* to experience ambivalent emotions, we should leave open the possibility that ambivalent emotions directed at the same object exist. Account One does not successfully explain a situation in which Bethany feels happy and sad *about the same object*, namely Alice's success. This is because, in account One, the emotions of happiness and sadness are not directed at the same object.

---

[38] Similar cases have been discussed by Greenspan (1988), Calabi and Santambrogio (2018).

[39] There is no empirical evidence currently available that clearly demonstrates that it is possible to experience two co-occurring contrasting emotion at the same time *directed at the same object*. This is because psychologists often define ambivalent emotions as two co-occurring emotions that have an opposite valence, that is, that have a positive and negative connotation, e.g. (Larsen et al., 2001). This contrasts with the way in which philosophers define ambivalent emotions, namely two conflicting and co-occurrent emotions directed at the same object, e.g. (Kristjánsson, 2010). In the philosophical literature, the contrast is not only in the valence of the mental state, but in how the object is represented.

According to account Two, the emotions of happiness and sadness are directed at the same object. This account, therefore, seems better equipped to explain a situation in which Bethany feels truly ambivalent emotions, which are directed at the same object. The contradiction between the contents of the emotions explains the tension that Bethany experiences about Alice's success. However, account Two does not successfully explain why Bethany ends up feeling happy and sad about the same event at the same time. It seems mysterious why emotions with *contradictory* causal evaluations are triggered at the same time.

Here I suggest that the best explanation of Bethany's ambivalent emotions is that the content of the emotion and the content of the causal evaluation differ. In what follows, I argue that this explanation better fits with the intelligibility of the activation of the emotion, as well as with the phenomenology of the subject's experience. I propose that Bethany's ambivalent emotions are described by the following emotions' and causal evaluations' contents:

Account three:

Content of the causal evaluation1: My friend's success is good.
Content of the causal evaluation2: My failure is a loss.
Content of the emotion1: Alice's success is good.
Content of the emotion2: Alice's success is a loss.

Account Two does not explain satisfactorily why each of the two contrasting emotions is activated. However, account Three is able to provide a more compelling explanation than account Two. Kristian Kristjánsson argues that emotions are sensitive to the subject's perspective and they depend on the *aspects* through which the subject grasps the situation (Kristjánsson, 2010). According to account Three, the causal evaluations are about two different aspects of the same object: Alice's success. Alice's success is apprehended by Bethany as the success of Bethany's friend, and it is also apprehended as Bethany's failure. There is no tension between the two aspects. This better explains why the two contrasting emotions are triggered at the same time.

Account Three also better explains the phenomenology of Bethany's experience. When experiencing ambivalent emotions, the subject feels pulled in two opposing directions. This phenomenological feature of the experience cannot easily be captured by account One, since in account One the intentional objects are different, and the tension is missing. Account Three better explains the contrast present in cases of ambivalent emotions (the emotions' contents are in contradiction), and the intelligibility of the emotions: the causal evaluations' contents are responding to different aspects of the situation that are not in contradiction. This distinction may be helpful also for assessing whether there is some irrationality involved in ambivalent emotions: the irrationality seems not present at the level of the causal evaluation since Bethany's happiness and her sadness are triggered by two different aspects of the same events. However, her emotions are motivating her to act in very different ways. Should she call Alice and congratulate her for her success with enthusiasm, or should she tell Alice how sad she is for having lost a wonderful opportunity? The distinction between the content of the emotion and the content of the causal evaluation is useful to narrow down the source of irrationality that may be involved in ambivalent emotions. The irrationality, if any, seems to be located in the emotions' contents, and not in the causal evaluation's contents. This is because the emotions' contents, but not the causal evaluation's contents, have a *global content*: they apply to the object as a whole (Abramson, 2010). In particular, this precludes the appreciation of other evaluative properties of the object. I do not have space here to offer a fully-fledged account of the irrationality involved in ambivalent emotions, and I will leave it for future work. What matters here is that distinguishing the content of the emotion from the content of the causal evaluation can help us not only to understand why certain emotions are triggered but also to assess their rationality status.

2.5. Summary and Conclusion

This chapter provided an insight into the relationship between the object of the emotion and its causal evaluation. I argued that the object of the emotion could not be considered identical to the object of the causal evaluation that triggers it. This is because the content of the causal evaluation might diverge from the object to which the emotion is directed. This result is interesting because it shows that the evaluation that constitutes the object of the emotion may bring about an element of originality, which is not inherited from previous mental states. I defended this thesis with arguments presenting situations in which the emotion and the causal evaluation's content differ. I also showed an interesting application of my argument for understanding ambivalent emotions.

This discussion has interesting applications for the rationality of emotion. A clear understanding of the relationship between the content of the emotion and the content of the causal evaluation helps us to identify sources of irrationality. In chapter 6, I will outline epistemic rationality norms that could apply to the emotions. We may expect that epistemic norms of emotions may be flouted by constitutive evaluations, and not by causal evaluations, or by causal evaluations and not by constitutive evaluations.

This chapter could also have interesting applications for understanding the conditions of correctness of the emotion, which I will discuss in chapter 5. This discussion suggests the possibility of divergence between the correctness of the emotion and the correctness of the causal evaluations.

CHAPTER 3: EMOTIONAL CONTAGION AND CAUSAL EVALUATIONS

ABSTRACT

Emotional contagion challenges the thesis that causal evaluations are necessary for emotional experience (Robinson 2009). This is because the most prominent model of emotional contagion, the mimicry-synchrony model (Hatfield, Cacioppo, and Rapson, 1994) explains emotional activation as a result of emotional contagion without any intervention of causal evaluations. In this chapter, I argue that this challenge can be met. I offer an account of associations that may trigger causal evaluations in cases of emotional contagion. This solution offers a better understanding of emotion activation as a result of emotional contagion and enhances our understanding of the role of causal evaluations in setting off emotions.

INTRODUCTION

In the previous chapter, drawing on existing literature from philosophy and psychology, I have relied on the claim that causal evaluations[40] are *often* involved in emotional experiences as their triggers. However, some have advanced a more ambitious thesis, according to which causal evaluations are not only *often* involved in the activation of emotions but are a *sine qua non* condition for it (Robinson, 2005). I have not relied on the latter claim and I am not committed to it in the other chapters of this thesis. But are there any reasons for denying that causal evaluations are always involved in triggering of the emotions? The claim about the necessity of causal evaluations is interesting because it provides explanatory advantages.

[40] By causal evaluation I mean causal *emotional* evaluations. In this chapter, every time I mention causal evaluations, I will be referring to causal emotional evaluations.

However, "primitive" emotional contagion (Hatfield, Bensman, Thornton, and Rapson, 2014; Hatfield et al., 1994) represents a challenge to the thesis that causal evaluations are necessary for emotional experience (Robinson, 2009) because emotions set off by emotional contagion do not seem to involve any causal evaluations. If the challenge is effective, then either emotional contagion does not trigger proper emotions, or the thesis of the necessity of causal evaluations for emotional experience must be rejected.

Surprisingly, emotions set off by emotional contagion have been understudied by philosophers. These emotions demand a further elucidation because a better account of emotional contagion could help us to illuminate the role of causal evaluations in setting off emotions. In this chapter, I argue that the challenge posed by emotional contagion to the claim that causal evaluations are necessary can be met by suggesting there are causal evaluations involved in emotions set off by emotional contagion.[41] If, at the end of this chapter, my readers think that this suggestion is convincing, then they can be less sceptical about the claim of the necessity of causal evaluations. Otherwise, the arguments developed in the remaining part of the thesis can be read as relying only on the modest claim that emotions are *often* set off by causal evaluations.

The chapter will be structured as follows: in section 3.1, I will introduce the thesis of the necessity of causal evaluations. In section 3.2, I will draw on the mimicry-synchrony model of emotional contagion (Hatfield et al., 1994) to explain the key tenets of the process of emotional contagion, clarifying why it poses a challenge for the thesis of the necessity of causal evaluations. In section 3.3, I will present a new account of emotional contagion that is

[41] Emotional contagion may pose another interesting challenge to emotion theories that assume that emotions are necessarily object-directed. This is because emotions set off by emotional contagion seem to lack a particular object. This point has been discussed by Robinson (2009), and I will not discuss it here.

consistent with the hypothesis of the necessity of causal evaluations. In section 3.4, I illustrate two main independent explanatory advantages of my account.

3.1. The Necessity of Causal Evaluations

Causal evaluations are evaluations involved in setting off emotions (Ben-Ze'ev, 2000, p. 72). They are evaluations of the situation inwhich the subject is in, according to the subject's concerns (see chapter 5 for an account of concerns). Many authors across philosophy and psychology seem to suggest that causal evaluations are *necessary* for setting off emotions. Cognitivist theories, for instance, claim that emotions are always the result of a belief (e.g., Nussbaum, 2001; Solomon, 1988). Appraisal theories also assume that emotions require evaluations, or "appraisals" in their terminology (Lazarus, 1991; Scherer et al., 1984; Smith and Ellsworth, 1985). Psychoevolutionary theories of emotions seem to support the idea that emotions are always the result of either an "appraisal mechanism", where the evaluation is the product of this mechanism, or an evaluative belief (e.g. Griffiths, 1997).[42] As a result, the emotion theories that assume the necessity of causal evaluations deny that fully-fledged emotions can be triggered by processes that do not involve causal evaluations (Robinson, 2009). Such processes may include emotional contagion, but also emotions set off by direct-

[42] Feeling theories would probably deny that causal evaluations are necessary for the triggering of an emotion. As William James puts it "the bodily changes follow directly from 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). According to one interpretation of James' theory, emotions are triggered without mediation of cognitive states. This interpretation of James has been questioned by authors such as Jenefer Robinson, who argue that the feeling theory may be compatible with the thesis that emotions may involve causal evaluations, or appraisals (Robinson, 2009). Even under Robinson's interpretation, it is not clear whether the feeling theory can endorse the claim of the necessity of causal evaluations, as this claim seems overly demanding for the feeling theory.

brain stimulation, drugs and other psychoactive substances such as coffee. In this chapter, I will focus only on the challenge from emotional contagion.[43]

It is possible to identify three main motivations for the thesis that causal evaluations are necessary for the triggering of an emotion: (1) causal evaluations explain the intelligibility of emotions; (2) causal evaluations explain differences in the emotional activation between different subjects; (3) causal evaluations account for the regularities between emotions and environment.

Causal evaluations explain the intelligibility of emotions because causal evaluations explain the extent to which something – let it be an object, an event or a state of affairs – has a certain degree of evaluative significance for a subject.[44] This can be illustrated with the following situation: suppose that Joanne is sad because her train has been cancelled and she will not be able to participate in a hiking trip she was craving. Causal appraisals are able to explain the activation of her sadness: this emotion is triggered because not being able to participate in the hiking trip has been evaluated as loss-involving. Explanations that appeal only to the mental state's phenomenology and bodily changes seem insufficient to fully explain the activation of emotion: a mere combination of bodily feelings and autonomic responses, such as cardiovascular and muscular activity, does not offer a complete explanation of why Joanne feels sad. One may want to explain the intelligibility of emotions by appealing to the emotions' cognitive bases. The cognitive bases are mental states, such as perceptions, beliefs, memory or imaginings, which ground the emotions, providing them with relevant information (Deonna

[43] See chapter 4 for a discussion of the rational assessability of emotions aroused by these means.

[44] This is only part of the explanation. We may further inquire about the reasons for the causal evaluations. To answer why-questions about the reasons for causal evaluations, it will be important to understand whether and how something may be significant for the subject, and for doing that, we also need to know what the subject's concerns are (see chapter 5 for a discussion of concerns).

and Teroni 2012, p. 5). For instance, Joanne's sadness for her cancelled train is based on her belief that she will not be able to participate in the hiking trip. Similarly, consider Steve, who fears a spider. His fear may be based on his perception of the spider, on imagining a spider, or on a belief that there is a spider. Which mental state acts as a cognitive basis for Steve's emotion depends on the actual situation. However, the emotions' cognitive bases also seem insufficient to explain why emotions are triggered. This is because we can easily think about scenarios where subjects such as Joanne and Steve have the same cognitive bases but lack an emotional experience. Without causal evaluations, we face an explanatory gap.

Different people react differently to the same situation. For instance, suppose that Max sees that his train to the Peak District has been cancelled and, as a result of this, he will not be able to go hiking with his friends. Instead of feeling sad, Max feels relieved (you may imagine that he was too embarrassed to admit to his friends that he was not fit enough for that particular hike). Suppose that Joanne and Max have the same[45] cognitive bases: they both believe that the train has been cancelled, and that they will not be able to participate in the hiking trip. However, they react differently to the same situation. Causal appraisals explain – at least in part – this difference: while Joanne evaluates the event as loss-involving, Max evaluates it as a relief from an embarrassing situation. A full explanation of this will require discussing the subjects' concerns (I discuss this issue in chapter 5).

Causal evaluations explain why different instances of the same emotion type are activated across different contexts. Different instances of fear, for instance, are all caused by an evaluation of something as dangerous. Similarly, different instances of sadness are caused by an evaluation of something as loss-involving.

[45] Two or more subjects have the same cognitive bases when their cognitive bases have the same contents.

3.2. THE CHALLENGE FROM EMOTIONAL CONTAGION

Emotional contagion is a psychological process where, via mimicry, one subject appears to experience the same type of emotion as another subject. Elaine Hatfield, John Cacioppo and Richard Rapson characterise emotional contagion as "the tendency to automatically mimic and synchronise facial expressions, vocalisations, postures, and movements with those of another person, and consequently to converge emotionally" (Hatfield et al., 1994, pp. 153-154).[46] The mimicry of somebody else's behaviour may range from motor movements, posture, vocal changes, speech gestures, laughter and yawning to what Prochazkova and Kret call "autonomic mimicry", which is convergence in the subject and other person's physiologies such as heartbeat, breathing rhythm and hormonal levels (Prochazkova and Kret, 2017). It has been argued that the mimicry can occur not only between two or more people, but also between two or more animals, and that vocal mimicry could happen between humans and animals (Briefer, 2018)[47]. It has also been argued that people may experience emotions as a result of emotional contagion by music (see, for instance, Davies, 2011; Justlin and Västfjäll, 2008; Robinson, 2005). However, the possibility of emotional contagion by music is still quite controversial, partly because it is unclear whether emotional contagion, in these particular situations, can arouse proper emotions.[48] In the following part of the chapter, I will use the neutral expression "target" to qualify the entity that is mimicked by a subject. I will leave open whether the target will be a person or an entity expressing emotion that is not a person (e.g. expressive music or non-human animals).

---

[46] Emotional contagion does not need to be initiated voluntarily.

[47] It would be interesting to know whether emotional contagion can occur between human-animal interactions, and between robots-human interactions. However, this area is still underexplored, and my argument does not depend on the result of this research.

[48] For criticisms against the possibility of emotional contagion by music, see Kivy (1990) and Robinson (2009).

Emotional contagion should not be confused with other similar psychological processes that may lead to emotion as a result of social interaction. Consider the following example: Max visits Joanne, who is sad because she cannot participate in the hiking trip she previously planned and craved for. Suppose that Max knows that Joanne coveted the trip for a long time and that she undertook an intense three-month training for it. He imagines that, if he were Joanne, he would be sad as well. In this scenario, Max's sadness may be considered the result of an empathic process (Goldie, 2000).

The term "empathy" embodies a multitude of different psychological processes and has been notoriously difficult to define. It is out of the scope of this chapter to provide an in-depth discussion on empathy. Here, I will only try to clarify why emotional contagion and empathy are different processes. Both emotional contagion and empathy involve two entities: a subject and a target, and a sort of "affective matching" between the subject's and the target's emotion (Coplan, 2011). Emotional contagion differs from empathy in a number of important respects: it is not an imaginative process (e.g. projecting oneself in the other's situation, or understanding what it is like to be in someone else's situation), and it does not require a clear awareness of the distinction between the subject and the target: the emotions are experienced not necessarily as somebody else's emotion, but may also be experienced as one's own (Coplan, 2011, p. 5; Preston and de Waal, 2002). More generally, we can understand emotional contagion as a primitive kind of process that is "relatively automatic, unintentional, uncontrollable and largely inaccessible to the conversant awareness" (Hatfield et al. 1994, p. 5).[49]

[49] By "inaccessible to the conversant awareness", Hatfield and colleague mean that the party involved in social interaction are not able to consciously access the process of emotional contagion.

The model of emotional contagion proposed by Hatfield and colleagues has been called the mimicry-synchrony model, and is composed of three stages, each of one described by a proposition. As they put it:

> "Proposition 1. In conversation, people tend automatically and continuously to mimic and synchronize their movements with the facial expressions, voices, postures, movements, and instrumental behaviour of others.
> Proposition 2. Subjective emotional experiences are affected, moment to moment, by the activation and/or feedback from such mimicry […]
> Proposition 3. Given proposition 1 and 2, people tend to "catch" others' emotions, moment to moment" (Hatfield et al., 1994, pp. 10-11).

According to the mimicry-synchrony model, the causal chain of emotional contagion proceeds as follows: firstly, the subject mimics the target; secondly, she is affected by the feedback of her behavioural changes; finally, subject and target converge emotionally, that is, they happen to experience an emotion of the same type. Hatfield and colleagues provided experimental evidence for their account (Hatfield et al., 1994).[50] Here, I do not have the space to discuss in depth the available empirical evidence for each proposition. I will rather focus on whether emotional contagion challenges the thesis of the necessity of causal evaluations.

Emotional contagion poses a challenge to the claim that causal evaluations are necessary because the emotions triggered as a result of emotional contagion seem to not involve any causal evaluation. In what follows, I consider each stage of the mimicry-synchrony model and I argue that none of the stages of emotional contagion described by the mimicry-synchrony model seems to involve a causal evaluation.

[50] More recent empirical data is reviewed by Prochazkova and Kret (2017).

According to Hatfield and colleagues, mimicry and synchrony initiate instantaneously and automatically.[51] The mimicry-synchrony model does not explain why we imitate, and it is unclear whether this tendency to imitate has a structural condition for its manifestation, that is, a certain property that grounds the disposition to imitate. This aspect is not explored by the model and the nature of the disposition to imitate is left underdeveloped. So, it may be objected that the instantaneity and automaticity of the process do not rule out the possibility of automatic evaluations at the first stage of the process. Indeed, appraisal theories are open to the possibility of automatic causal evaluations, e.g. (Clore and Ortony, 2000; Moors, 2010). Could causal evaluations always be involved in the first stage of the mimicry-synchrony model?

There are reasons to doubt so. The tendency to mimic and synchronise with the target seems to be present across people with different backgrounds, regardless of how the situation is evaluated in accordance with the subjects' concerns. There is evidence that subjects' emotions, moods and background beliefs may determine the extent to which subjects mimic targets, but they do not always determine the initiation of the mimicry. Hatfield and colleagues, for instance, present an empirical study conducted by McHugo and colleagues that shows that individual differences in emotions, values and background beliefs do not result in a difference in the extent to which subjects exhibit facial mimicry (McHugo, Lanzetta, Sullivan, Masters, and Englis, 1985). They consider a study in which college students that were classifying themselves either as Republicans or Democrats watched excerpts displaying President Ronald Reagan expressing happiness, anger and fear in a television show. Researchers obtained data regarding the facial automatic reactions of participants using facial electromyography. Subjects reported to have very different emotions in response to the video, and this changed depending

[51] It is currently debated whether this tendency is innate (Meltzoff and Moore, 1977) or acquired (Oostenbroek et al., 2016).

on their previous attitude towards Reagan. However, both Republicans and Democrats mimicked Reagan's changes in facial expressions (Hatfield et al, 1994, pp. 24-26). Even if students reported that they had different emotions as a result of seeing the excerpts, their facial expressions were synchronised with Reagan. So, it does not seem that causal evaluations were contributing to the activation of the mimicking process. We can therefore conclude, at least in this case, that there were no causal evaluations involved at the first stage of the process. We may need additional data to understand how generalisable this result is. Nevertheless, the thesis that causal evaluations are necessary for initiating mimicry seems doubtful.

The mimicry-synchrony model does not assume any causal evaluation at the second stage: all that the model says is that emotional experiences are affected by the feedback from the facial, verbal and postural mimicry. Similarly, the model does not seem to suggest the thesis of the necessity of any causal evaluation at the third stage. According to the model, the emotion is triggered by the target's behaviour alone.

I conclude that the emotions experienced as a result of emotional contagion pose a challenge to the thesis of the necessity of causal evaluations. This is because emotions aroused as a result of emotional contagion appear not to be triggered by any causal evaluation. Some have argued that these affective states are not "proper" emotions (Kivy, 1990, 2002; Robinson, 2009).[52] In what follows, I answer the challenge from emotional contagion by defending a new strategy that is available for theories of emotions that defend the thesis of the necessity of causal

[52] This is especially true in the context of a discussion about the possibility that music sets off emotions as a result of emotional contagion. One of the main reasons for arguing against the possibility that music triggers emotions as a result of emotional contagion is that emotions triggered as a result of emotional contagion seem to lack causal evaluations, or causal appraisals.

evaluations. My strategy consists in offering a modification of the mimicry-synchrony model, suggesting that causal evaluations are triggered by an associative mechanism.

3.3. My Proposal: Associations and Causal Evaluations

In this section, I argue that the mimicry-synchrony model can allow for causal evaluations to be present in the emotional contagion process, and I offer an account of how they are generated.[53] The theories that defend the necessity of causal evaluations would be able to use my account as a theoretical model that looks promising for addressing the challenge from emotional contagion.

Firstly, we need to consider the format of the causal evaluation that may be involved in emotional contagion. Since the emotional contagion process is automatic,[54] it is plausible to think that, if there were a causal evaluation involved in the emotional contagion process, it would be conceptual but not propositionally structured (see chapter 2 for a distinction between conceptual and propositional content). The function that I am going to offer here will thus individuate the associative mechanism that is responsible for the activation of causal evaluations with such a structure.

---

[53] Perhaps my solution to the challenge could be applied also to solve the challenge from emotions set off by drug and brain stimulation, but I will not discuss this here.

[54] Emotional contagion is an automatic process. Subjects may voluntarily control the initiation of the emotional contagion process by employing an emotion regulation strategy such as "situation selection". This strategy has been described by James Gross (2015). Subjects may voluntarily put themselves in a situation where they know there will be people expressing emotions, to increase the likelihood that emotional contagion is automatically initiated. For instance, a subject who wants to experience happiness as a result of emotional contagion may decide to join a birthday party with their cheerful friends.

I will now characterise the associations that a defender of the thesis of the necessity of causal evaluations can use to meet the challenge from emotional contagion. These associations are between an experience of a certain pattern of bodily changes and an affective evaluative property. The associations I am going to propose are not arbitrarily individuated, but they are formed in a particular way: they are individuated by a certain function. In what follows, I present the function that individuates associations and I argue that these associations are responsible for triggering causal evaluations.

Let *F* be a function that identifies associations, and let *S* be the set of all bodily states. Let *U* be the set that has as elements R_i, which are sets of bodily states[55] that have a certain property, namely to occur when the subject undergoes each basic emotion (see *Figure 1*, p. 85). A detailed characterisation of the types of bodily changes that occur when a subject undergoes a particular emotion has been given by researchers such as Ekman and colleagues (Ekman, 1972, 1977; Ekman, Levenson, and Friesen, 1983; Nummenmaa, Glerean, Hari, and Hietanen, 2014). Moreover, let *P* be the set of affective evaluative properties that are related to each basic emotion. For instance, assume that the elements of the set R_1 of bodily states that occur when a subject undergoes the emotion of anger are sweating and increased heart-rate.[56] Let *P* be composed of affective evaluative properties (I discuss affective evaluative properties in chapter 5). In this chapter, I have relied on the claim that emotional contagion triggers only "basic emotions", which are, as Griffiths suggests, the emotions of fear, anger, disgust, surprise, sadness and joy (Griffiths, 1997). I will discuss Griffiths's account of basic emotions in more

[55] In figure 1, for instance, the elements of U are R_1, R_2, and R_3.

[56] In a real case scenario, the bodily states that occur when a subject experiences anger are more than the above suggested ones. But let's assume that there are only two, for the sake of simplicity. Once that the mechanism I am suggesting has been understood, it will be easy to generalise it to more complex cases.

detail in chapter 7. So, the elements of *P* would be, respectively, danger, offence, noxious stimuli, unexpected events, irrevocable loss, and the progress towards the fulfilment of a goal (Lazarus, 1991).

I define the function *F* that associates to each element of *U*, one and only one element of *P*, according to the following rule: an element R_i is associated with an element P_i if all the bodily states that belong to R_i have been previously paired with P_1. For instance, in the toy model represented in *Figure 1* (p. 85), the bodily states that belong to R_1 are c_1 and c_2. If c_1 and c_2 have been previously paired with p_1, then if c_1 and c_2 occur, p_1 will occur too. Now, in what circumstance may c_1 and c_2 have been paired with p_1? In previous occasions, where c_1 and c_2 were activated by the causal evaluation of p_1.

For instance, assume that the elements of the set of bodily states R_1 that occur when someone experiences anger are sweating and increased heart-rate. Assume also that the set of affective evaluative properties *P* is composed of the following three elements: danger, offence, and loss. The function *F* associates to R_1 one of the three above-mentioned evaluative properties if all the bodily states that belong to R_1 have been previously paired with that evaluative property. Assume that, when we experience sadness, we do not usually have any increase in heart-rate and we do not sweat.[57] If this is the case, R_1 will not be associated with the value "loss". Assume that when we experience anger, we do manifest an increase of heart-rate and sweating, instead. In this case, R_1 may be associated with the affective property "offence".

The proposed model presents a rule for identifying associations between patterns of bodily states and evaluative properties. We can consider these associations to trigger proper causal evaluations. This is because what is triggered is an evaluation of a certain situation – the

[57] Again, here I am trying to simplify the situation for the sake of clarity, but I do acknowledge that the situation may be different.

situation in which the subject is in – which is relevant to the well-being of the subject. The function gives us a way to characterise the causal evaluations that explains why we have an emotion as a result of emotional contagion. Note that there may be bodily changes – like c_{11} in the picture or sets of bodily changes that are not associated with affective evaluative properties.

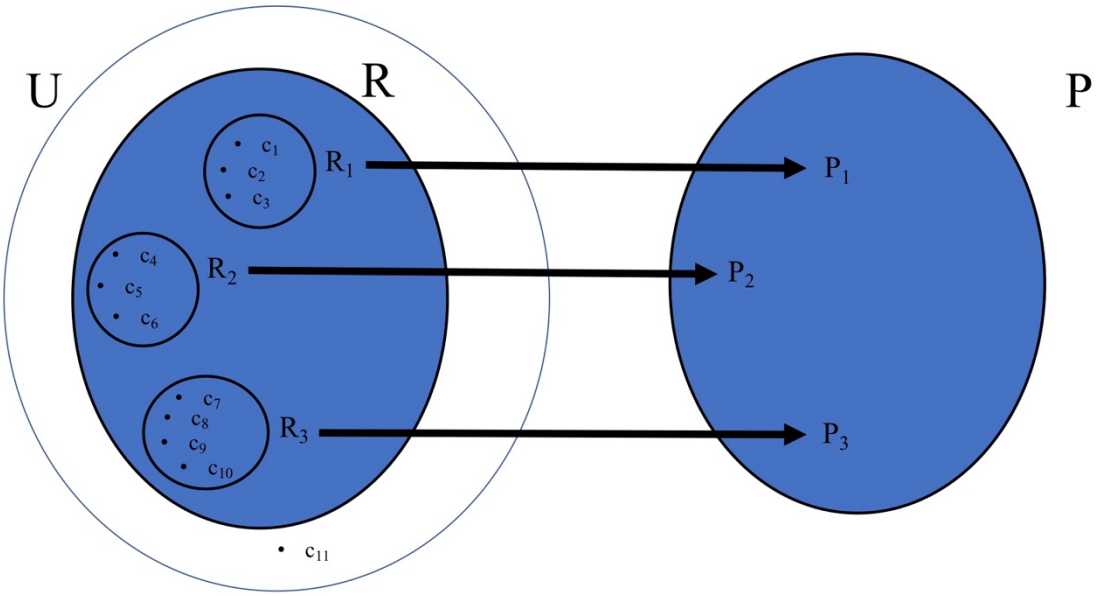

Figure 1: This figure represents the function that generates appraisal in my account.

An interesting question, which it is not possible to fully answer without further empirical study, is whether these associations vary between subjects. What we currently know about patterns of bodily changes associated with each emotion render plausible the idea that these associations are subject to only a very minor variation between subjects. Indeed, Ekman and Cordaro (2011) argued that typically the patterns of bodily changes associated with each emotion are similar across different subjects; only subjects experiencing traumatic life events during their development seem to have unusual levels of activations.

My proposal will be attractive to theories of emotions that accept that there are patterns of bodily changes associated with each emotion and that these patterns of bodily changes are pancultural. Psychoevolutionary theories, for instance, would share this assumption

(DeLancey, 2004; Griffiths, 1997). However, my account will be less attractive to theories that are sceptical about the possibility of associating patterns of bodily changes to each emotion type such as constructivists, like Lisa Feldman Barrett (Barrett, 2009, 2017).[58]

To sum up, a subject, say Joanne, sees another person, call him Mark, who is expressing an emotion. Suppose that, in this case, Mark is scared. According to the mimicry-synchrony model, when Joanne sees that Mark is scared, she automatically mimics his bodily changes, e.g. his fearful expression and posture. She also mimics his internal physiology. Suppose that Joanne experienced fear before. In previous instances of fear that were not resulting from emotional contagion, she experienced a configuration of bodily changes R_1 that is similar to the one she is experiencing now. Previous instances of fear were triggered by causal appraisals of danger. So, she learned to associate R_1 with danger, and she formed a disposition to transition between the experience of bodily changes R_1 and danger. As a result of mimicking Mark, Joanne's configuration of bodily changes ends up being similar to the configuration that, in her previous history, was associated with danger. The causal evaluation "there is danger" is triggered via association and she feels fear.

"There is danger" is an evaluation because it assesses the impact of the situation for the subject's well-being: in particular, in this case, it is presenting the situation as being dangerous. It is a *causal* evaluation because it plays a role in triggering fear. It may be argued that this evaluation is not genuine because it is the mere result of a pairing process. However, there are many other emotions that are triggered by causal evaluations that are the result of associative processes. Some instances of recalcitrant emotions, that is, emotions that persist despite the subject's better beliefs, may be examples of such cases. Consider, as an example, Alice's fear of stepping on cracks on the pavement. She knows that there is nothing dangerous in doing so,

[58] Discussing this debate is outside of the scope of this chapter.

but when she sees a crack, she feels fear. One possible explanation of Alice's emotion is that Alice developed an association between cracks and danger. Perhaps Alice survived an earthquake where the danger (the earthquake) was paired with cracks (the pavement of her house was cracked during the earthquake). In her history, cracks have been paired with events that negatively affected her well-being. As a result, an association between cracks and danger has been built. So, when she sees this crack in the pavement, the representation of danger is activated and, she automatically evaluates the crack as being dangerous. Causal evaluations involved in emotional contagion seem similar. So, whoever denies that causal evaluations in emotional contagion are genuine should also deny that Alice's emotion has non-genuine evaluations too. However, doing so seems revisionary. Emotions like Alice's have not been used as a counterexample to the necessity of causal evaluation thesis precisely because there seem to be a causal evaluation involved in setting off these emotions, even if such an evaluation is the result of an associative process. Therefore, I suggest that the evaluations I am proposing are genuine. Like causal evaluations that are not the product of associations, the proposed evaluations are responsible for setting off emotions and they also assess the impact of the situation for the subject's well-being.

In ordinary emotional experiences (see *Figure 2*), bodily changes do not precede the triggering of causal evaluations but follow them. However, when emotions are triggered by emotional contagion, mimicry causes bodily changes, and these changes *precede* causal evaluations (see *Figure 3*). Once that the emotion has been set off by emotional contagion, further bodily changes may also occur. For instance, once Joanne begins to experience fear as a result of emotional contagion, her heartbeat may accelerate further, and she may start experiencing shortness of breath.

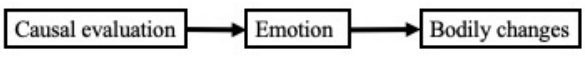

Figure 2: Emotion triggering in ordinary cases

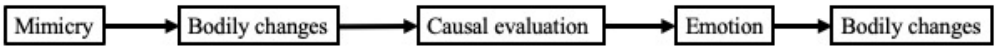

Figure 3: Emotional contagion

In my account of emotional contagion, the proximal cause of the emotion is a causal evaluation – such as the evaluation "there is danger!". This evaluation is able to explain why emotions are triggered via emotional contagion. For instance, in the example above, fear has been triggered: and this is because danger has been detected.[59] The emotions set off by emotional contagion are generated through genuine evaluations of the impact of the situation for the subject's well-being. However, there is still a sense according to which emotions aroused by emotional contagion are "vicarious", that is, they seem to be somebody else's. The bodily changes that are generated at stage I of the mimicry-synchrony model are still initiated by imitation, and not by an appraisal of the actual situation.

The class of causal evaluations that, in my account, are triggered in cases of emotional contagion seem to lack an object. It does not seem plausible to consider that the evaluation is *about* the pattern of bodily changes. When experiencing an emotion triggered by emotional contagion, we are not concerned about our bodily changes. The evaluation is triggered by the association between a pattern of bodily changes and the presence of danger but is not *about* these changes.

The object of the emotion may be different from the object of the target's emotion. For instance, consider again Joanne, who experiences fear as a result of emotional contagion: she does not seem to be scared about her bodily changes, and her causal evaluation seems not to

[59] Another interesting question regards whether the causal evaluation is correctly representing the situation. However, for space reasons, I will not address this issue here.

be about them. Similarly, there is no guarantee that the object of causal evaluations will be the same as the target's emotion. Joanne may feel an emotion via emotional contagion of Mark's emotion without having any knowledge about the intentional object of Mark's fear. However, the causal evaluation seems to be directed at *something*, which is taken to be in the "outside" world.

In my proposal, the causal evaluations that activate emotions through emotional contagion are not propositionally structured. Theories of emotions that argue that full-blown propositionally-structured beliefs are necessary for the activation of the emotions may not be able to employ my account as it is. Examples of such accounts are cognitivist theories (Nussbaum, 2004; Roberts, 2003; Solomon, 1977, 1988).

However, cognitivists may also be able to employ my solution, adding a slight modification of my account. Associations may subsist between bodily changes and propositionally structured evaluations. Cognitivists could argue that in cases of emotions triggered by emotional contagion, associations are not between bodily changes and an evaluative property, but between bodily changes and evaluative beliefs.[60] In this way, these evaluative beliefs act as causal evaluations, in a way that is consistent with cognitivist accounts.

3.4. Advantages of my Proposal

This account provides two independent advantages: (1) it can explain why, in some cases, mimicry is not sufficient to set off a convergence in emotion between subject and target and (2) it can provide an answer to the challenge from voluntary emotions that could be accepted by theorists who support the necessity of appraisals. I explain these advantages below.

---

[60] I am not the only one to suggest that associations may subsist between propositionally structured mental states. The possibility that associations may subsist between propositionally structured mental states has been discussed also by Eric Mandelbaum (2016).

The mimicry-synchrony model encounters difficulties in explaining why and how in some cases we mimic the target's emotional state, but we do not end up having an emotion of a similar type as a result. According to the model, it is problematic to explain what grounds this difference. This is because if we accept that we do have a tendency to mimic, and the feedback from the bodily changes is responsible for setting off the emotion, we would expect that an emotion of the same type will be the end result of the mimicry process. However, sometimes we mimic, but we do not respond with the same emotion. Cases like this have been explored by Dezecache and colleagues (Dezecache et al., 2015). As Dezecache and colleagues argued, the mimicry-synchrony model is wrong to assume that emotional contagion is always a replicative process. According to them, this assumption is simplistic because it does not take into account more complex relations between the subject and the mimicked target, e.g. whether the two parts are in competition or in cooperation. Dezecache illustrates some research directions that may serve to explain why emotional contagion propagate in some cases and not in others (Dezecache, Jacob, and Grèzes, 2015).

My model could contribute to this research explaining how the mechanism that underlies emotional contagion sets off an emotion in some circumstances and not in others. In particular, the associations that I proposed here explain why sometimes the feedback from the body causes the emotion *e* to activate, and sometimes it does not. When the mimicry causes a subject *S* to have an internal state, which matches R_i and the association between R_i and p_i is already stored in the subject's memory, the causal evaluation is triggered and the emotion e_i arises. On the other hand, when the bodily changes do not match the bodily changes that belong to the set R_i, the causal evaluation does not occur, and, as a result, the subject does not experience any emotion. Imagine, for instance, that Sam is very nervous for a job interview. His anxiety makes him very irritated about any noise he hears in the office. When his colleague Emily enters the room and laughs with her friend, instead of being "infected" by Emily's happiness, Sam starts

to be angry at her. Even in this situation, Sam could have mimicked some of Emily's emotional expression and bodily posture. However, given that Sam did not end up experiencing happiness, we could suppose that the bodily changes that Sam registered were not enough to match the ones that usually occur with happiness.

Another advantage of my account is that it can be used to solve another potentially problematic phenomenon for the thesis of the necessity of causal evaluations: the voluntary arousal of emotions. Actors, for instance, may try to voluntarily entertain the posture, the tone of voice and the expression of a person who feels sad, and they may end up feeling sad as a result of this. Levenson and colleagues' experimental results illustrate that voluntarily producing a certain facial configuration for basic emotions may generate levels of subjectively experiencing these emotions (Levenson, Ekman, and Friesen, 1990). This evidence may be prima facie problematic for theories that assume the thesis of the necessity of appraisals. When an actor voluntarily exhibits a certain facial expression of sadness, for instance, she knows that she is not experiencing any personal loss, since she is aware that she is not her character. However, she may feel sad. The account for individuating appraisals can also be used to explain such cases.

One of the assumptions of my proposal is that we can consider causal evaluations as evaluations of the impact that the situation has on our well-being. It may be argued, though, that the associations I am suggesting are not proper causal evaluations because causal evaluations are usually directed at objects in the world and not at bodily changes internal to the subject. There are reasons to think that bodily feedback can affect our well-being: physical illness, tensions in our body and pain may impinge on our wellbeing. It seems plausible to think that other feelings may also be accessed by an appraisal system that generates causal evaluations. And even if our bodily states did not really affect our well-being, this would not imply that associations involving our bodily changes are not appraisals. Causal evaluations

may incorrectly represent the relevance of that particular bodily state for our well-being, and these associative causal evaluations generated by the proposed function may be incorrect.

3.5. Summary and Conclusion

In this chapter, I showed that theories of emotions that assume the necessity of appraisals face a challenge from emotional contagion. I argued that this challenge could be met by revising the mimicry-synchrony model of emotional contagion. I suggested that causal evaluations also occur in cases of emotional contagion and I presented an account of the associations that trigger causal evaluations. I argued that my account is at least consistent with most non-cognitivist theories of emotions, and that, with a slight modification of my account, also cognitivist theories can use my solution to the challenge from emotional contagion.

In the first section of the chapter, I argued that the necessity of causal evaluations is interesting, and it delivers important explanatory advantages. I then presented the mimicry-synchrony model and I presented the challenge from emotional contagion (section 3.2). In section 3.3, I presented my own tentative solution to the challenge from emotional contagion. I argued that the mimicry-synchrony model can allow causal evaluations to be present in the emotional contagion process, and I suggested a function that individuates the associations that generate the causal evaluations that are responsible for setting off emotions as a result of emotional contagion. In section 3.4, I presented two main independent advantages of my account: it can explain why emotional contagion may not trigger emotions, and my account could provide an answer to the challenge from voluntary emotions.

This result is significant for the emotion literature because it improves our understanding of emotional contagion and it offers a better explanation of why emotions may be triggered as a result of emotional contagion. Moreover, my account is significant for current debates regarding the necessity of causal evaluations because it offers a new defence of the necessity of causal evaluations from the challenge from emotional contagion.

CHAPTER 4: THE RATIONAL ASSESSABILITY OF EMOTIONS

ABSTRACT

Sometimes we feel emotions that are disproportionate or do not respond to the evidence. We get angry despite knowing that no offence is intended. We fear things even if we believe they are not dangerous. It is tempting to consider these emotions irrational. Although the issue of whether emotions can be epistemically rationally assessable has been addressed before (see, for instance, Helm 2001; D'Arms and Jacobson 2003; D'Arms 2005; Döring 2015; Deonna and Teroni 2012; Brady 2009), there is no systematic argument for the rational assessability of occurrent emotions that fits with the epistemic rational assessability of mental states in general. The aim of this chapter is to provide a set of criteria for the epistemic rational assessability of mental states and show that most emotions satisfy these criteria.

In this chapter, I present four conditions for the epistemic rational assessability of mental states, using a reflective equilibrium methodology. I argue that a mental state *m* is epistemically rationally assessable if and only if the following conditions are met: (1) *m* is informationally plastic; (2) the mechanism *P* that maintains *m* is generally receptive to the evidence; (3) *m* has a descriptive content; (4) *m* contributes to the subject's view of the world.

INTRODUCTION

Recall the first scene of Shakespeare's tragedy *King Lear*, that I illustrated in chapter 1. In this act, the old King, looking for flattery, asks his daughters to express their love for him. While the King's eldest-born Goneril and his second-born Regan praise their father, exaggerating their affection, the third-born Cordelia refuses to meet his request. Instead of singing his praises, as the other sisters did, Cordelia answers sincerely that her love does not need to be

supported by flattering words. Listening to this answer, King Lear reacts in a wave of anger and pride, vehemently disclaiming his parental care for Cordelia. Despite the attempts of the Earl of Kent to make King Lear understand his overreaction, the King stands firm in his position.

The Earl of Kent, and also the audience of Shakespeare's tragedy, think that not only is King Lear's action inappropriate but also that his emotion of anger is utterly disproportionate. More precisely, King Lear's anger appears to be epistemically irrational: the content of his anger is not well supported by the evidence. There is nothing offensive in Cordelia's speech that justifies such tremendous rage towards her.[61]

We know that sometimes, like King Lear in Shakespeare's tragedy, we feel emotions that are disproportionate or that do not fit with the evidence.[62] We get angry for no reason. We fear things despite knowing that these things are not dangerous. We feel guilty for actions even though we know that we could not have done otherwise. It is tempting to consider these emotions irrational. And indeed, the idea that emotions can be epistemically irrational has gained popularity in the last twenty years (see, for instance, D'Arms, 2005; Solomon, 2007; Nussbaum, 2004). However, as I have explained in chapter 1, the rational assessability of emotions, and in particular the emotions' *epistemic* rational assessability is difficult to explain.

According to cognitivists, the rational assessability of emotion is explained by appealing to the rational assessability of the belief that is the necessary component of the emotion. However,

---

[61] It is possible to identify reasons that explain why the King reacts in this way – e.g. he wanted flattery. However, being able to understand the King's anger does not make his anger less epistemically irrational. As Lisa Bortolotti shows, irrational cognitions do not always compromise the possibility to understand or explain the subject's behaviour (Bortolotti, 2018).

[62] Although the term "emotion" is sometime used in English to refer to occurrent and dispositional states, in this chapter, when I use the term "emotion" I only refer to occurrent emotions.

cognitivist accounts face difficulties when explaining the persistence of recalcitrant emotions, which are emotions that conflict with our evaluative judgment (D'Arms and Jacobson, 2003). I will come back to this issue again in chapter 7. Cognitivist accounts also face difficulties in explaining the irrationality involved in recalcitrant emotions. Cognitivists assume that recalcitrant emotions necessarily involve an evaluative belief. Thus, according to cognitivists, if recalcitrant emotions exist, they involve two contradictory beliefs. This raises a problem: either the subject is irrational because of having two contradictory beliefs, and this seems counterintuitive, or recalcitrant emotions seem impossible. Non-cognitivist accounts are better able to explain why recalcitrant emotions persist. (On this point, see also chapter 7, where I discuss a non-cognitivist explanation of the persistence of recalcitrant emotions). However, non-cognitivists struggle to explain the rational assessability of emotions (Helm 2001, p. 42; Tappolet 2016, p. 31; Benbaji, 2013, p. 580).[63] They cannot explain the epistemic rational assessability of emotions appealing to the rationality of the belief or the judgment that is a constitutive element of the emotional experience.

As I claimed in chapter 1, there is no systematic argument for the rational assessability for instances of emotions that is compatible with both cognitivist and non-cognitivist theories of emotions and fits with the epistemic rational assessability of mental states in general. In this chapter, I aim to fill this gap.

I will adopt a new strategy for identifying the epistemic rational assessability of emotions. I will not investigate whether emotions, *as types*, are epistemically rationally assessable. I take the epistemic rational assessability to be a property that is not intrinsic to the psychological type, but it is intrinsic to the particular mental state's *occurrence*. In this chapter, I identify the conditions for the epistemic rational assessability that apply to *occurrences* of mental states. In

[63] See chapter 1, section 1.5.3.

this way, the epistemic rational assessability of emotion occurrences will fit with the epistemic rational assessability of other mental states. The aim of this chapter is to present four conditions for epistemic rational assessability of mental states and to show that most emotion occurrences fit this requirement.

In section 4.1, I outline the central assumptions of this chapter. In section 4.2, using a reflective equilibrium methodology (see, for instance, Goodman, 1954; Rawls, 1999), I present my account of the epistemic rational assessability of mental states. Finally, in section 4.3, I discuss the problem of another account of the epistemic rational assessability by Kate Nolfi (Nolfi, 2015).

4.1. Norms and Rational Assessability

Mental states can be subject to different types of norms. Depending on the context, we can identify, for instance, norms of morality, aesthetics, and rationality. Here, I focus only on rationality norms.

Consider an epistemically rationally assessable mental state *m* of a certain type. Norms of mental states distinguish between mental states that comply with the norms and mental states that do not comply with them. In our everyday discourse, we consider some mental states to be obviously rationally assessable, whereas we do not consider other mental states to be so. Paradigmatic examples of rationally assessable mental states are beliefs, and examples of a-rational mental states are perceptual experiences[64] and bodily sensations, e.g. feeling itchy,

[64] In this chapter, I am going to adopt the assumption, widely shared in the philosophy of emotions literature, that perceptual experiences are *a-rational*. However, I am aware that Susanna Siegel offered an influential account according to which perceptual experiences are epistemically rationally assessable (Siegel, 2017). See also footnote 65.

dizzy, tingling, hungry etc. Other mental states that are neither obviously epistemically rationally assessable, nor obviously a-rational, are moods.

One possible strategy for developing an account of the epistemic rational assessability of mental states may be to use belief as a prototypical example of an epistemic rationally assessable mental state. According to such a strategy, it would be necessary to investigate the epistemic rational assessability of beliefs first, and after such investigation, it will be possible to develop a more generalised account of the epistemic rational assessability of mental states that could also be applied to the emotions. Although this is a respectable strategy, it has the drawback of rendering the minimal criteria for the rational assessability of mental states difficult to capture. This is because beliefs are usually understood as propositionally structured mental states. Modelling the rational assessability of mental states only on the rational assessability of beliefs may mislead us into thinking that an over-demanding feature is central for the rational assessability of emotions (e.g. having a propositionally structured content), even when a "thinner" and a less demanding alternative is available. In what follows, I do not pursue this strategy. I employ a reflective equilibrium methodology instead (Goodman, 1954; Rawls, 1999). In the following section, employing this methodology, I present intuitions about paradigmatic cases of epistemically rationally assessable mental states and a-rational mental states, and I formulate a theory that accounts for our intuitions, identifying the necessary properties of epistemically rationally assessable mental states. I check back and forth whether intuitions and theory agree and ascertain the coherence with other background beliefs until an equilibrium is reached.

4.2. The Epistemic Rational Assessability of Mental States

In this section, I present four conditions for the epistemic rational assessability. While the first two conditions regard the aetiology of the mental state, the second two conditions regard properties of the mental state once it has been triggered.

4.2.1. Informational Plasticity

Compare paradigmatic instances of a-rational mental states with rationally assessable ones. A-rational mental states and rationally assessable mental states seem to have a different aetiology. One common feature of a-rational mental states seems to be that they are formed as a result of a process that displays some sort of rigidity. Conversely, central instances of rationally assessable mental states seem to be formed as a result of flexible processes. My belief that a glass of water is on my right may be formed as a result of many different inputs. For instance, it could be formed as a result of my visual perception of the glass, as a result of my remembering that such glass was on my right one minute ago, or as a result of your testimony that the glass is on my right. Compare this with my feeling thirsty. To feel thirsty, I have to receive a specific input, something like a "drinking signal" from my brain (Vineet et al., 2018). I cannot infer thirst, nor can I feel thirsty solely as a result of testimony or as a result of perception. Reading that I am thirsty or being told that I am thirsty will not make me thirsty. Feeling thirsty is, in most cases, the result of a fixed process. On the contrary, beliefs seem to display some sort of flexibility.

Perceptions are also mental states that are often considered to be a-rational and the product of a fixed process. As an illustration, consider a subject who sees the Müller Lyer illusion. She sees two lines as being different in length, even if she is aware that they are the same length. Her belief about the actual length of the two lines does not modify her perception, even when she tries to change her visual perception so that it corresponds with her background knowledge. If perception is indeed the product of a fixed process, my account suggests that it is a-rational.[65]

[65] Recent understanding of perception may challenge the folk conception of perception outlined here. According to novel accounts of perception (see, for instance, Pinker, 1998), the relation between the input and the final perceptual experience is not explicable only by a mere disposition to respond. The stimuli impinging on the retina

Christine Tappolet and Carolyn Price argued that flexibility is one of the key differences between emotion and perception: while perception is rigid, as the Müller Lyer case shows, emotions can be trained over time and are more sensitive to socio-cultural information (Tappolet, 2016, pp. 30-36). Tappolet suggests that plasticity is the key factor that grounds the rational assessability of emotions (Tappolet, 2016). According to her, emotions can be influenced by a wide range of mental states, such as expectations, beliefs, desires, imaginings and preferences (Tappolet 2016). For Tappolet, this difference explains why our perceptions are not rationally assessable, but our emotions are so.

In this section, I use Zenon Pylyshyn's account of *informational plasticity* to explain what grounds the flexibility involved in the aetiology of epistemically rationally assessable mental states. I firstly explain Pylyshyn's account of informationally plastic *behaviour*, then extend his account to informationally plastic *mental states*. Finally, I argue that a necessary condition for the epistemic rational assessability of mental state is their informational plasticity.

According to Pylyshyn, a behaviour can be informationally plastic when its activation is not the mere result of a rigid causal process, which is not sensitive to different sources of information (Pylyshyn, 1980, p. 120). Behaviour is informationally plastic, in Pylyshyn's sense, when the behavioural response can be produced by a variety of stimuli and may differ

---

is often ambiguous: the same stimulation may be consistent with almost infinite possible surface arrangements. Perception, then, needs to be mediated by representational states. A different understanding of perception may lead to a novel understanding of its rational assessability. Susanna Siegel, for instance, assumes that perception is the output of an inferentially mediated process and argues for the rationality of perception (Siegel, 2017). This claim may be compatible with my account. If perception were the result of an inferentially mediated process, it would satisfy the condition of informational plasticity. In my account, if all conditions presented in section 4.2 are met by a token of a perceptual state, such perceptual state will be rationally assessable. Cognitively penetrated perceptual experiences may satisfy these requirements.

when new information is integrated. Pylyshyn explains the notion of informational plasticity with the following example: "seeing that the building you are in is on fire, smelling smoke coming in through the ventilation duct, or being told by telephone that the building is on fire, can all lead to similar behaviour, and this behaviour might be radically different if you believed yourself to be performing a play at that time" (Pylyshyn, 1980, p. 121). Compare this with an example of a rigid process, such as the three-spined stickleback's behaviour in spring, as described by Niko Tinbergen (Tinbergen, 1974). Tinbergen observed that every time a male stickleback saw another fish with a red underbelly, it attacked the fish. This attacking behaviour can be explained as a brute disposition to respond to a certain stimulus – the perception of a red underbelly. This behaviour is executed even if the fish does not resemble a male stickleback at all. The stickleback is not able to modify its behavioural response so that it integrates other sources of information, such as the shape of the fish. Conversely, when I am told over the phone that the building I am in is on fire, a number of alternative actions are available to me. I may ignore the testimony or run out of the building. My behavioural response may also change depending on my background knowledge: if I have just been told that the announcement is a fire drill and that there is no need to undertake any action, I will probably keep calm, and remain in the building. If I believe that the announcement is not a drill, and I do not have reasons to doubt the reliability of the testimony, I will probably proceed to the designated fire assembly point. In general, my behaviour cannot be described as a brute disposition to respond to the event of hearing the phone ring or of answering the phone.

Since here I am concerned about the rational assessability of mental states, and not about the rational assessability of the behaviour that follows a mental state, I extend Pylyshyn's account of informational plasticity to *mental states*. Saying that a mental state *m* is informationally plastic means stating certain facts about the aetiology of *m*, namely that *m*'s aetiology is informationally plastic. A mental state is informationally plastic when (i) the

mental state can be acquired by several different processes, and (ii) the same mental state's aetiology could give rise to a different mental state if new information is integrated. Most instances of beliefs, for instance, are informationally plastic. Reading that today is sunny, being told by your friend that today is sunny, and seeing that today is sunny are all inputs that may lead you to believe that today is sunny, but this might be different if you are reading the weather forecast in an unreliable newspaper, if you know that your friend is speaking sarcastically, or if you know that what you are seeing is an illusion. This flexibility is consistent with the intuition that beliefs are paradigmatic examples of rationally assessable mental states. On the contrary, physical sensations, such as feeling thirsty, are generally not regarded as displaying informational plasticity. They seem the output of a fixed process. Again, this is consistent with the intuition that physical sensations, such as feeling thirsty, are not informationally plastic. I can, therefore, state the first condition of rational assessability of mental state:

- *C_1:* A mental state *m* is epistemically rationally assessable only if it has an informationally plastic aetiology.

Condition one predicts that beliefs are epistemically rationally assessable, but that physical sensations are not. However, it also predicts that some occurrences of beliefs are not rationally assessable. Suppose that a neuroscientist mechanically induces belief *p* in me, or that an evil daemon zaps me to believe *p*. In my account, these two instances of belief are not epistemically rationally assessable because they are not the output of a plastic aetiology. This is because the aetiology of your belief *p*, in the zap scenario, does not give rise to a different mental state if new information is integrated. So, the belief *p* in the zap scenario is not epistemically rationally

assessable. I consider this to be a strength of the theory: mechanically induced mental states do not seem epistemically rationally assessable.[66]

Emotions display different degrees of complexity, and this may go in hand with different degrees of informational plasticity. Most of the emotions may be informationally plastic.[67] They could be acquired by different processes, and the same aetiology may give rise to a different emotion if new information is integrated. Suppose that a subject experiences fear of arriving late to a meeting, which has been scheduled for 10 a.m. This fear could have been acquired by different processes, e.g. being told that is already 10 a.m., reading that it's 10 a.m., seeing that there is unusual traffic. The same aetiology could have led to a different emotion if the subject knew that the person was speaking sarcastically, the clock was 20 minutes fast, and that the meeting had been postponed.

At the low end of the spectrum of emotion's complexity, there are startle responses. Startle responses are automatic and intense bursts of fear of sudden or seemingly dangerous stimuli (Simons, 1996). Consider, for instance, startle reactions caused by very loud noises, such as a gunshot.[68] Paul Ekman and colleagues observed that the startle response rigidly follows the perception of sudden loud noise (Ekman et al., 1985). In their experiment, participants were asked to inhibit their startle response to the noise of a gunshot. The subjects knew when the shot was going to happen, and they also knew that the situation was not dangerous.

---

[66] According to one – here simplified – version of process reliabilism, a mental state's formation process is rational only if the mental state is caused by a reliable cognitive process (Goldman, 2012). This claim, as it is, is consistent with the claim that a mental state caused by a non-cognitive process (such as the demon's zap) is a-rational.

[67] Here I am appealing to our everyday experience. Further research may investigate the frequency of emotion with informationally plastic aetiology.

[68] There is disagreement about whether the startle response is an emotion. The thesis that the startle response is an emotion has been defended by Silvan Tomkins (1962). See also Ekman, Friesen, and Simons (1985).

Nevertheless, the startle response occurred. We can, therefore, infer that startle responses do not have an informationally plastic aetiology: they rigidly follow the registration of a certain stimulus. Startle reactions would then be a-rational in my account. This is in line with our intuitions. We rarely attribute rationality or irrationality to startle reactions.

Antonio Damasio describes cases in which emotions are triggered by direct stimulation of the brain (see, for instance, Damasio, 2004). Emotions triggered by direct stimulation of the brain enjoy a very low degree of flexibility. Again, these emotions do not qualify for rational assessment because they are not the result of an informationally plastic aetiology. The same holds for emotions that can be entirely explained by the effects of drugs. As an illustration, consider Lysander's belief that Helena is the "worthier maid" in Shakespeare's *A Midsummer Night's Dream*. Lysander's belief is the result of his contact with an enchanted love potion. In the play, a person whose eyelids come into contact with this potion will fall in love with the next living creature they see. The potion is poured onto Lysander's eyelids while he sleeps, and the first person he sees on waking is Helena. Lysander's love is the mere result of the spell that has been cast on him. According to my account, Lysander's love does not display a level of plasticity sufficient to be rationally assessable. This is because, in this scenario, Lysander's love is the mandatory result that the potion produces. Lysander's love would qualify as a-rational rather than irrational. Some may have the intuition that Lysander's love is irrational. However, I do not think that this intuition should lead to a revision of the theory. This intuition may rest on the idea that Lysander is able to provide reasons for his love. Lysander confabulates when he tells Helena that "the will of man is by his reason sway'd, and reason says you are the worthier maid" (Shakespeare, 2016). However, the reader knows that the only reason that Lysander falls in love with Helena is the spell. The irony of this line resides precisely in the idea that Lysander's reasons are in fact artificial, and that his love is deaf to reason. But being deaf to reason is not sufficient for irrationality. Suppose that Lysander was tricked by the spell

to see the leaves of the first tree he saw on awakening as brownish. I suspect that my challenger would agree that Lysander's perception, in this scenario, would not be irrational, but a-rational. Moreover, whoever criticises my argument would also have difficulty explaining the rational assessability of similar scenarios. Lysander's situation is analogous to emotions produced as a result of electrical brain stimulation. Emotions may be produced as a result of direct electrical stimulation of the amygdala (Lanteaume et al., 2007). In this case, the emotion is triggered by a brute law relating the brain-stimulation with an emotional response. The law ensures that if the zap or the electrical stimulation are present, the emotion is triggered. It follows that my challenger would also need to be ready to accept that emotions felt as a result of direct stimulation of the brain are irrational as well, however counterintuitive this may seem.

An interesting class of emotion is what Nico Frijda calls emotional "urges", that is, emotions that seem to mandate an expressive action about which we seem to be completely passive (Frijda, 1986). These emotions usually have high intensity and are triggered after a short latency time. For instance, suppose that while you calmly walk the street, a thief tries to steal your bag, grabbing it from behind your back. You may immediately experience an intense fear: you may start to scream before you even realise that a thief is trying to steal your bag. Although we seem to be passive to the activation of these emotional urges, they are the output of informationally plastic behaviour. Suppose that, for instance, in the scenario above, you knew that the thief had no intention of harming you. You would probably react differently. Therefore, emotional urges satisfy this condition.

Other interesting emotions that appear not completely flexible are recalcitrant emotions such as phobias. However, even if phobias persist against the subject's beliefs, they have an informationally plastic aetiology. Different sources of information may activate the same emotion of fear, e.g. seeing a drawing of a spider, thinking about a spider approaching, perceiving it, being told that there is a spider. Moreover, the emotional response can be

different if the subject has learnt to relate to phobogenic stimulus differently, for instance, if she has learnt to associate the stimulus with something positive.

4.2.2. Receptivity to the Evidence

Suppose that there is a person who has mental states similar to the ones you have. However, unlike you, this person cannot understand the import of new evidence for her "view of the world", that is, her understanding of what the world is like. In other words, her mental states are not receptive to the evidence. Her mental states do not seem epistemically rationally assessable. Suppose, for instance, that this person believes that there is a glass of water on a desk. However, sometime later, a new incoming perceptual state represents that the glass of water has been moved to a cupboard. Although such perception enters into her view of the world, her processes underlying the maintenance of her belief that the glass of water is on the desk do not recognise that this incoming perception is relevant for maintaining the above-mentioned belief into her view of the world. In this case, this belief does not seem epistemically rationally assessable.

When we consider paradigmatically rationally assessable mental states, such as beliefs, we observe that these states can be relevant for our view of the world, but also that these mental states can be put in a certain relation to the evidence. Whereas the property of contribution to the subject's view of the world moves from the evidence to the inclusion of *m* into the subject's view, the receptivity to the evidence of *m*'s maintenance processes regards the other side of the coin: it regards the maintenance of *m* in relation to changes in the view of the world. Consider again King Lear's tragedy. At the beginning of the tragedy, King Lear believes that Goneril loves him. Goneril's hyperbolic speech is taken at face value, and King Lear endorses his belief that his daughter loves him. This belief has a high epistemic-force value for King Lear. However, sometime after, the King realises that Goneril's actions do not reflect her

declarations of love for him. Lear's view is receptive to the evidence, and the epistemic-force value of his belief decreases in accord with it.

Lear's belief revision process relies on a crucial aspect of epistemic rational assessability: *m*'s epistemic-force value is assessed against the new incoming evidence. This receptivity to the evidence, in its minimal sense, consists in the possibility of a certain plastic relation between the mental state *m* that constitutes the subject's view and another mental state[69] that is carrying information about the incoming evidence. This relation does not need to be overly complex. It consists of the possibility to assess the compatibility between the evidence and the mental states forming the subject's view of the world. We can, therefore, state the following requirement for the epistemic rational assessability of the mental state *m*:

- *C_2:* the mechanism *P* that maintains *m* is generally receptive to the evidence.

In my account, receptivity to the evidence consists of the capability of the mechanism that maintains *m* to compare incoming evidence with *m*, detecting any conflict with it. Receptivity to the evidence does not imply responsiveness to the evidence. The mechanism maintaining *m* is responsive to evidence when it is both receptive to evidence, and revises *m*'s content accordingly. For instance, when any incongruence is detected between my belief that the plane is flying safely and my perception that the plane is crashing, the mechanism maintaining either my belief or my perception should not only recognise the conflict but also change my mental state in accordance with the evidence.

A state can be receptive without being responsive to the evidence. Consider a paradigmatic case of fear of flying. The subject believes that flying is safe, and therefore she gets on the plane. However, when the plane is taking off, she feels scared of flying. She realizes that there is an incongruence between her previous belief and the new incoming emotion. She sees that

[69] This mental state needs to be descriptive. I will discuss descriptive mental states in section 4.2.3 of this chapter.

everybody else is seated, calm, and relaxed. She also knows that statistics suggest that flying is safe. She knows that she should trust her belief, not her emotion, but her emotion persists against her better judgment. The mechanisms maintaining her emotion are *receptive* to evidence; we can infer this because she detects incongruence between her emotions and her belief, and she understands that based on the available evidence, she should revise her emotion. However, the mechanism that has led to the emotion does not *respond* to the evidence. She continues to experience fear for the whole duration of the flight. My account leaves open the possibility that this emotion of fear is open to rational assessment. I take this to be a strength of my account because paradigmatic examples of rationally assessable mental states lack this property.[70] For instance, it is commonly noticed that people often stick to their beliefs, irrespective of evidence that undermines the basis of these beliefs. 41% of registered Republicans maintain their belief that Obama was not born in the US despite being presented with his actual US birth certificate (Clinton and Roush, 2016). Moreover, psychologists have discovered a widespread *belief perseverance* bias (Ross, Lepper, and Hubbard, 1975). A belief is perseverant when it is maintained despite robust evidence against the basis of the belief. In the empirical study conducted by Ross and colleagues, for instance, participants were shown cards displaying fake and real suicide notes and were asked to identify which notes were real (Ross et al., 1975). Participants were then given false performance results. After some time, one group of participants was "debriefed", i.e. informed that their task scores were randomly assigned and that the results were not correlated with their ability. The other group was not given this information. Interestingly, participants in both groups maintained their previous belief regarding their ability to perform the task. Even when participants were "debriefed",

[70] This reply has been inspired by Siegel's account of the rationality of perception (Siegel, 2017).

they maintained their previous belief about their ability to distinguish fake from real notes to some extent.

Reason receptivity and reason responsiveness are two different properties. However, it is difficult to establish whether a mental state is receptive without being responsive to the evidence. Although there are clear epistemic manifestations of the property of reason responsiveness, e.g. the subject revises her mental state in accordance with her reasons, it is less clear what the epistemic manifestations of receptive mental states that are not reason responsive are. Lisa Bortolotti suggests that we can infer reason receptivity of a belief that is not reason responsive when we do not reject it, but we slightly revise it in light of the evidence (Bortolotti, 2010). Moreover, she infers that some delusions are reason-receptive (although not reason responsive) from the tendency of the subject to argue for the content of the delusion when challenged (Bortolotti and Miyazono, 2015).[71]

Let us apply C_2 to the emotions. Emotions are occurring states. Their appraisal mechanism is responsible for maintaining the disposition to form the emotion *e* when a certain stimulus occurs. I assume that the mechanism *P* that maintains *e* is the appraisal mechanism. I will discuss the appraisal mechanism in more detail in chapter 7.

The mechanism *P* that maintains the emotions usually display some levels of receptivity to the evidence. A clear case of this is when a subject has the disposition to form a certain emotion *e* in a certain context *C*. When the evidence shows a change in the evaluative character of *C*, the disposition to form *e* in *C* changes. Consider, for instance, King Lear's love[72] for Goneril.

---

[71] Bortolotti argues that reason receptivity, or, in her terminology, reason sensitivity, is a necessary feature of belief but reason-responsiveness is not (Bortolotti, 2010; Bortolotti and Miyazono, 2015). Examples of beliefs that are reason receptive but not reason responsive are some instances of superstitious beliefs, delusions and religious beliefs.

[72] Assume, for the sake of the argument, that King Lear experiences an emotion of love for Goneril.

At the beginning of the play, King Lear loves her. After some time, Lear realises Goneril is offending and harming him. Goneril deliberately makes him feel weak and inferior, and leaves him to wander outside during a storm. As a result, Lear's disposition to feel love for Goneril is revised; after these actions, he no longer feels love for her. In this case, Lear's emotion is responsive to the evidence. The appraisal mechanism maintaining Lear's emotion of love for Goneril changes it and, respectively, Lear's disposition to feel love for his daughter. Given that responding to the evidence entails receptivity to the evidence, we can infer that in this case, Lear's love was receptive to the evidence. This is a fictional scenario, but many emotions are revised in this way. For instance, I may be annoyed at my friend who turned up very late for a meeting, but my irritation ceases when I realise that she had an unexpected setback. Again, in this case, the receptivity to the evidence is inferred from the responsiveness to the evidence. There are other types of emotions that seem not to be responsive to evidence. This is the case of recalcitrant emotions. In these cases, we cannot infer the receptivity to the evidence from the emotions' responsiveness to it. However, recalcitrant emotions, namely emotions that conflict with our evaluative beliefs may be overcome (Paquette et al., 2003). Even if recalcitrant emotions are not immediately revised, the disposition that maintains them can be changed over time in response to evidence, to a certain extent. After being exposed to many safe flights, for instance, the disposition to fear flights changes to a certain extent. The best explanation for recalcitrant emotions' ability to be diachronically changed is their receptivity to the evidence. Therefore, although recalcitrant emotions are not immediately responsive to the evidence, they are receptive to it.

4.2.3. Descriptive Representational Content

Mental representations can be of two kinds: (1) descriptive and (2) directive. Descriptive representations present facts about the world (Anscombe, 1957). Consider, for instance, Lear's belief that Cordelia is his daughter. This describes a state of affairs, namely a state in which

Cordelia is his daughter. Conversely, directive representations are representations that bring about the disposition to modify the world so as to fit the content of the mental representation. Desires are commonly regarded as examples of directive representations. Consider, for instance, Lear's desire that Cordelia flatters him. This desire brings the disposition to change the world, in such a way to fit the mental representation. For instance, Lear decides to ask Cordelia to express her love to him, expecting flattery in return.

Intuitively, descriptive representations can be open to rational assessment, but directive representations cannot. This can be seen in our ordinary practice. Our ordinary practice considers paradigmatically descriptive states, such as beliefs, as being open to epistemic rational assessment. For example, we may say that Lear's belief that Cordelia is offending him is irrational. Conversely, in our ordinary practice, we do not consider paradigmatically directive states like desires to be open to epistemic rational assessment. The following remark – Lear's desire to be flattered by Cordelia is irrational – does not seem meaningful at least in the epistemic sense.[73] A similar observation was made by Ronald de Sousa, who argued that rationally assessable emotions require descriptive content (de Sousa, 1980, 127-151). It thus seems plausible to assume that a mental state is epistemically rationally assessable only if my third condition holds:

- *C_3: m* has a descriptive representational content.

[73] It may be argued that some mental states, such as Ruth G. Millikan's *pushmi-pullyu* representations (Millikan, 1995) have both a descriptive and a directive representational content. Evaluative beliefs – such as the belief that the Rosetta's stone is interesting – and emotions could be considered mental states with descriptive *and* directive contents (see, for instance, Solomon, 2007). It may be also argued that desires also have a descriptive and directive representational content. If desires had both directive and descriptive representational content, then epistemic rational assessment would only apply only to the desire's descriptive content.

This condition is consistent with background beliefs about epistemic norms. Epistemic norms seek an accord between our description of the world and the state of the world, whereas practical rationality norms do not.

Emotions satisfy this condition: emotions are often considered to have an *evaluative* description, that is, a description of how the world is impinging the subject's well-being (Lazarus, 1991; Helm, 2001). Although there is disagreement about what emotions are, the thesis that emotions have a descriptive representational content is widely accepted by many cognitivists and non-cognitivists (see, for instance, de Sousa, 1987; Nussbaum, 2004; Prinz, 2004; Solomon, 1988). An exception to this is the feeling theory of emotion, according to which emotions do not have a descriptive representational content (James, 1984). If you endorse a feeling theory of emotions, then emotions are not epistemically rationally assessable because they do not fulfil C_3.

4.2.4. Epistemic Force

Not all mental states with descriptive representational content appear to be epistemically rationally assessable. Imaginings and daydreaming states, for instance, look a-rational. Consider Cordelia's imagining that King Lear forgives her. Although the content of her imagining is descriptive, it does not meet the requirements to be epistemically rationally assessed.[74] To present the necessary and sufficient conditions for epistemic rational assessability, I need to explain what property grounds the distinction between beliefs and imaginings. Here I suggest that the crucial feature that is shared by beliefs but not by imaginings is the mental state's *epistemic force*, and that this is the property that explains why

[74] This claim regards only the epistemic rational assessability of mental states and leaves open the possibility that imagination or daydreaming are mental states that are *practically* rationally assessable.

beliefs qualify as being epistemically rationally assessable, but imaginings and daydreaming states do not.

One preliminary observation is that beliefs and perceptions, but not imaginings, contribute to the subject's "view of the world". Consider, for instance, a subject who believes that a glass of water is in front of her. Such a subject may take her belief at face value.[75] As a result, she may move her arm, grasp the glass, and drink the water. Compare this with a subject who imagines that a glass of water is in front of her. Imaginings are not usually taken at face value and are not usually acted upon. Imaginings and daydreaming states are not useful, *per se*, to gain an accurate view of the actual world. This is because the contents of what we imagine or daydream are largely determined by us, and not by the world (Gendler, 2016).

It might be objected that, in some exceptional circumstances, such as when a subject uses imagination to solve a certain task, e.g. assessing whether a table will fit in a room, the subject's imagining seems to contribute to her view of the world: her imagining seems to tell her something about how the world actually is, providing a reason for believing that the table will (or will not) fit in the room. However, in this case the contribution is not direct, but indirect, because mediated by the belief that there is a mapping relation between the imagined room and the real one.[76]

---

[75] In philosophy of perception, the property of mental states that explain why we take perceptions, but not imaginings as face value is often called "assertivity" (e.g. Briesen, 2015 pp. 2245-2252; Tucker, 2010 p. 530), or "forcefulness" (Huemer, 2001 pp.77-79). However, except Briesen (2015), the accounts of this property have been very concise and underdeveloped.

[76] Realistic imaginings may be classified as contributing to the subject's view of the world. For instance, imagining that a glass of water is standing in front of somebody is more realistic than imagining that the philosopher's stone is standing in the same place. However, realistic imaginings, *per se*, do not tell us anything about the world.

Imaginings may still contribute indirectly to our view of the world. Consider a subject who is prone to imagining a bright future for herself and to imagine good things happening for her and her loved ones; it is easy to predict that this imaginative disposition has a long-term impact on her view of the world. She will tend to form beliefs that describe a positive future for her. However, also in this case, the contribution of the imaginings to the subject's view of the world is only indirect: the imaginings themselves are not incorporated in the subject's view of the world. We usually quarantine imaginative descriptions from the mental states that comprise our view of the world.[77]

So far, I have argued that there is an important distinction between beliefs and imaginings that seems relevant for the epistemic rational assessability of mental states: beliefs *directly* contribute to the subject's view of the world, but imaginings do not. It is therefore natural to suppose that rationally assessable mental states require a direct contribution to the subject's view of the world. I now need to qualify what I mean by such a contribution.[78]

A mental state directly contributes to our view of the world when it disposes us to take its content at face value. I call "epistemic force" the property of a mental state *m* that grounds the subject's disposition to take the mental state content at face value. For instance, consider again a subject who believes that a glass of water is in front of her. Her belief brings about the

---

[77] See Nichols and Stich's theory of imagination and pretence (Nichols and Stich, 2000).

[78] One possibility would be to consider a mental state to contribute to a view of the world if and only if it *assertively* represents the world as *actually* being in a certain way (Briesen 2014). I do not have space here to closely discuss this account. For the purposes of this chapter, it is sufficient to note that his account of assertivity of a mental state *m* already presupposes that *m* is procedurally rationally assessable. In the rest of the chapter I aim to offer a less-theoretically demanding understanding of assertivity that does not explain the condition under which a mental state directly contributes to the view of the world by appealing to a property that already presupposes the (epistemic or procedural) rational assessability of that mental state.

disposition to take the content of the belief at face value: she is disposed to act as if the content of her belief was true. She may also be disposed to act on the basis of her belief, e.g. grasping the glass and drinking its water. Conversely, imaginings do not usually dispose us to take their content at face value. Therefore, they lack epistemic force.

The epistemic force of the mental state is not determined only by the content of the mental state, but also by the mental state's mode. The strength of the disposition is proportional to the degree of epistemic force of the mental state. Some mental states' modes may, by default, have larger epistemic-force value than other mental states' modes. Consider a situation in which two mental states of different types are in conflict, for instance when a subject fears something and at the same time believes it to be safe. In this situation, it may be the case that one mental state type brings about a disposition to be taken at face value that is stronger than the other. For instance, it may be the case that perceptions usually have larger epistemic-force values than beliefs,[79] or that fear occurrences have, in general, larger epistemic-force values than beliefs about the dangerousness of objects. When a subject has a phobia of spiders, for instance, she may believe that spiders are not dangerous, and, at the same time, fear them. It may be more difficult for the subject to keep her fear out of her view of the world than her belief. Consider, for instance, the belief of a subject with fear of flying: suppose that she believes that flying is safe when the plane is ready for take-off. At this very moment, her belief has some degree of epistemic force: she takes the belief at face value, for instance, when allowing her son to take an intercontinental flight. However, when she is on a flight herself, she may trust her fear instead and she may refuse to enter the plane. In this case, both her belief and her emotions have a significant degree of epistemic force: the belief that flying is safe still brings about her

[79] This seems consistent with Jesse Prinz's hypothesis of a perception trumping mechanism that makes perception always trump beliefs when these two mental states come into conflict (Prinz, 2006b).

disposition to act accordingly in certain contexts, such as allowing her son to travel. Her fear of flying disposes her to act as if flying was dangerous in other contexts, such as when she enters the aircraft. In this situation, the subject is torn between her belief and her fear precisely because both of them have epistemic force.

The exact degree of epistemic force of a mental state does not need to be accessible to the subject's consciousness. Epistemic force is a property *of the mental state*, and not a second-order belief about the reliability of the mental state. Beliefs, in general, have a significant degree of epistemic force. Consider King Lear's belief that Goneril is his daughter. This belief is taken at face value and incorporated in the subject's view of the world. On the contrary, imaginings lack epistemic force. As I argued at the beginning of this section, they do not tell us anything directly about the world. Imaginings do not dispose us to take their contents at face value. Therefore, the fourth necessary property for the epistemic rational assessability of a mental state *m* is the following:

- C_4: *m* displays epistemic force

This condition can be used to explain the epistemic rational assessability of emotions. Emotions often have epistemic force: they dispose us to take their content at face value. Emotions brings about motivational tendencies (Frijda 1986; de Sousa 1987; Brady 2013; Tappolet 2007). For instance, fear and worry motivate us to flee away from the danger. They also usually motivate us to take their content at face value.[80] Fearing an object *x* brings about the disposition to take the dangerousness of *x* at face value; similarly, being angry at *x* disposes us to think that *x* is offensive. Emotions may differ in the strength of their epistemic force: some recalcitrant fears, for instance, may have a low level of epistemic force because we may know that the information that the fear is giving to us does not match what we know about the dangerousness

[80] This has been discussed by William Alston (1969) and Jérôme Dokic (2018).

of the emotions' intentional object. Other emotions are immediately taken at face value and acted upon. If there was an earthquake and I felt fear, I would probably assign a large epistemic-force value to my fear because I know that earthquakes may put human life at risk.

Desires, wishes and perceptions may also have epistemic force. Wishful thinking cases seem to suggest that conative states such as wishes and desires may bring about the disposition to believe the content of the desire.[81] Perceptions may also have epistemic force: my perception of the rain, for instance, may have epistemic force: it often brings about the disposition to take at face value its content.

To sum up, in this section I argued that a mental state *m* is epistemically rationally assessable if and only if *m* satisfies *all* the following four conditions:

- *C_1: m* is informationally plastic
- *C_2: m* the mechanism *P* that maintains *m* is generally receptive to the evidence
- *C_3: m* has a descriptive content
- *C_4: m* displays epistemic force

Each of these conditions is necessary for epistemically rational assessability.

4.3. Epistemic Rational Assessability and Responsibility

In this section, I discuss Kate Nolfi's account of the epistemic rational assessability of beliefs (Nolfi, 2015).[82] Although her account does not specify whether the rational assessability is epistemic or practical, it is possible to read it as an account of the epistemic rational assessability of mental states. According to this interpretation of her account, a mental state is

[81] This does not make wishes and desires epistemically rationally assessable states, since wishes and desires do not meet C_3.

[82] Nolfi uses a different terminology: she uses the term "rational evaluability".

epistemically rationally assessable only if the subject who has this state is answerable for it, that is, only if the subject can be asked why-questions for her having this state. For Nolfi, the answerability of the subject is grounded in her responsibility for being in that state, which, in turn, depends on the subject's ability to shape her disposition to believe.

Nolfi's account differs from mine in that it ties together questions of rational assessability and questions about responsibility. In her account, assessability goes hand in hand with responsibility because we use rational evaluation to hold someone accountable for her mental states; as Nolfi puts it: "when we evaluate a subject's belief as rational or irrational, justified or unjustified, our evaluation paradigmatically, if not universally, serves to hold the subject to account for believing as she does" (Nolfi 2015 p. 5). Nolfi rightly argues that the subject's responsibility depends on facts about the extent to which the subject can regulate her own beliefs. However, the epistemic rational assessability of a mental state does not imply that the subject is responsible for this state. When we assess the rational assessability of a subject's mental state, we are not necessarily considering this subject accountable for being in that state.

I think that we need to disentangle responsibility and rational assessability because this better explains our evaluative practice. In her paper, Nolfi considers a scenario in which a subject is terrified of flying. When the subject is waiting for take-off, she believes that the plane is going to crash despite her better contrary belief (Nolfi, 2015). In this scenario, the subject is not able to regulate her belief through reasoning. Nolfi argues that in our evaluative practice, we would consider the subject's belief that the plane will crash as irrational, but we would also excuse the subject for her belief.

However, Nolfi's account seems to predict the opposite: the a-rationality of the belief that the plane will crash. This is because the subject in this scenario, is not answerable for her belief because she is not able to regulate her own belief with her reasoning. Therefore, according to Nolfi account, she is not responsible for her belief that the plane will crash. And without the

subject's responsibility for her fearful belief, her belief would be a-rational. Despite this, Nolfi maintains the subject's fearful belief is rationally assessable because she considers this belief the exception, rather than the rule. Her strategy consists in showing that beliefs, by and large, are states for which why-questions apply, and the rational assessability is enjoyed by all the occurrences of a type whose tokens, by and large, satisfy such requirements.

In this chapter, I have presented a different picture, in which the rational assessability criterion applies to mental state tokens, and not to types. This is because, as I argued in section 4.2, some instances of paradigmatically rationally assessable mental states may seem a-rational, e.g. zapped and mechanically induced beliefs. My account is able to distinguish between rationally assessable and a-rational tokens of a mental state *M*, and this is more in line with our intuitions about such exceptional cases. It seems strange to consider a belief with a zap-like aetiology epistemically rationally assessable, simply because it belongs to a type that is generally so. Moreover, cases of fearful beliefs suggest that questions regarding rational assessability and responsibility ought to be kept distinct. I could in principle agree with Nolfi that the subject is in some sense excused for her belief: after all, she is not immediately able to take into account the rational criticisms of her belief.[83] However, the extent to which the subject is excused in this situation would require more careful consideration. It is not clear why facts about the indirect causal efficacy of our judgments or habits do not play a role in defining the subject's responsibility for her belief.

Existing accounts of the rational assessability of beliefs rely on the claim that rational assessability requires some level of answerability (Hieronymi, 2008; Nolfi, 2015). However, there are cases in which a subject is not able to offer reasons for a mental state, but the mental

[83] In this thesis, for space reasons, I will not present an account of excuses for actions based on epistemically irrational emotions.

state is understood as a paradigmatic example of irrationality, thus is a rationally assessable mental state. This is the case with people with delusions who admit the implausibility of their beliefs (Coltheart, 2007). My account is, therefore, better able to explain the rational assessability of such cases. This is because delusion occurrences fulfil the rational assessability conditions presented in this chapter.

Moreover, there may also be reasons to doubt the reliability of the subject's ability to provide reasons for paradigmatic rationally assessable cases, such as belief. This is because empirical studies in psychology show that in many cases subjects confabulate about their reasons for beliefs (Nisbett and Wilson, 1977, Wegner, 2002).

4.4. Summary and Conclusion

In this chapter, I presented an account of the epistemic rational assessability of mental states, and I showed that most emotions are epistemically rationally assessable. With this framework, it is finally possible to defend the epistemic rational assessability of King Lear's anger. His anger satisfies all four conditions. It is informationally plastic: the King's anger at Cordelia could have been experienced as a result of different input, and the King could have reacted differently to Cordelia's speech. Moreover, the mechanism that maintains King Lear's anger is receptive to the evidence. We can infer this because, towards the end of the play, the King ceases to be angry at Cordelia. The anger has a descriptive representational content: Cordelia is represented as offensive. Furthermore, the King's anger is contributing to his view of the world. It is on the basis of his anger that Lear decides to deny paternal care to his daughter: the anger has epistemic force, and so the fourth condition is satisfied. King Lear's anger is therefore epistemically rationally assessable. Since his anger is not in accordance with the evidence, it is epistemically irrational.

My account can be accepted by cognitivist and by most non-cognitivist theories of emotions. According to my account, many emotion occurrences are epistemically rationally assessable in

virtue of their above-mentioned four features that they share with other typically epistemically rationally assessable mental states, e.g. beliefs. Moreover, the epistemic rational assessability of emotions is not *sui generis* but, as my account shows, it fits with the epistemic rational assessability of other mental states.

CHAPTER 5: EMOTIONS AND CORRECTNESS CONDITIONS

ABSTRACT

Emotions inform us about things that matter to us. They tell us about the extent to which objects display affective evaluative properties. But under what conditions are affective evaluative properties instantiated? And is there an objective way to determine whether an emotion is correct? In this chapter, I argue that it is possible to objectively establish whether an emotion is correct. I argue that the intensity of emotions could be objectively quantified, and I present a relativistic account of affective evaluative properties according to which the instantiation of evaluative properties depends on the subject's concerns.

INTRODUCTION

In the previous chapter, I presented my account of the epistemic rational assessability of mental states, and I argued that most emotions meet these conditions. But what are the epistemic norms that could be applied to emotions? How are they generated? And should we always try to form emotions that are in accordance with the evidence? Before getting into these questions, which I will discuss in chapter 6, I need to explain what it is for an emotion to be correct, and in which sense evaluative properties are objectively instantiated. This is a necessary step for this discussion because epistemic rationality norms will be selected as the norms that will increase the likelihood to enjoy correct emotions. Without a firm grasp of the notion of correctness, I cannot provide a satisfying account of how epistemic rationality norms of emotions are generated.

The starting point of this discussion is the evaluative content of emotions. In previous chapters, and especially in chapter 2, I claimed that emotions are *intentional* and *evaluative*

mental states: they are about an object,[84] and they present such an object as having a certain affective evaluative property. For instance, consider Othello's anger at Cassio. Othello's anger is directed at a *particular object*, namely Cassio. The anger is also presenting evaluative information to Othello: the information that Cassio is perpetrating an offence against him. Different types of emotions give different types of evaluative information: anger represents something as being offensive, disgust represents something as being noxious, etc.[85]

In this chapter, I will explore the claim that when an emotion occurrence is *correct*, the evaluative information the emotion gives to us tracks how things really are: the emotion's object displays the relevant evaluative property to the degree indicated by the emotion's intensity.[86] Consider, as an illustration, Shakespeare's play *The Tempest* (Shakespeare 2008, [1610]). The play begins with a violent storm that hits a ship. The boatswain is gripped by fear and tries to prevent the sinking of the vessel spurring his crew to work hard and take back control of the ship. There is a sense according to which the boatswain's fear is correct: the storm is an imminent danger to him. When an emotion occurrence is incorrect, it mistakenly represents the evaluative import of the subject's situation. Consider, for instance, Shakespeare's tragedy *Othello* (Shakespeare, 2005[1622]). In this play, Othello is deceived by Iago that his wife Desdemona is having an affair with Cassio. Othello's anger at Desdemona

[84] I use the term "object" to refer to whatever entity, event or state of affairs that is the intentional focus of emotion.

[85] See also chapter 2 for a discussion about the evaluative content of emotions.

[86] Correctness conditions for emotions seem comparable to truth conditions (Tappolet, 2016c, p. 88). In this chapter, I focus solely on correctness conditions and I will not discuss whether they can also be considered truth conditions for emotions. It has been argued that correctness conditions of emotions may differ from the truth conditions of an emotion (Salmela, 2014, chapter 5). Accounts of emotional truth have been put forward by Salmela (2006, 2014), and de Sousa (2002). Criticisms have been raised by Morton, (Adam Morton, 2002), Döring (2009) and Gunther (2003).

and his jealousy of Cassio are both incorrect emotions: Desdemona is loyal to Othello and Cassio did not have an affair with Desdemona. As evident in this play, incorrect emotions may be pernicious: they may lead subjects to form false beliefs and to make bad decisions. Conversely, in a large variety of contexts, subjects are better off if they form emotions that are correct. They are better able to understand what really matters to them and they are better equipped to foster their concerns.

This chapter will be structured as follows: in section 5.1, I present an account of correctness, inspired by D'Arms and Jacobson's norms of a "shape" and "size" (D'Arms and Jacobson, 2000b). This account of correctness opens up new questions regarding the intensity of emotion. What does the intensity of emotion represent? Can we quantify the emotion's intensity? And can we assess whether an emotion has the right intensity? I will address these questions in section 5.2. In section 5.3, I will consider existing accounts about the correctness of emotions and I present two desiderata for an account of affective evaluative properties. One important desideratum is that affective evaluative properties are mind-dependent. In section 5.4, I discuss existing accounts of concerns, and I suggest that we should favour a desire-based account. Finally, in section 5.5, I explain my own relativistic account of affective evaluative properties.

5.1. "SHAPE" AND "SIZE" OF EMOTIONS

D'Arms and Jacobson argue that emotions are correct only if they have the right "shape", that is, the right formal object, and "size", i.e. the right intensity (D'Arms and Jacobson, 2000b).[87] Following D'Arms and Jacobson's account, I suggest that an emotion *E* with an intensity *I* directed at the particular object *O* is correct if and only if:

[87] An alternative account of correctness has been given by Mikko Salmela (Salmela, 2014).

(a) the object *O* has the affective value *v* represented in *e*'s formal object, relative to the subject's concerns

(b) the degree according to which the object has *v*, relative to the subject concerns, is proportional to the emotion's intensity

For example, Othello's anger at Desdemona is correct if and only if Desdemona is offending Othello, relative to Othello's concerns – in this case, his wife's fidelity – and it is an offence to the degree indicated by the emotion's intensity value. However, Desdemona is not offending Othello at all. Since Desdemona is faithful to Othello, Othello's anger at her is not correct. Emilia's sadness at the death of Desdemona is correct: Emilia cares about Desdemona's life. The death of Desdemona represents a loss for Emilia, and a loss of great value. Given that Emilia is concerned about Desdemona's life, her sadness is correct. If Emilia did not care about Emilia's life and company, her death would not have been a loss for her, and her sadness would not have been correct.

On the one hand, condition (a) ensures that correct emotions give accurate information about the evaluative import of the situation. The boatswain's fear, for instance, satisfies condition (a): the situation is really dangerous. On the contrary, Othello's jealousy at Desdemona does not satisfy the first condition because Desdemona is loyal to Othello: Othello's concern about his wife's fidelity is not affected by any of Desdemona's actions. Condition (b) captures the idea that the emotion needs to be proportional to the extent that its object is affecting the subject's concerns. Suppose that a subject is concerned about her own safety and about her own professional reputation. One day, she risks arriving late at a business meeting. In this situation, arriving late might involve some level of danger for her: it may damage – to some extent – her reputation among her colleagues and her business partners. However, let us stipulate that arriving late just once will not destroy her professional credibility. Now, suppose that she feels an extremely intense fear of arriving late. Her fear may satisfy condition (a): it

will correctly inform her that there is some level of danger involved in arriving late. However, the intensity of her fear would be incorrect: arriving late will not be dramatically dangerous for her. Condition (b) is placed to account for cases like this, where the emotion is incorrect not because the object lacks the evaluative property represented by the emotion's formal object, but because the intensity of the emotion is incorrect. This aligns with our everyday talk of emotion correctness. We criticise someone's emotion not only when it is attributing the wrong value to a certain situation or object – e.g. you shouldn't fear spiders! – but also when it is too intense – e.g. you shouldn't be *so* angry with her!

The intensity informs us about the extent according to which objects have affective evaluative properties that figure in the content of emotions (Tappolet, 2016c, p. 24). The emotion vocabulary reflects this variation (Prinz, 2010, p. 524). Not all emotions that present an object as being offensive, for instance, are presenting the object to be offensive to the same degree. When we experience severe anger, we say that we are enraged; when the anger is moderate, we say that we are angry; when the anger is mild, we say that we are irritated. I take the intensity of the emotion to depend on the extent according to which the following emotional components are affecting the subject's experience:[88] the bodily changes that accompany the emotion, the force of the motivational components and the degree according to which emotions control deployment of attention.[89]

Condition (b) requires that the intensity of the emotion can be quantified. Quantifying the intensity of an emotion raises some problems, such as how to weight the contribution of each

[88] For an alternative account according to which the intensity of an emotion is given by the extent to which the emotion modifies affordances, see (Shargel and Prinz, 2018). Daniel Shargel and Jesse Prinz use an enactivist account to explain the intensity of emotion in this way. Given that I am using a non-enactivist account, I will not discuss this approach here.

[89] For similar accounts, see (Ben-Ze'ev, 2000, chapter 5).

emotion component to the emotion's intensity, and how to identify a common scale for the intensity contribution of each component (Ben-Ze'ev, 2000, pp. 118-119). Regardless of these difficulties, both psychologists and ordinary people quantify the intensity of emotions when discussing emotional experiences. Psychologists often measure the intensity of emotion employing a variety of methods, e.g. asking participants to self-report the intensity of their emotion using graded-scale questionnaires (Mauss and Robinson, 2009; Meiselman, 2016). They also employ autonomic nervous system measures such as skin-conductance, cardiovascular response and facial electromyographic activity to infer the variation in the intensity of the emotion experienced (Reisenzein, 1994).

Condition (b) requires not only that the intensity of emotions can be quantified, but also that affective evaluative properties admit of degrees. This is not an unorthodox view: this possibility has been already discussed e.g. (Tappolet, 2000). This is also often acknowledged in our daily practice: a remark can be offensive to a greater or lesser degree, food can be more or less noxious, losing an object could be a minor or a major loss for us etc. Public institutions often distinguish between different levels of danger, e.g. the department of health and human services distinguishes between four levels of danger involved in working in biomedical and clinical laboratories (Chosewood and Wilson, 2009). A correct emotion, however, does not need to meet the *exact* degree according to which a situation is impinging on the subject's concerns: we could expect some margin of tolerance. This is because, even with training, we do not have an extremely fine-grained ability to discriminate the variation of intensity of our emotions. We might able to distinguish with some levels of confidence, say, between 5-7 different levels of emotional intensity, but the more fine-grained the scale, the more difficult it becomes to discriminate between each level. Although what matters for correctness is the *actual* intensity of the emotion – and not the perceived intensity – it would be over-demanding to require our emotions to exactly meet extremely precise and fine-grained levels of emotional intensity.

Which fine grained-scale is adopted may depend from context to context, but it should not be over-demanding to meet.[90]

A final caveat needs to be added to this discussion. People may have different sensitivities, and different characters (Ben-Ze'ev, 2000, p. 152). A person that is generally cheerful will tend to display a greater intensity of happiness for something that is fostering her goal than a grumpy person. The exact degree of intensity of someone's emotion needs to be calibrated according to the subject's sensitivity and character. What is informative is the relation between the change and the person's baseline: the baseline of a grumpy person will be different from the baseline of a cheerful person.

To recap, in this section, I have outlined the notion of correctness of emotion. To assess whether an emotion is correct, we need to understand under what conditions an object presents an affective evaluative property, and we need to understand what emotional intensity is. I will discuss these issues in the following two sections.

5.2. Emotional Intensity and Importance

In this section, I focus on the emotion's intensity and I discuss whether intensity represents the degree according to which something is important to us.

There is a widespread consensus that emotions are mental states that are directed at things that are important to us (see, for instance Helm, 2001; Monteleone, 2017; Nussbaum, 2004;

[90] The difficulty to discriminate different levels of emotional intensity suggests that there may be edge cases in which is unclear whether the emotion is correct or not. However, the existence of such cases is not, *per se*, a problem for an account of the correctness of the emotions. Vagueness is a ubiquitous phenomenon (Williamson, 1994) and generates similar problems not only in philosophy of emotions but also in other contexts, such as in logic, philosophy of language, epistemic metaphysics, see, for instance, Keefe and Smith (1996). How to best deal with this these sorts of cases is a question that deserves consideration but falls beyond the scope of this chapter.

Roberts, 2003).[91] Martha Nussbaum argues that the variation in the intensity of the emotion is explained by the degree of importance that the subject assigns to the object (Nussbaum, 2004, p. 55). Consider Ann, who happens to break one glass. Depending on the circumstances and the importance of the broken glass, Ann might or might not experience emotion. We can imagine that, if the glass was made of crystal and was a special present for her birthday from her beloved aunt, she would not be indifferent to the breakage. She would rather experience intense sadness, proportional to the extent to which she considered the glass important to her. Conversely, if the glass was made of cheap material, was not a gift, nor particularly beautiful, Ann would maybe react with little or no emotion. Thus, the importance of the object of the emotion seems proportional to the intensity of the emotion.

A critic of this account might argue that there are cases where a subject believes that an object or a state of affairs, *x*, is not important to her, but when *x* occurs, she nevertheless feels an emotion of a certain intensity. She may use cases like this to insist that believing that something is important is not necessary for an emotion to be set off. For instance, Maria may believe that it is not important to her whether or not it is going to rain this afternoon. However, when her colleague tells her that sun is forecast, she feels happy. Cases like these seem to happen often. The case of Maria's happiness shows that the importance consciously assigned to objects or events, such as the weather forecast need not match the intensity of her happiness. However, emotions may represent something as important to us, regardless of our existing beliefs about it. Cases like Maria's happiness about the weather suggest that emotions can assign importance to states of affairs independently of the subject's beliefs about the

[91] One exception to this is the feeling theory, e.g. (James, 1884). For a discussion of the feeling theory and its main objections see chapter 1.

importance of that particular state of affair.[92] For emotions to represent something as important, it seems sufficient that the object of the emotion bears upon the subject's concerns. I will discuss existing accounts of concerns in section 5.4.

Nussbaum also argues that the intensity of the emotion is proportional to the degree to which the object of the emotion exemplifies the evaluative property given by the emotion's formal object (Nussbaum, 2004, p. 55). It is not clear whether, according to Nussbaum, the intensity of the emotion represents (1) the extent to which the intentional object exemplifies the evaluative property given by the emotion's formal object or (2) the extent to which the intentional object is important to the subject. Apparently, (1) and (2) seem to deliver different values; consider, for instance, the evaluative property "being a loss". Evaluating something as being "a certain loss" does not seem, necessarily, to imply evaluating it as "being important" to the same extent. Carolyn Price, for instance, accommodates the possibility that losing one hair may be considered "a loss" for the subject, although such a loss is of no importance for the subject (Price, 2013). This is a case where something seems to exemplify an evaluative property, being a loss, although this loss does not seem to be important for the subject. In this case, the extent to which the intentional object exemplifies the evaluative property "being a loss" does not appear to be identical to the extent to which the intentional object is important to the subject.

I believe that, usually, losing one hair is not important to a subject.[93] However, this example does not show a distinction between the intensity values delivered by (1) and (2). This is because the evaluative property "is a loss", in this case, is purely descriptive. Typically, losing

[92] Note that even a cognitivist theory would have to accept this issue for explaining recalcitrant emotions.

[93] Things may be different if that hair was assigned an important meaning – e.g. it was the last hair, it was a sign of an illness, etc.

one hair is a loss, in this context, merely in virtue of descriptive facts about the world: the hair, which once belonged to the subject, is not "owned" by the subject anymore. However, when we say that the death of a loved one is "a loss", the property at stake is not merely descriptive, but *evaluative*: the death of a loved one is a loss for me precisely because it has a certain impact on my concerns (see section 5.4). And if something is affecting my concerns, then, *necessarily*, it ought to be important for me. Price's wider point, that being a loss does not imply being important to the subject to the same extent, however, still holds: there can be losses regarded as not very important that give rise to an intense feeling of sadness. With this distinction in mind, we can say that, necessarily, if something is represented as having an *affective evaluative* property, for the subject *S*, then it is *also* represented as important to *S*, but this degree of importance does not need to match the degree of importance that the subject consciously believes that the object has.

5.3. EVALUATIVE PROPERTIES

To understand whether an object *O* has the affective evaluative property *v*, we need to understand the conditions according to which these properties are instantiated. In this section, I examine accounts of affective evaluative properties – response-dependent accounts, detectivist accounts and fitting-attitude accounts. I will argue that (i) response-dependent accounts are not compelling and (ii) I identify two main desiderata for an account of affective evaluative properties.

An influential account of affective evaluative properties is the *response-dependent account* (Elgin, 2008; Hume, 2000 [1738]). [94] An account is called response-dependent when the

[94] Examples of response dependent accounts of moral evaluative properties that are response-dependent have been put forward by Smith (1989) and Lewis (1989).

instantiation of a property depends on a certain response (a judgment, a perception, an emotion, etc.) in a subject *S*. According to a basic version of response-dependent accounts of affective evaluative properties, no evaluative property is instantiated independently of an emotional response. On this basic version of the response-dependent account of affective evaluative properties, an object *O* has the evaluative property *e* for a subject *S*, if, and only if, the subject responds with the emotion *E* to *O*. Depending on which evaluative property we are considering, a different emotional response will be relevant for this account. If, for instance, we are considering the property "being dangerous", the emotion of fear will be the relevant emotional response. If we are considering the property "being a loss", the emotion of sadness will be the relevant response. More generally, given an affective evaluative property *e*, the relevant emotional response will be the emotion that has the affective evaluative property *e* as a formal object. If we apply this version of a response-dependent account to the example considered above, Desdemona's death is a loss for Emilia. This is merely because Emilia grieves for her death. If this were the case, then Desdemona's death would represent a loss, relative to Emilia, only in virtue of response-dependent features.

Under this account of evaluative properties, objectivity is precluded. Suppose that Emilia did not grieve at Desdemona's death at all. According to a response-dependent account of values, Desdemona's death would simply cease to be a great loss for her. In this scenario, we cannot say that Emilia's failure to feel grief facing the death of Desdemona would be incorrect. However, this consequence is unsatisfying. In our ordinary practice, we do think that emotions can be incorrect, and this basic version of the response-dependent theory cannot easily accommodate incorrect emotions. This is perhaps more evident if we consider phobias. Consider, for instance, a subject with arachnophobia who is scared of a European spider. According to the response-dependent account of evaluative properties, this particular spider would be correctly ascribed the property of being dangerous to the subject, simply because the

subject is fearing it. However, this does not seem right. We want to be able to say that, in this situation, the spider is not dangerous to the subject (or that at least it is not as dangerous as the intensity of the fear suggests). Another account of evaluative properties is required.[95]

However, more sophisticated versions of the response dependent account can accommodate this problem. One possibility is to consider evaluative properties that feature in the correctness conditions of emotions as dispositional properties.[96] I call this refined version of the response-dependent account "dispositionalist" account. According to the dispositionalist account: an object *O* has the evaluative property *φ* if, and only if, in normal conditions, it elicits the emotion *e* in a subject *S*. An object *O* elicits *e* when *O* is the cause of *E*. For instance, for Othello, something has the property of being offensive, as long as, in normal situations, it is disposed to elicit anger in him. If the situation is not normal, say Othello is sleeping, he may not feel angered by the same object or situation. However, he may still have the disposition to be angered by it: if in most of the possible worlds where he is faced with Cassio's having Desdemona's handkerchief, Othello is angered by it, then Cassio's having Desdemona's handkerchief is offensive to Othello. Unlike the response dependent account, the dispositional

---

[95] The idea that sometimes emotions may be "excessive" or incorrect, precisely because the intensity of the emotion does not reflect the extent to which the object of the emotion has the relevant evaluative property does not only guide our everyday practice, but also diagnostic criteria of mental disorders, for instance, the diagnostic and statistical manual of mental disorder (DSM V) describes anxiety disorders as "excessive fears" (APA, 2013).

[96] David Hume's sentimentalist theory can be considered an example of this view (Hume, 2000 [1738]). He focussed on moral properties, but we could consider these properties as a special case of evaluative properties. One example of an evaluative property he discussed is the property of being a "crime", in recent terms, the property of being "morally wrong". For Hume, something is morally wrong in virtue of a sentiment of blame that the object is disposed to arouse in us. Generalising this insight, we could say that, according to a Humean account, an object *O* has the evaluative property *φ* if, in normal conditions, it is disposed to elicit the emotion e that has *φ* as a formal object in a subject S.

account can accommodate the possibility of error. Suppose that, in normal situations, you are not scared of European spiders. According to the dispositional account of evaluative properties, European spiders are not dangerous to you. If one day you feel scared of a European spider, then the fear would be incorrect, since, in normal conditions, European spiders do not elicit fear in you. The dispositional account can, therefore, admit a certain level of objectivity of evaluative properties, where objectivity is understood as a state of affairs that occurs independently of the subject's response on a particular occasion.

However, the dispositional account is problematic. This theory fails to account for apparently incorrect emotions. Consider now a person with arachnophobia. She has the disposition to fear spiders; thus, the dispositional account would predict that spiders are dangerous for her. Yet, this does not seem right. The response-dependent theorist may refine her theory further, arguing that *O* has the evaluative property *e* if and only if it has the disposition to produce the emotion *E* in "the evaluative experts". I call this version of the response-dependent theory respectively "expertist". An expertist theory needs to explain who the "experts" are. According to Catherine Elgin, emotional experts are people who tend to feel correct emotions in their area of expertise. They do so because they have developed (1) their emotional sensitivities to discriminate evaluative properties and (2) the ability to calibrate the intensity of the emotion to the features of the situation (Elgin, 2008). For instance, an expert of sadness, by attending to her emotional states learns to discriminate between sadness and melancholy and learns to calibrate her sensibility to the magnitude of the loss involved in the situations she faces.

However, the expertist account would struggle with determining who the "experts" may be. Suppose that society designates a group of people *E* as experts of offensiveness. When an event that may be appraised as offensive occurs, the society looks at emotional reactions of such experts, and if they have the disposition to be angered by that event, such an event is regarded as offensive. However, this approach is problematic. There is no guarantee that "experts" will

be able to be disposed to feel the emotion of anger when a situation that concerns other people is offensive to them. Emotions are tied to *personal* concerns (Nussbaum, 2001; Roberts, 2003). We therefore need to modify Elgin's account as follows: An object *O* has the affective evaluative property *v* (e.g. being offensive) if and only if there is an expert *E* who developed (1) her emotional sensitivities to discriminate evaluative properties and (2) her ability to calibrate the intensity of the emotion to the feature of the situation, and (3) had such an expert been put in the same situation as *S*, and had the person had the same concerns *S* has, she would have experienced the emotion *e* (e.g. anger) at *O*.

However, this expertist account of evaluative properties is not compelling because it generates a problem of circularity.[97] What makes an agent an "emotional expert" is her sensitivity to evaluative properties, and her ability to calibrate the intensity of the emotion to the feature of the situation. To determine what makes a person *develop* a *better* sensitivity to evaluative properties – and an ability to calibrate the intensity of their emotion – we need an account of evaluative properties. This generates a problem of circularity: the conditions for acquiring a better sensitivity to evaluative properties seem to depend on an account of evaluative properties that this account is meant to explain.

And even if it was possible to identify "experts" without employing evaluative properties, the crucial problem of the dispositionalist is that it shifts the focus of the evaluation from the evaluated object to the evaluator: when the dispositionalist asserts that the spider is not dangerous, she is not concerned about the actual features of the spiders, but she is absorbed with the subject's emotional responses to the spiders. Mark Johnston calls this distortion of

[97] This argument is analogous to Crispin Wright's argument against response-dependent accounts of colour and moral properties that appeal to the concept of "best opinion" in the optimality conditions (Wright, 1992, pp. 120-123).

topic the "pornographic attitude": an attitude similar to a person who, instead of viewing the other with pleasure, becomes focussed entirely with her own pleasure (Johnston, 2001). According to Johnston, the "pornographic attitude" consists in considering emotions, rather than an epistemic route to unveiling the evaluative import of things, all there is for understanding such import. However, this attitude is misleading: there is a sense in which objects can be bearers of evaluative properties independently of whether they cause emotional reactions in us (Prinz, 2004). Suppose, for instance, that you did not know that exposure to asbestos may increase the risk of developing some types of cancer. We are inclined to think that asbestos may be dangerous to someone even if she lacks the disposition to fear it. It may be dangerous to you regardless of your knowledge that it is conducive to cancer. Whether our attitudes – such as fear – are correct, is an open question.

As an alternative, Johnston proposes a *detectivist* account of evaluative properties (Johnston, 2001). According to this account, evaluative properties exist mind-independently: whether something is dangerous does not depend on the mental states a subject has, and in particular, it does not depend on the subject's emotions: something has a particular evaluative property in virtue of its mind-independent features. Johnston's central idea is that the world is, in itself, not evaluatively neutral. For the detectivist account, emotions have an important role: they help us to detect evaluative properties. Johnston supposes that emotions are attuned to such properties because this attunement gave humans an evolutionary advantage in the past.

According to this account, we would be able to claim that – for a human being – European spiders are not dangerous since they are not venomous. Similarly, we would be able to claim that asbestos is dangerous regardless of the emotions it is disposed to cause in a subject since it increases the possibility that subjects who have been in contact with it develop cancer. Johnston's account does not display any "pornographic attitude". However, it does not seem to be fully satisfying. This is because this universalisation of evaluative properties does not

take into consideration important differences between the subjects' mental states. Different people may care about different objects and this modifies the extent to which objects have evaluative properties. Consider again the difference between Donalbain and Malcolm, and Lady Macbeth facing King Duncan's death. Independently of the subject mind, it would be difficult to understand whether something is a loss or not for a subject. There seems to be no way to objectively ascertain whether the King's death is a loss – or (practically) good – independently of the subject's mind.[98]

It may be argued that we could ascertain whether King Duncan's death is a loss for Malcolm and Donalbain by looking at their objective interests, such as the objective interest of having their father alive. In this case, the objection goes, we would be able to explain why King Duncan's death is a loss for Donalbain and Malcolm appealing, for instance, to the above-mentioned objective interest. However, characterising concerns as objective interests in this way is not compelling, as it would not be fine-grained enough to characterise some situations. For instance, there may be a situation in which a son hates his father and is not concerned about his well-being. In such situations, it seems at least questionable whether the death of a father would really be a "loss" for his son (see also section 5.4. on this issue).

A response-dependent account that is neither detectivist, nor dispositionalist – yet still response-dependent is the fitting attitude analysis (henceforth, FA) of evaluative properties (see, for instance, McDowell, 1985; Wiggins, 1998). John McDowell, in his paper "Values and Secondary Qualities*"* develops a very complex account of moral properties. This account can

[98] The dispositionalist may be better equipped than the detectivist to solve this problem. According to the dispositionalist, King Duncan's death is an irrevocable loss for Donalbain since he is disposed to grieve that event. At the same time, the King's death is a great advancement in Lady Macbeth's plans, since she has the disposition to rejoice it.

be extended to an account of affective evaluative properties, although the focus of McDowell's paper is moral properties. According to McDowell, an object *O* has the evaluative property φ for the subject *S* if and only if *O merits* or *fits* a certain emotion *e*. Consider, for instance, Othello's anger at Cassio. Cassio is offensive for Othello if and only if Cassio merits *e*. In this way, the theory avoids the dispositionalist "pornographic attitude", bringing back the focus of the inquiry onto the object. But in virtue of what is this merit earned? For McDowell, merely causal explanations cannot explain in virtue of what an emotion is fitting. This is because causal explanations do not rationalise emotions. Consider, for instance, Othello's anger. It seems questionable whether Cassio merits fear or not. Fear may be criticised – or approved – not in virtue of the causal processes that lead to its activation, but in virtue of *normative* reasons for the fear. McDowell's strategy is to consider normative standards as explanatory of the conditions by which evaluative properties are instantiated. Note that the explanation should avoid circularity: the feature in virtue of which an emotion *E* fits the object cannot be explained by appealing to the evaluative property φ. This is because the fitting attitude analysis explains φ, and φ cannot be, at the same time, *explanans* and *explanandum.* For instance, in the above-mentioned example, Cassio's meriting Othello's anger cannot be established by appealing to the alleged offensiveness – or respectfulness – of Cassio. There are many different norms that may explain the fittingness of emotions (moral, prudential, epistemic, etc.), and it is unclear what normative standards are relevant for establishing how fitting the emotional response is. Different normative standards are often conflated (D'Arms and Jacobson, 2000a). The FA theorist may choose among different options: she may ground the fittingness of emotion on teleological considerations, on practical or epistemic norms (see, for instance, neo-sentimentalist accounts such as D'Arms and Jacobson, 2003).

Instead of discussing all possible variations of FA theories of value, here I want to highlight the main problem that FA theories of value face. FA theories explain evaluative properties by

appealing to normative considerations. The difficulty for the FA theorist is then to spell out what the relevant norms are, without falling into circularity (Deonna and Teroni, 2012). The norms need to be logically independent of any claim regarding affective evaluative properties. Moreover, the emotion needs to be understood without employing affective evaluative properties. Epistemic norms are especially pernicious: an account of the correctness of the emotion such as the one developed in section 5.2, for instance, will face the circularity problem. It seems evident that at least some non-epistemic normative considerations will not be apt for this role.[99] Suppose that I have been threatened by a demon to fear my neighbour. If I do not comply with her commands, she will destroy my well-being. I have therefore prudential reasons to fear my neighbour. Under this menace, it seems practically rational, for me, to do as the demon commands. However, in this case, it seems implausible to think that my fear is fitting. My reason for feeling fear does not have anything to do with the neighbour's danger.[100] So far, the difficulty of identifying the relevant norms remains a pressing problem for the fitting-attitude analysis and a compelling fitting-attitude account of the norms that can be used to analyse evaluative properties has not been given yet (Jacobson, 2011).

This discussion shows that a good account of evaluative properties has to satisfy the following desiderata: (i) evaluative properties are instantiated independently of the subject's emotional responses and (ii) the instantiation of evaluative properties by objects is subject-dependent. The first desideratum will accommodate the possibility that an object *O* lacks the relevant affective evaluative property φ even if the subject – or the members of a designated community – has the disposition to feel *E* when confronted with φ. This will ensure, for

[99] See Deonna and Teroni (2012) for a discussion about the problem of the FA analysis.

[100] On the distinction between practical and epistemic reasons and on the illustration of the evil daemon problem, see Schroeder (2012).

instance, that objects such as European spiders do not qualify as dangerous for a subject who has a phobia of spiders, and this will hold even if most of the people who are presented with a similar situation are disposed to feel fear. Moreover, it will allow the possibility that objects, say asbestos, have a particular evaluative property regardless of the subject's emotional response. The second desideratum will be compatible with the fact that different subjects assign different import to things: the instantiation of evaluative properties is relative to the subject, and in particular to the subject's concerns. An account that satisfies these two desiderata will not be response-dependent, but subject-dependent. Nevertheless, it will be an *objective* account of evaluative properties, where the term "objective" is meant here to signify independence of any particular emotional experience or disposition.

5.4. An Account of Concerns

In the previous section, I argued that a good account of affective evaluative properties requires the instantiation of the property to be relative to the subject's concerns. There is disagreement on what concerns are. Two different types of account have been proposed so far: externalist accounts, according to which concerns are objective goods (Parfit, 1984); and internalist accounts according to which concerns are determined by subject's mental states (see, for instance, Ben-Ze'ev, 2000; Monteleone, 2017; Nussbaum, 2004; Price, 2013; Roberts, 2003).

Here I will briefly survey these two accounts and I concentrate on desire-based account of concerns (Monteleone, 2017). This is because, as I will I argue in this section, the desire-based account is the most compelling account of concerns we currently have.

Objective-list accounts consider concerns as *objective* goods,[101] that is, goods that are independent of the attitudes of the subject (see, for instance, Parfit, 1984). According to this account, concerns are goods that are required for the subject's well-being. They may comprise goods like education, health, shelter, social interaction, etc. These items are not selected in virtue of a subject's mental states such as beliefs, preferences or desires (Fletcher, 2016). They are rather selected in virtue of features inherent to the object itself, and, more precisely, in virtue of the object's normative features (Rice, 2013). Examples are mutual loving relationships, that benefit people not merely instrumentally, but because loving relationships are inherently good.

Although the objective-list account may be a viable normative theory of well-being, this theory is not a convincing account of concerns.[102] This is because, if the objective-list theory was used as an account of concerns, then it would face a circularity problem: we would explain the conditions under which affective evaluative property are instantiated by employing a theory of concerns reliant upon evaluative considerations (perhaps on affective evaluative properties).

An even if it was possible to address the circularity problem,[103] there are individual variations in what is good for us that need to be explained by a theory of concerns. Objective-list theorists argue that the variation in what is good for us is always *objective*. They could allow some variation across individuals, for instance they could accept that people at different stages in their life have different concerns. However, it is questionable that the best explanation

---

[101] Parfit uses the term "interest". For consistency, I will continue to use the expression "concern".

[102] In this section, I do not want to argue for a particular theory of well-being. The aim of this section is to understand the concerns that are involved in determining whether something has a certain affective evaluative property. I will not discuss whether the account of concerns here presented is also good theory of well-being.

[103] One possibility to do so would be to assume that something is valuable if we would be disposed to like it, under an idealised condition (e.g. under the veil of ignorance).

for individual differences regarding concerns can be traced back to only *objective facts* (Railton, 2003). It seems plausible that what matters for a subject need not be what matters to all people, nor what matters to people that are similar to the subject in virtue of mental state-independent features. Conversely, it seems more compelling that concerns bear a connection to our mental states.

Internalist accounts do not face the same problem. Many versions of internalism have been discussed, e.g. concerns as desires (Lazarus, 1991), preferences (Ben-Ze'ev, 2000), as values (Nussbaum, 2004; J. Prinz, 2004), as likes and dislikes (Price, 2013). Concerns also have been considered as more complex states such as the sentiment of caring (see, for instance, Frankfurt, 1969; Seidman, 2016). All these accounts have in common the idea that concerns are individuated by the subject's mental states, and not by features intrinsic to the object of concern itself.

Carolyn Price put forward the like-based account of concerns, according to which concerns are likes and dislikes (Price, 2013). She defines a "like" as a settled disposition to experience something as pleasant, and a dislike as a settled disposition to experience something as distressing. Experiences of something as pleasurable may include physical sensations like the satisfaction of hunger and thirst, emotions, moods, etc. Given that the subject will not always be able to fully know what she likes, she may be mistaken about her real concerns. However, this is not problematic. Sometimes, we use our affective states *also* as an epistemic route to understand what our concerns are. Likes and dislikes may be inherited, e.g. disliking losing support, or learned, e.g. liking to receive a job offer. When a subject acquires a like (or a dislike) that does not reflect the subject's personal experience, Price argues that the like is *mislearned*. As an example of a mislearned like, Price asked us to consider a person who dislikes crowds just because she had a distressing episode in her childhood when she became briefly separated by her mother in a crowd (Price, 2013, p. 192). In this case, the person's fear of the crowd is

mislearned, for Price because it is the result of an overgeneralisation: one single episode should not be enough to shape a person's disposition to find crowds unpleasant. Note that here Price is assuming that there are norms that regulate the mechanisms according to which we acquire likes, that are different from norms that regulate the correctness of an emotion. According to the likes-based account, the fear of the crowd, in the case here considered, would be *correct*, under the assumption that it was proportional to the extent to which the subject actually dislikes crowds, but it would be mislearned.

However, the likes-based theory face problems in handling the following cases: Maisie is angry at Ann, her boss, for slamming her door. Assume that Maisie has a settled disposition to experience shutting doors as distressing. Suppose that the likes-based account is correct. How can we establish whether Ann's slamming the door is offensive or dangerous for Maisie? One strategy would be to understand how slamming the door is impinging on Maisie's concerns. However, Maisie's concerns are so coarse-grained (they are just likes and dislikes), that it is unclear how to establish if Maisie's anger is correct. Why would her anger be correct, and her fear, for instance, incorrect?

Even if it was possible to discriminate between likes that support the correctness of one type of emotional state over another, the likes-based theory would face a further challenge. Consider a subject who starts to feel scared of ladybugs. Initially, her emotion is incorrect: none of her concerns is affected by the ladybug. However, suppose that this person acquires the disposition to dislike ladybugs since she is scared by them. Once the disposition has developed, the mere presence of a ladybug will impinge on the subject's concern: on her dislike of ladybugs. After having acquired the dislike of ladybugs, if the fear of the ladybug is of the right intensity, this will be correct. And this seems an unhappy consequence of the account. Note that this is not an isolated case. Take any emotion *e* directed at one type of object *O*. Once a disposition to

experience *e* directed at that *O* has been created, then, the emotion *e* would be correct, according to Price's like-based account, provided that it has the right intensity.

An alternative to the likes-based account has been proposed by Richard Lazarus. He considers concerns to be desires, or, in his words "goal commitments" (Lazarus, 1991). We have personal goals that express what the subject desires in a particular situation, e.g. the desire that the radiologist does not see any abnormality in one's CT scan, or, more generally, what the subject desires in his life, e.g. being healthy. These desires are "commitments" because they have a strong motivational force that the subject is moved to act upon. However, sometimes a subject does lack a desire for something, say *x*, even though *x* seems, nevertheless, a concern for the subject (Roberts, 2003). Consider your best friend. You care about her, and you want her to be safe. But you may not have an occurrent desire that your friend is safe right now. Nevertheless, her safety seems one of your concerns. This problem may be addressed appealing to dispositions to desire.

Roberts argues that concerns are not identical to having a collection of desires regarding your friend. Suppose, for instance, that your friend is doing a hiking trip in the Alps. Assume that to continue her walk she needs to wade across a torrent with a quite strong current. You may desire that your friend wades in and continues her hiking trip, if a thunderstorm is approaching her location and she needs to find shelter quickly. You may also desire that she does not wade into the torrent if it is a sunny day and an alternative but longer path is available. But you don't need to have these specific desires to be concerned about your friend's wading into the current.[104] For Roberts, in this case, your concern is given by your attachment to your friend. In his account, concerns are not only desires: they are a heterogeneous attitude that comprises desires, aversions and attachments (Roberts, 2003). As he puts it, a concern denotes

[104] A similar example was discussed by Roberts (Roberts, 2003, p. 114).

"desires and aversions, along with the attachments and interests from which many of our desires and aversions derive" (Roberts, 2003, p. 142). Examples of concerns are the attachment to one's own mother and health, the aversion to casinos, and the desire to eat a tasty pizza. Contrasting with desires, attachments are affective dispositions that have "diffuse", or general goals (Roberts, 2003). For instance, in the case of your friend walking in the Alps you care about your friend's safety, and this would mean you would be disposed to desire her to wade through the torrent when a thunderstorm is approaching, helping her to wade through the torrent if she needs, being happy if she succeeds, or desire that she does not wade through the torrent if she could avoid it without problems.

Roberts's account of concerns is heterogeneous, and it's not clear why these diverse things belong together in a single category called "concerns". John Monteleone suggests that what these different mental states have in common is their being all states of *non-indifference* that are the result of a *disposition* to desire (Monteleone, 2017). As he puts it, *concerns* are "a stable disposition to have a range of desires in different counterfactual situations" (Monteleone, 2017, p. 194). What unites these different desires, for Monteleone, is their sharing a common *focus*. The focus of a desire is the background object that renders intelligible the relations between different desires. For instance, Iago's desire to help Roderigo to trust Desdemona, Iago's desire that Emilia takes the handkerchief that Othello gave to his wife and Iago's desire for Cassio's death share a common focus: Othello's downfall. The common focus explains the relationship between these apparently unrelated desires.

Desires-based accounts like Roberts's and Monteleone's may be criticised on the grounds that there may be things that are important to us even if we do not desire them. For instance, a person who is in a depressed mood may not care about her cleanliness or her health, although, it may be argued that health and cleanliness are important for her well-being. Although this may be considered a problem for a theory of well-being, it does not seem a pressing problem

for a theory of concerns. One could endorse a desire-based theory of concern, and, at the same time endorse an objectivist account of well-being. In this way, we could still say that cleanliness and health are important for the well-being of the subject, understood in an objective way, while at the same time accepting that the subject's concerns differ from the elements that ground the subject's wellbeing. Discussing which theory of wellbeing we should adopt falls beyond the scope of this chapter, and I will not discuss it here.

To conclude, in this section, I have argued that objective-list theories of concerns are not compelling because they face the circularity problem and do not satisfactorily account for individual variations. I have claimed that internalist accounts fare better than externalists accounts, and I have suggested that we should favour a desire-based account of concern.

5.5. Relativism About Evaluative Properties

I defend a form of relativism about evaluative properties. As in any relativist account, the possession of some properties, in this case, the possession of affective evaluative properties, is relative to some parameter *P*. I argue that the parameters that determine whether something has an evaluative property are determined by a subject's *concerns*. Here I am going to assume that a subject's well-being is defined by the subject's disposition to desire, following Monteleone's account (Monteleone, 2017). Generalising from this, I propose that an object *O* has the evaluative property φ, for a subject *S*, if and only if it has an impact μ on the subject's *concern c*. The type of impact μ describes how things in the world impact on our concerns.

Consider, again, King Duncan's death. His sons, presumably, enjoy their father's company. Enjoying the King's company is, therefore, one of Donalbain and Malcolm's concerns. When the King dies, this event has an impact on Donalbain's concern: namely, his desire to enjoy his father's company is frustrated. It is then possible to say that, given Donalbain's concerns, the King's death is a loss for him. More generally, when a particular subject's concerns have been spelled out, given the subject's circumstances, it will be possible to understand the evaluative

property of a situation, depending on the variable μ, that spells out how the situation affects that subject's concerns.

To reiterate this with the example previously used, whether something is a loss for Lady Macbeth or Donalbain and Malcolm depends on descriptive facts about the world and how these facts affect their concerns. It is a descriptive fact that King Duncan's death brings about his impossibility to enjoy life and their sons' impossibility to enjoy his company in future. This descriptive fact has an impact on Donalbain and Malcom's concerns: they care about their father's safety and the King's death affects their concerns about their father. This descriptive fact, in this account, is described by the variable μ.

Conversely, the parameters of the evaluation are given in relation to the subject's concerns that account for the significance of the situation. The impact of King Duncan's death, for instance, is clearly significant: it is not merely the impossibility to enjoy company with a person randomly chosen, but the impossibility to enjoy the company with a person they deeply care about. Something is significant if and only if, given the actual state of the subject, it impinges on the subject's concerns. Something is not significant if it does not have such an impact.[105] Once the descriptive facts about a situation are known, and the subject's concerns are identified, it will be possible to identify how the situation in question impinges on the subject's concerns. Suppose that when King Duncan dies, Malcolm is kept in the dark about his death. Suppose, again, that the King's life is one of his concerns. Under this assumption, King Duncan's death is a loss for him, even if he does not know that King Duncan is dead.

Without specification of the parameters of the evaluation, which are the subject's concerns, the property attribution is indeterminate. Consider, as an illustration, a situation where the

[105] The problem of understanding under what condition a situation is personally significant has been called the "problem of emotional significance" (Price, 2013).

subject's concerns are in tension with each other. Rose, for instance, may be concerned on the same night about socialising and also about receiving some intellectual stimulation. If she stayed at home reading a book, the intellectual stimulation concern would be fostered, whereas the socialisation concern would not. If she decided to meet up with a friend instead, the reverse would have happened. In a situation like this, the evaluative property of a situation will not be an *all things considered* good *and* bad for the subject. Without specifying the parameters of the evaluation, it is not possible to determine whether meeting up with a friend will be all things considered good. I take this to be a positive feature of the account because it helps to handle difficult cases. The situation would be good only relative to the former concern, and bad relative to the latter concern.

This discussion has shown that we can objectively ascertain whether an emotion is correct: once the subject's concerns are known, and the descriptive facts about situations are settled, it is possible to determine whether the object of the emotion has the relevant evaluative property to the extent indicated by the intensity of the emotion.

5.6. Summary and Conclusion

In this chapter, I have offered an account of the correctness conditions of emotions that is consistent with the objectivity of evaluative properties. In section 5.1, I presented an outline of the correctness conditions of emotions, expanding on D'Arms and Jacobson's norms of "shape" and "size". In section 5.2, I argued that the intensity of emotions can be quantified, and that our ability to discriminate the intensity of emotions constrains the norms of correctness that regards their intensity. In section 5.3, I examined the main existing approaches to understanding evaluative properties and I outlined some desiderata. In section 5.4, I compared internalist and externalist accounts of concerns and I argued that internalist accounts better capture what concerns are. Finally, in section 5.5, I defended a relativist account of evaluative properties.

CHAPTER 6: TOWARDS EPISTEMIC RATIONALITY NORMS

ABSTRACT

How should epistemic rationality norms be generated? In this chapter, I argue that epistemic rationality norms should be selected as the norms that are conducive to entertain correct emotions. Although subjects are *often* better off having epistemically rational emotions – at least from an epistemic point of view – this does not always hold. Epistemically rational emotions are not always practically beneficial. Nor do epistemically irrational emotions always disrupt the subject's ability to achieve their practical goals. Moreover, epistemically rational emotions may, in some cases, impair the subject's "epistemic functionality", which is the subject's ability to achieve her epistemic goals. Focusing on the emotion of worry, I argue that when this happens, the worry is "epistemically guilty".

INTRODUCTION

In the previous chapter, I offered an account of the correctness conditions of emotions, and I argued that the correctness of emotions could be mind-dependent *and* also objective. We may not be in the position to know all the relevant data for assessing whether an emotion is correct, for example, we do not always know exactly what concerns a subject has, and we may not be able to know the intensity of a subject's emotion, nor all the relevant facts about the situation the subject is in. However, I argued that *if we had* all the relevant information, we could establish whether a given emotion occurrence is correct, objectively.

However, aspiring to have correct emotions is difficult without clear guidelines. How can we improve our chances to have emotions that are correct? We are rarely in the position to know all the relevant data required to establish whether an emotion is correct or incorrect. We need other principles to help us understand which emotions are likely to be correct in our everyday life. We need guidelines that could help us to form emotions that are likely to be

correct. Here, I suggest that epistemic rationality norms could be such guidelines. In this way, epistemic rationality norms would function as a "regulative ideal" that we should follow, at least in epistemically relevant contexts.

Following epistemic rationality norms of emotions helps us to experience emotions that are more likely to be correct. Although epistemic rationality norms, if followed, often ensure that the subject is in a better epistemic position, this does not always hold. There are cases in which epistemically rational emotions may affect the subject's ability to function well epistemically. Consider, for instance, a subject who experiences rational fear of having a deadly disease. This fear may affect the subject's ability to engage with epistemic tasks such as solving a philosophy puzzle. The motivational and attentional component of the emotions may disrupt the subject's ability to function well epistemically. Moreover, in certain situations, epistemically irrational emotions could give important practical and epistemic benefits. Consider a surgeon who has to perform a difficult medical procedure. The success of the operation depends on her ability to keep her hands steady and take difficult decisions. Experiencing fear with less intensity than what epistemic rationality requires may be conducive to the success of the operation. In this chapter, I consider similar cases and I specify the conditions under we are not obliged to follow epistemic rationality norms.

In section 6.1, I argue for a particular strategy for identifying epistemic rationality norms; a strategy that, if followed, will help us to increase the possibility of entertaining the correct emotions. In section 6.2, I present an example of an epistemic norm of emotions. But should we always follow rationality norms? Following rationality norms is *often* epistemically and practically beneficial. However, in some circumstances, epistemically *irrational* emotions may be not only practically, but also epistemically beneficial. I discuss such cases in section 6.3. *Rational* emotions may also be epistemically problematic. In section 6.4, I provide a case for the claim that, in some circumstances, epistemically rational emotions may have important

epistemic costs. I focus on a particular type of emotion: the emotion of worry. I present a scenario where the worry is epistemically rational and yet it impairs the subject's cognitive functioning, that is, the subject's ability to achieve her epistemic goals. When this happens, I will argue, the worry is epistemically "guilty" because it disrupts the subject's ability to function well epistemically.

6.1. How Should Epistemic Norms be Generated?

John Pollock distinguished between two approaches that may guide the selection of epistemic norms: norm internalism and externalism (Pollock, 1987). According to norm internalism, the norms are selected in virtue of facts that are internal to the subject. Conversely, according to norm externalism, the norms are selected in virtue of facts that are not limited to mental states that are internal to the subject. Here, I follow an externalist approach: the norms would be selected because they are conducive to the formation of correct emotions. Thus, these epistemic norms are not selected only in virtue of facts regarding the subject's mental states, but also in virtue of facts that are external to the subject, specifically, in this case, the norm's conduciveness to the epistemic goal of having correct emotions.

The norms are prescriptive. They dictate how emotions should be formed *if the subject wants to have epistemically correct emotions*.[106] In this sense, these norms are action-guiding and have an important directive function: they indicate what the subject should do if she wants to maximise her possibility of having correct emotions. When the epistemic norms are all met, the emotion is epistemically rational. When the epistemic norm is not satisfied, the emotion is epistemically irrational. But what are the conditions under which we are not obliged to follow the norms? This issue will be addressed in section 6.4.

[106] This entails that it is not the case that we always have the obligation to follow epistemic rationality norms.

In what follows, I suggest that the epistemic rationality status of an emotion depends on facts about its aetiology. To argue for this dependence relation, I use a strategy similar to the one adopted by Susanna Siegel (Siegel, 2017). Siegel argued that perceptual experiences might be the product of a mental process that resembles an inference. As conclusions of inferences are epistemically dependent on the rationality of their premises, or grounds, so psychological processes that have an inferential aetiology are epistemically dependent on their inputs, which act as premises. When a perceptual experience is the result of an inferential process, for Siegel, the irrationality at one level of the aetiology of a perceptual experience "infects" further elements of the aetiology on which the mental state is based until the epistemic rationality of the mental state is compromised (Siegel, 2017).

I do not need to endorse Siegel's claim that the rationality of perceptions is affected by facts about its aetiology for my argument to work. What I need to endorse is the claim that if a psychological process that leads to an emotion resembles an inferential structure, the epistemic rationality status of the result of the process – its "conclusion" – depends on the epistemic rationality of its inputs, or "premises".

Schematically, we can represent the aetiology of emotions as follows: first, a mental state, for instance, perception, belief, memory, or imagining – the emotion's cognitive basis – serves as input for the appraisal system. The appraisal system then stores the disposition to transition between certain classes of cognitive basis and an affective value, such as danger, loss, and offensiveness. When the input matches the antecedent of the disposition, an occurrent evaluation is activated and the emotion corresponding to the affective value is triggered. It is possible to represent the appraisal system as operating using inferential rules having the structure <if *x* then *y*>, where the antecedent is an object or state of affairs and the consequent is an affective evaluative property. Note that this does not mean that causal evaluations have always propositional content (see chapter 2 on this point). In chapter 7, I will use Nick Shea's

taxonomy (Shea, 2014) to explain the structure of the appraisal system and the causal evaluations stored in the "automatic appraisal mechanism", one of the systems that psychoevolutionary theorists claim could trigger emotions. In that chapter, I will argue that the appraisal system is best understood, using Nick Shea's taxonomy, as involving implicit representations, where implicit representations are dispositions to transition between two mental representations. Implicit representations do not have propositional content, as they are merely dispositions to transition between two mental representations. Nevertheless, they may be represented as having an <if *x* then *y*> structure, as the activation of representation *x* is disposed to trigger representation *y*, where *x* is the cognitive basis of emotion and *y* is the affective evaluative property. When the antecedent is met, and the emotional disposition is activated, the relevant occurrent causal evaluation is tokened, and the emotion is triggered.

The claim that emotions are formed according to an *almost* inferential aetiology is not new, as it has been developed by Jesse Prinz (Prinz, 2004, pp. 237-238). According to Prinz, emotions are like conclusions of arguments, as it is possible to consider the elicitors of emotion as *reasons for* the emotion. However, Prinz does not recognise that in addition to the emotions' cognitive bases, also the emotions' causal evaluations can be considered reasons for the emotion. Suppose you are scared by a steep cliff. I may ask you why you are scared, and you may reply that you are scared because the cliff is steep (the emotion's cognitive basis). You may also reply that you are scared because if there is a steep cliff, then there is a danger (the appraisal system's implicit dispositional representation with conditional content). If this is the rule that is responsible for the triggering of your causal evaluation (this particular cliff is dangerous) then, this rule seems another important *reason for* the emotion. When the antecedent of the disposition is met, the occurrent causal evaluation is triggered, e.g. (the steep is dangerous), and the emotion is set off.

The emotion stage is composed at least in part by an evaluation, which can be seen as the conclusion of an argument. The aetiology of emotion is represented below.

Figure 4: the aetiology of emotion

So far, I identified three stages of the formation of an emotional experience: the cognitive basis, the appraisal system, which stores implicit dispositional representations and generates occurrent causal evaluations (see chapters 2 and 3, and chapter 7), and the emotion. Here I suggest that the epistemic rationality norms should apply to all three levels of an epistemically rationally assessable emotion's aetiology. As the process that leads to the formation of the emotion resembles an inferential process, the epistemic status of all the elements of the inference affects the emotion's final epistemic status.

6.2. Towards Epistemic Rationality Norms

In what follows, I will present an example of an epistemic norm that could be applied to emotions. I do not have space here to discuss all epistemic norms that may apply to the emotion. My aim here is to discuss a concrete example of an epistemic norm that applies to the three stages of emotional experience. So, for instance, we could say that an emotion *E* is epistemically rational if and only if:

- E's cognitive basis, appraisal system and emotional experience are evidence-responsive.

The cognitive basis of the emotion is therefore subject to requirements to be formed in accordance with the evidence. In this way, the rationality conditions of the emotion depend on the epistemic rationality conditions, if any, of the cognitive basis of the emotion. However, not

all the mental states that constitute the cognitive basis of the emotions are subject to epistemic rationality conditions. In chapter 4, I presented an account of the epistemic rational assessability of mental states according to which a mental state is epistemically rationally assessable if and only if it satisfies four properties: (1) the mental state *m* is informationally plastic; (2) the mechanism *P* that maintains *m* is generally receptive to the evidence; (3) *m* has a descriptive content; (4) *m* contributes to the subject's view of the world. This account could be used to assess whether the cognitive basis of the emotion is open to epistemic rational assessment. Perceptual states and imaginings, for instance, seem very often not rationally assessable. In this case, the epistemic status of the emotions is not modulated by the epistemic status of *an a-rational* cognitive basis. However, if the cognitive basis of the emotion is rationally assessable, and the rationally assessable cognitive basis are formed on poor grounds, then the epistemic status of the emotion will be compromised. This is because the irrationality of the cognitive basis is transferred to the further stages of the emotion process.

Consider, for instance, a subject who believes on ill grounds, that European spiders are venomous. When this subject sees a European spider, she feels fear. Note that her appraisal process, in this case, seems to follow the evidence: the subject has evidence that venomous objects are dangerous and activates fear accordingly. However, she lacks evidence that European spiders are venomous. However, the subject's emotion is based on the belief that European spiders are poisonous. In this case, the emotion is epistemically irrational because the belief on which the emotion is based is not responding to the evidence.

Note that the rationality conditions of emotions do not require that the mental states that constitute the emotion's cognitive basis are correct or true. As George Pitcher argued, emotions may be based on a false belief and yet be intuitively rational (Pitcher, 1965). To illustrate this Pitcher asks us to imagine a situation where a subject is scared by a man's voice that is threatening her. Even if the voice was produced by a fake recording created by the subject's

friends to play a joke on her – and therefore the belief would be false – the subject's emotion would be epistemically rational because it would satisfy the normative criteria presented above.

The epistemic rationality norms of emotion apply to the particular evaluation that is stored in the appraisal mechanism and becomes occurrent when a given emotion is triggered. This particular evaluation is responsive to the evidence if it is consistent with one's evidence at the time. For instance, suppose that Amy sees a spider, and believes that the spider she sees is not dangerous. Suppose that her belief is epistemically rational: she has seen many spiders in the past, and none of these spiders harmed her in any significant way. However, when she sees a spider, she feels fear. In this case, the cognitive basis of her emotion is her perception of the spider. Since the perception, in this case, is not irrational, the epistemic status of her belief is not affected. However, her appraisal mechanism seems to be faulty. Her appraisal system stores an implicit disposition to transition between spiders and danger. This disposition could be described as an "if then" rule: if there is a spider, then there is danger. Her appraisal system evaluates this particular spider as dangerous and fear is set off. Amy's appraisal system seems defective: Amy lacks evidence for the implicit disposition to transition between spider-like features and danger. This fear seems to be epistemically irrational precisely because the emotion's appraisal system is not responding correctly to the evidence.

Lastly, the epistemic norms also apply to the final level of the aetiology of the emotion: the emotional experience. Suppose that Sarah has to deliver a talk, which she is worried about doing. Her cognitive basis is not faulty: her belief that she is about to deliver a talk is epistemically rational. But is her appraisal mechanism, in this case, responding to the evidence? Suppose that her appraisal system associates delivering a talk with danger. Suppose that Sarah has some evidence that conveying a bad impression to the public may be harmful to her career. The disposition to transition between delivering a talk and danger *seems prima* facie responding to the evidence. However, let us suppose that Sarah knows that she is very skilled

and that she has already proved to be an expert in her field. Consider a scenario where Sarah feels terrified of giving a talk. In this case, the intensity of her emotion is not responding correctly to the evidence. The emotion is epistemically irrational as a result of it.

Typically, emotions are directed to an object. However, as I argued in chapter 2, the intentional object of the emotions may be different from the intentional object of the causal evaluation that triggers the emotion. Epistemic norms could regulate the selection of the object of the emotion. The emotion responds to the evidence when its object is selected in accordance with the evidence. Suppose, for instance, that Sabrine hears a loud noise coming from the kitchen and she feels scared. Suppose that her auditory experience is not irrational and that her appraisal mechanism is responding to the evidence, since she has evidence that loud sound may be associated with danger. However, Sabrine's appraisal system triggers an emotion of fear directed at a mouse in her apartment, which she considers to be the cause of the noise. Suppose that she has good grounds to believe that the noise she just heard is not caused by a mouse. In this case, again, the emotion is not responding correctly to the evidence. Therefore, it would be epistemically irrational.

The rationality conditions of emotions do not require the actual existence of the object at which the emotions are directed. Emotions directed at objects that do not exist have been called unfounded emotions (Taylor, 1975, p. 392). Suppose, for instance, that Susan hears an unexpected noise from her kitchen, and she fears that an intruder may be in her apartment. Assume also that Susan's appraisal system forms this emotion in accordance with the evidence. Despite Susan's suspicion, there is not an intruder in her apartment, and the noise is actually produced by a cat knocking over a pack of sugar. Susan's emotion seems intuitively rational. Conversely, suppose that Susan* is an anxious person and is obsessed with the possibility that intruders will sneak into her apartment. When Susan* hears a noise coming from the kitchen, she fears that an intruder may be in her apartment. Again, in this situation, there is no intruder

in the apartment. The difference in the epistemic status of Susan's and Susan*'s emotion is not predicted by the emotion's correctness, but by the extent to which the emotion and its psychological precursors – the emotions' cognitive basis and the appraisal system – respond to the evidence. In this way, the rationality norm of emotion qualifies as *structurally internalist*: the structure of the epistemic norms is spelt out entirely in terms of a relationship between mental states.

Having epistemically rational emotions is often beneficial. Epistemically rational emotions are likely to be correct emotions. An accurate understanding of the evaluative import of the situation, in a large variety of contexts, leads to better decision making: subjects are better able to address the impact that the world has on their own concerns, and they are also better at fostering their concerns. Suppose, for instance, that I want to safely navigate in the environment: it is important for me to recognise dangerous objects or situations so that I can be sure not to put my life at risk. Recall the boatswain's fear in Shakespeare's play *The Tempest.* The fear is informing him that he and his crew are in a serious situation. Having had this information, the boatswain can then act to address the danger: he spurs his crew to work strenuously to avoid the sinking of the ship.

In contrast, when emotions are epistemically irrational, subjects are prone to make mistakes about the evaluative import of the situation that they are in. Their actions will tend to be ineffective because they are less tailored to the evaluative import of the situation that the subjects are in. Moreover, they will also be likely to form bad evaluative beliefs. Consider, for instance, Othello's jealousy at Desdemona. The emotion is incorrect since Desdemona is faithful to Othello. Moreover, Othello's emotion is also epistemically irrational: Othello's emotion fails to be formed in accordance with the evidence. His appraisal mechanism overlooks a large amount of evidence of Desdemona's faithfulness, and overrates the scant evidence – the handkerchief in Cassio's hand – that she may have had an affair with Cassio.

The intensity of the jealousy is not formed in accordance with the evidence. This jealousy is taken at face value by Othello, who forms, on this basis, judgements about Desdemona's actions and about her character. Moreover, Othello's jealousy leads him to make extremely bad decisions, such as insulting and threatening her. These terrible consequences culminated in his decision to kill Desdemona.

The potentially dramatic consequences of epistemically irrational emotions are also often highlighted in children's education. This theme is often present in fairy tales. For instance, in *Little Red Riding Hood*, the protagonist fails to have epistemically rational fears at the right time. Little Red Riding Hood does not feel fear of wandering around in the wood alone, nor is she scared by the wolf, when she encounters him in the wood, despite having some evidence that wandering in the wood and talking with strangers may be dangerous, since her mother explicitly warned her not to do so.

I am aware that I did not provide a fully developed account of the epistemic rationality of emotions, and that I have left many questions unanswered, which could not have been considered given the limited space available for this thesis. In particular, I have not addressed questions that surround the nature of the relevant type of evidence required at each stage of the emotional experience. The account offered in this chapter should be understood as a first attempt at providing a sketch of how epistemic rationality norms should be generated.

6.3. Can Epistemically Irrational Emotions Be Beneficial?

Although we are often better off with epistemically rational emotions, this is not always the case. In some circumstances, epistemically irrational emotions may be practically beneficial. Here I discuss the benefits of epistemically irrational emotions focussing on medical practice. A careful examination of the epistemic rationality of emotions in the medical context falls beyond the scope of this chapter. My aim here is more modest: using realistic examples, I want

to suggest that epistemically irrational emotions, on some particular occasions, may be beneficial for the subject.

I draw my examples from Danielle Ofri's book *What doctors feel.* The author suggests that doctors ought to experience a certain level of fear while practising medicine because of the risk of harming the patient is almost always present (Ofri, 2013).[107] Ofri relays the story of her own child's birth, where suddenly the values of her baby were not in the normal range. She describes how her doctor's expression and manners immediately changed from relaxed to scared, now seriously focussed on the problem just found. The doctor had evidence that the baby he was caring for, at that moment, seemed to have a serious problem and his emotion was responding to that evidence. The doctor's emotion, in this case, would be epistemically rational: the emotion is formed in accordance with the evidence. Ofri suggests that fears like this are not only correct – or "appropriate", as she puts it – but also beneficial for medical practice. As she puts it:

> "Appropriate fear, as I witnessed in my obstetrician, can be crucial for good medical care, especially during critical situations. All eventually worked out well for me and my baby, but I gained an appreciation for how fear in the right dose might serve the doctor – and the patient – well. Being aware of our fear and figuring out how to titrate it appropriately is a vital skill for a doctor. Our patients' lives may depend on it" (Ofri, 2013, p. 94).

For doctors, feeling a certain intensity of fear of harming the patient may be epistemically rational on some occasions. Despite being epistemically rational, the fear may interfere with the doctor's ability to make decisions. Ofri describes an episode of her life in which, after only one week of experience with the medical consult service, she had to attend a person in cardiac arrest (Ofri, 2013, p.66). In this circumstance, she had to direct the resuscitation procedure, but she was overwhelmed by the fear of making the wrong decision; she was unable to understand

[107] An assumption here is that the life of the patient is one of the doctors' concerns.

the exact nature of the problem of the patient – the medical evidence was not clear-cut – and she was hesitant to instruct staff on the action to be taken. It is difficult to ascertain with certainty whether her emotion, in this circumstance, was epistemically rational. But suppose, for the sake of the argument, that her emotion was so: assume that she cared about the life of the patient and the intensity of the emotion she experienced was proportional to the danger of the situation. This would be a scenario where an epistemically rational emotion obstructs the ability to make decisions.

Emotions do not only provide us with cold evaluative information about what is significant to us, but they also involve a series of bodily changes that prepare us to respond to such significance (Griffiths, 1997). Fear, for instance, does not only inform us that we are facing a certain danger. It also involves autonomic responses that prepare us to fight, freeze or flee in the face of a threat (LeDoux, 2015). These changes mobilise our cognitive resources: they consume our attention, increase our sensitivity to elements in the world that are emotionally salient and incline us to accept the information emotions bring to us (Brady, 2009). Here I want to suggest that the bodily and cognitive changes that emotions bring about may impede efficient action on some occasions, which include some occasions when the emotion is epistemically rational.

Because of these effects of fear, during a critical situation like the scenario mentioned above, it may be beneficial for a doctor to experience emotions that are of less intensity than the ones required by epistemic rationality norms. Ofri, for instance, narrates that another doctor successfully took over the resuscitating procedure she was struggling with. This doctor is described as acting calmly and with confidence and calm temper. In a dramatic situation like this, where it is necessary to make very important decisions in a short amount of time, it may be beneficial for a doctor not to feel overwhelmed by fear to be better able to perform better the medical procedure. The benefit, in this case, would be a practical benefit. The subject will

be better able to carry out her own practical goal of performing well the resuscitation procedure. However, the benefit may also be epistemic. Feeling fear with less intensity than the subject ought to experience may free cognitive resources that may help to obtain a better diagnosis of the medical problem.

In the situation just described, Ofri struggled to understand whether the electrocardiogram showed signs of elevated potassium levels. Understanding whether the patient has elevated potassium levels is an epistemic task. Her intense fear, however, partly interfered with this task: she was distracted by the possibility of a mistaken diagnosis. She describes her experience as follows:

> "Okay, the T waves. Maybe they looked a little peaked. Peaked T waves were indicative of elevated potassium levels, except when they weren't. And except when they looked peaked but weren't actually peaked. They did look sort of peaked, I thought to myself, but maybe they were just hyperacute T waves, or maybe they were enlarged from early repolarization. Or maybe they were just big. I was too terrified to trust myself on anything" (Ofri, 2013, p. 66).

Experiencing fear to a lesser degree may also bring epistemic benefits: in a situation similar to the one Ofri experienced, for instance, feeling fear to a lesser extent may reduce the number of cognitive resources mobilised in considering interpretations of the electrocardiogram result that counter her initial hypothesis. These examples suggest that epistemically irrational emotions may, in some exceptional cases, have practical and epistemic benefits.[108]

Epistemic benefits may not occur not only when emotion is epistemically irrational because its intensity is less intense than it epistemically ought to be. They may also occur when it is irrational because it is more intense than it epistemically ought to be. For instance, scientific research shows that feeling happy improves productivity in the workplace (Oswald, Proto, and

[108] This holds in the short term. Whether irrational emotions like this are practically and/or epistemically beneficial in the long term is an open question.

Sgroi, 2015). In the experiment conducted by Oswald and colleagues, one group of participants viewed a comedy movie clip, where the control group did not see the clip. All participants were then asked to perform as many additions they could in 10 minutes. The participants who had viewed the clip were more productive than the control group. The authors of the paper concluded that higher happiness is associated with higher productivity. Now, suppose that a subject wants to be more productive than she is. She may have practical reasons to feel happy precisely because happiness would be beneficial for improving her productivity. This benefit would occur irrespectively of her evidence for feeling happy. Even if the subject lacked evidence that is consistent with high levels of happiness, feeling happy would be beneficial for the subject because she would be better able to improve her productivity.

In this section, I argued that, in some cases, irrational emotions may be beneficial. In the following section, I will focus on a single emotion type. I will discuss a scenario in which the emotion of worry is epistemically rational, but also epistemically problematic.

6.4. Can Epistemically Rational Emotions Be Epistemically Problematic?

In the previous section I argued that in some cases, irrational emotions may be practically beneficial. Consider a subject who is worried about the result of her MRI test. Her worry seems rational: she has good reasons to be worried because her oncologist told her that there is a 50% probability that she has cancer. In this situation, the worry is epistemically rational: it is responding correctly to the subject's available evidence. Although the worry is epistemically rational, it also seems epistemically problematic. When she worries about her health, she is not able to concentrate, her problem-solving abilities are impaired, and she underperforms at work. In cases like this, I shall argue, the worry affects the subject's "epistemic functionality", i.e. the subject's ability to gain knowledge.

At the moment, in philosophy of emotions, we lack an account of the relationship between the epistemic rationality of emotions and the subject's epistemic functionality. In the following sections, I will contribute to the understanding of this relationship, arguing that the emotion of worry has the potential to be "epistemically guilty" when it is rational but impairs the subject's epistemic functionality. Firstly, I will present the notion of epistemic functionality (section 6.4.1). The emotion of worry has close ties with motivation and attention. In section 6.4.2, I will make a case for the claim that the emotion of worry could impact the subject's epistemic functionality in virtue of its attentional and motivational components. In section 6.4.3, I argue that an epistemically rational emotion that impairs our epistemic functionality should be considered "epistemically guilty".

6.4.1. Epistemic Functionality

The epistemic functionality is the agent's ability to achieve their own epistemic goals (Bortolotti and Sullivan-Bissett, 2018). Examples of epistemic goals include the ability to acquire, retain and use information (Bortolotti, 2018).

A subject is able to function well when they can employ their cognitive resources for addressing the problem at hand. A full and detailed account of epistemic functionality falls beyond the scope of this chapter. Here I only highlight some of the features that we commonly recognise as required to function well epistemically. For instance, we recognise that to work well epistemically, the subject needs to have good attentional capacities, the ability to concentrate on their epistemic task, and the ability to exchange information with other people. Moreover, the subject also needs to remember what they need to do, and they need to be able to prioritise tasks effectively.[109]

[109] This list is not meant to be exhaustive.

It has been already shown that the subject's epistemic functionality may be compromised in patients suffering from psychopathologies such as schizophrenia (Bortolotti, 2016) and Alzheimer's disease (Bortolotti and Sullivan-Bissett, 2018). Here I focus on subjects for whom the experience of worry is rational, and in the following section, I will argue that in some contexts, the worry may also disrupt the subject's epistemic functionality.

6.4.2. Can Rational Worry Disrupt the Subject's Epistemic Functionality?

In this section, I will focus on the emotion of worry. I will consider worry as an intentional mental state, directed towards a particular object. Like other emotions, the emotion of worry has close ties with attention and motivation (Brady, 2013; Faucher, Christine, 2002; Tappolet, 2016a). Worry directs our attention to the source of danger (de Sousa, 1987), limits the range of information and the options we consider when deliberating (Damasio, 2006), keep us interested in looking for evidence that the particular object is really dangerous, and motivate us to act for a specific goal, in this case the goal of avoiding danger (Frijda, 1986).

It has been argued that epistemically irrational cognitions may, in some contexts, being able to restore the agent's epistemic functionality (Bortolotti, 2015; Bortolotti and Sullivan-Bissett, 2018; Puddifoot and Bortolotti, 2018). For instance, Bortolotti argues that adopting elaborated and systematized delusions may restore the epistemic functionality of people with schizophrenia and provide relief from anxiety (Bortolotti, 2016). In this chapter, I focus on the flipside of this phenomenon: I make a case for the claim that epistemically rational worry may disrupt the subject's epistemic functionality and I will discuss the epistemic status of such a worry.

In what follows, I present a scenario of an epistemically rational worry that is related to the fear of cancer recurrence (henceforth, FCR). FCR is defined as "fear or worry that cancer will return or progress in the same organ or in another part of the body" (Simard, Savard, and Ivers, 2010, p. 361), and it is commonly experienced in cancer survivors and their carers (Simard et

al., 2013). Consider Jess, for instance, who was diagnosed with bowel cancer three years ago. After her successful treatment, she was told by her doctor there was a 50% probability that cancer would return. She is now waiting for the results of an MRI test for her yearly check-up. During her two weeks' waiting time, she experiences intense worry.

We could consider her worry as epistemically rational, as she has evidence that there is a significant risk that cancer will reoccur. If it does, her life projects and well-being will be severely affected: she will not be able to undertake the career she was craving and she will not be able to fulfil her desire to become pregnant. The worry is delivering accurate information regarding the risks of cancer: the risk should not be ignored, and heavily affects the subject's life and well-being. Experiencing this fear may be beneficial for Jess insomuch as it keeps her motivated to attend regular check-ups, thereby maximising the possibility of early diagnosis should the cancer return and may motivate her to entertain a healthy lifestyle. Worry may also be epistemically beneficial because it helps her to understand that her health and is highly important to her.

However, her worry may also prevent her from achieving her epistemic goals. This is because the emotion of worry, like other emotions, is closely tied with attention and motivational tendencies. Emotions direct subject's attention to aspects of the situation that are constructed as important to us (de Sousa, 1987). They motivate us to look for reasons that bear on their correctness (Brady, 2013). Although this has the benefit of facilitating our assessment of the evaluative import of the situation, it also comes with cognitive and epistemic costs. Our attentional and cognitive resources are limited, and, as Brady argues, emotions *consume* attention (Brady, 2013). Suppose that Jess is a mathematician and is trying to solve a difficult mathematical puzzle. Worry could render concentration difficult, taking attention away from the task (Sweeny and Dooley, 2017). Despite being a highly skilled mathematician, Jess is not

able to progress in her research. Her attention span is limited because her worry draws her attention to the possibility that she has cancer again.

Moreover, the mathematician's worry motivates her to look for reasons bearing on the risk involved in the possibility of cancer recurrence. She is not able to focus on her mathematical problem because she repeatedly considers the possibility that her doctor will give her bad news. It has been shown that people with intense worry have less residual working memory (Hayes, Hirsch, and Mathews, 2008). While she worries, Jess cannot easily memorise well what she reads, and she is not fully attentive when her colleagues discuss with her what strategy to adopt to tackle the problem.

In this example, we have seen that the worry, despite being rational, could compromise the subject's epistemic functionality: it is affecting Jess's attentional capacity, and her the ability prioritise and concentrate on solving the mathematical problem. Her worry also reduces her memory and her ability to exchange information with other people.

In the following section, I will argue that worries like Jess's ones are epistemically "guilty" because they impair the subject's epistemic functionality.

6.4.3. Is Worry Epistemically "Guilty"?

Epistemically rational emotions of worry respond correctly to the subject's evidence. As in the above-mentioned example, these emotions may also be correct: they are representing the threat that the possibility of having cancer again has for Jess's well-being. However, in cases like this, the emotion of worry is also epistemically problematic: it compromises the subject's ability to function well epistemically.

In consequentialist terms, we could say that epistemically rational worries that disrupt the subject's epistemic functionality deliver significant *epistemic costs*. They may compromise the subject's ability to concentrate, to weigh different options, to select the relevant information for the task at hand and may affect the social communication that is conducive to profitable

knowledge exchanges. In this sense, epistemically rational worries that compromise the subject's epistemic functionality are not free from faults. They are epistemically problematic because they impair the subject's ability to achieve their epistemic aims.

What could we learn from this discussion? Providing an epistemic assessment of the emotions is more complex than we previously expected it to be. This is due to attentional and motivational components that are usually involved in emotional experiences. These components affect the subject's ability to achieve epistemic goals in ways that could be independent of the emotion's epistemic status. On the one hand, on one occasion attentional and motivational components of irrational emotions may deliver important epistemic benefits, as the scenario presented by Danielle Ofri seems to suggest. On the other hand, in other contexts attentional and motivational components can disrupt the subject's ability to achieve epistemic goals, even if the emotion is epistemically rational. This discussion suggests caution: not all epistemically irrational emotions are detrimental for the subject, and some epistemically irrational emotions could be beneficial to the subject. It could be suggested, therefore, that in situations like Jess' scenario – where emotions are significantly affecting our epistemic functionality – we are not required to follow epistemic norms. Similarly, I suggest that in situations analogous to Ofri's scenario described in section 6.3 – where irrational emotions enhance our epistemic functionality – we are also not required to follow epistemic norms. We may also make room for the possibility that in certain contexts, subjects may be excused for having epistemically irrational emotions.

6.5. Summary and Conclusion

In this chapter I presented a tentative view about the generation of epistemic rationality norms, and I discussed costs and benefits of epistemically rational and irrational emotions. In section 6.1, I argued that epistemic rationality norms of emotions have been selected because they are conducive to the formation of epistemically correct emotions. In section 6.2, I presented an

example of an epistemic norm that applies to three stages of the emotional experience: the cognitive basis, the appraisal mechanism and the emotion. In section 6.3, drawing on personal stories in the medical sector, I provided a real-world example in which epistemically irrational emotions may be practically beneficial.

In section 6.4, I argued that epistemically rational emotions, such as the emotion of worry may deliver important epistemic costs that should not be overlooked when epistemically assessing the emotion. I showed that, in Jess's scenario, her epistemically rational worry compromises her epistemic functionality. I argued that this is mainly due to the attentional and motivational components of the emotion of worry. I made a case for the claim that rational worries may, in some contexts, be epistemically rational and yet and epistemically "guilty": they are guilty of compromising the subject's epistemic functioning in certain contexts.

Overall, this discussion showed that in certain contexts epistemically irrational emotions have the potential to deliver practical and epistemic benefits, and that epistemically rational emotions such as worry have the potential to deliver significant epistemic costs. I concluded suggesting that a complete epistemic assessment of emotions should take into account also contextual factors, such as the contribution of the emotion to the subject's epistemic functionality. I also suggested that in situations where irrational emotions deliver important epistemic benefits, and in situations where rational emotions are epistemically problematic, we are not obligated to follow epistemic norms.

CHAPTER 7: THE DIACHRONIC PENETRABILITY OF THE EMOTION-TRIGGERING SYSTEM

ABSTRACT

Sometimes we experience recalcitrant emotions, i.e. emotions that conflict with our beliefs. The persistence of recalcitrant emotions has been explained by the psychoevolutionary theory of emotions by reference to the informational encapsulation of the emotions' triggering system (Griffiths, 1997). A system is informationally encapsulated when its processes cannot be directly influenced by information flowing from other cognitive systems. However, this explanation is in tension with the possibility to overcome recalcitrant emotions. In this chapter, I argue that informational encapsulation of the emotion-triggering system is a more complex and subtler phenomenon than we realize. I propose a new way of understanding the persistence of recalcitrant emotions drawing on recent work on psychological taxonomy (Shea, 2014). My strategy consists in defending the diachronic penetrability of the emotion-triggering system.

INTRODUCTION

In the previous chapter, I discussed epistemic rationality norms. As we have seen, we often have emotions that are irrational. But are we epistemically responsible for experiencing emotions that flout epistemic rationality norms? Before addressing this question, that I will consider in chapter 8, I need to discuss whether we are able to control our emotions. As we will see in the following chapter, the kind of responsibility that we have for our emotions depends on the kind of control we enjoy.

Consider recalcitrant emotions. As we have seen in chapter 1, recalcitrant emotions are emotions that conflict with our beliefs. Imagine having had terrible nausea after having eaten rotten eggs three years ago. Today a friend offers you some scrambled eggs. This time you

believe that the eggs are fresh and not noxious, but despite this, you have a reaction of disgust. You try not to feel repelled, but your efforts are in vain: you cannot but feel disgusted at the scrambled eggs. Recalcitrant emotions like this are common and widespread. We are scared of objects that we know are not dangerous. We may feel jealous of our partner even if we know that they are not giving any undue attention to anybody else. And it is a common experience that emotions persist despite our better belief. We are all familiar with how difficult it is to change our emotions with reasoning: on many occasions, no matter how much effort we put into it, we fail to align emotions with our beliefs. But why do recalcitrant emotions persist?

Although emotional recalcitrance is a well-known phenomenon, the persistence of recalcitrant emotions is difficult to explain. The psychoevolutionary theory of emotion explains the persistence with the informational encapsulation of the system that triggers the emotions (Griffiths, 1997). A system is informationally encapsulated when it cannot be influenced by other systems (Fodor, 1983). The thesis of the informational encapsulation of the system that triggers emotions has then gained popularity[110] and has been endorsed by many philosophical accounts of emotions, e.g. (Goldie, 2000, Tappolet 2016, p. 21), and especially by the so-called *perceptual theories* of emotion,[111] e.g. (de Sousa, 1987; Tappolet 2000, Döring 2007). However, the existing view of the informational encapsulation of the system that triggers the emotion is overly simplistic and falls short of explaining in enough detail the possibility that

[110] Ronald de Sousa, before Paul Griffiths, suggested that emotions – and not the emotion-triggering system – are informationally encapsulated (de Sousa, 1987, pp. 152-156, p. 195). However, de Sousa did not offer a full-blown argument for the informational encapsulation of emotions.

[111] Perceptual theories of emotions conceive emotions as perceptions. For an overview of perceptual theories of emotions, see Brady (2013), and Salmela (2011).

recalcitrant emotions can be overcome *due to cognitive influences*.[112] Recalcitrant emotions exhibit a certain resistance but are not set in stone. Even if now you are not able to eat the eggs that your friend is offering to you, through time you may overcome your tendency to be disgusted by eggs. Crucially, this can be the result of your thinking that fresh eggs are not noxious. We seem to accept that indirect cognitive influences may affect our emotions.

The phenomenon of recalcitrant emotions creates a dilemma: on the one hand, recalcitrant emotions exhibit some resistance to our beliefs – they seem to be informationally encapsulated – but on the other hand, over time, they seem to be overcome by cognitive influence. We currently lack a proper account of the cognitive penetrability of the system that triggers the emotion that successfully addresses this dilemma. Understanding whether recalcitrant emotions can be overcome in virtue of cognitive influences, and whether the emotion-triggering system may be *cognitively penetrated* (Pylyshyn, 1980) is of the utmost importance for emotional regulation and clinical practice. If cognitive influences were not causally efficacious for overcoming recalcitrant emotions, then trying to convince ourselves – and others – not to feel a recalcitrant emotion, using merely cognitive means, would not be an effective strategy for changing what emotion is triggered.

We need a new account of the cognitive penetrability of the emotion-triggering system that is able to explain the phenomenology of recalcitrant emotions. The aim of this chapter is to present a novel account of the cognitive penetrability of the emotion-triggering system that can account for the persistence of recalcitrant emotions, as well as the possibility of overcoming them. First, I briefly review the psychoevolutionary theory of emotions. Then, I propose a

[112] Philosophical theories of emotions agree that we may be able to control to some extent our recalcitrant emotions, but regard control more as the product of exposure or habituation, rather than the result of cognitive influences. See, for instance (Goldie, 2000, p. 115; Helm 2001).

solution to the challenge from emotional learning using Nicholas Shea's taxonomy (Shea, 2014). Next, I discuss the argument for the informational encapsulation of such a system (Griffiths, 1997). I then present the distinction between synchronic and diachronic penetrability (Churchland, 1988), and show how this can illuminate the debate on the cognitive penetrability of the emotion-triggering system. Finally, based on this distinction, I offer a new explanation of the persistence of recalcitrant emotions that does justice to empirical data and folk conception about how emotions work.

7.1. The Psychoevolutionary Theory of Basic Emotions

In this chapter, I focus on the psychoevolutionary theory of emotions (DeLancey, 2004; Griffiths, 1997), since the informational encapsulation thesis has been originally presented within this framework. According to Paul Griffiths's psychoevolutionary theory (Griffiths, 1997, p. 78), there are six basic emotions: surprise, anger, fear, disgust, sadness and joy. These emotions are called *basic* in this context because they are evolutionary adaptations and have a specific pattern of bodily changes associated with them. The emotions that have been left out, like for instance piety, pride or love, are not considered *basic* because they are better accounted for as higher-order emotions rather than evolutionary adaptations. In this chapter, for the sake of the argument, I will assume the adequacy of the psychoevolutionary account of basic emotions.

According to such an account, emotions are evolutionary adaptations in the sense that they have been selected throughout the evolution of the species for bringing about behaviour that helps cope with the challenges posed by the environment to the organism. The methodology employed by the psychoevolutionary theory can be characterised as "reverse engineering". It consists in explaining the evolutionary history of a known trait, in this case the emotions' bodily changes, through its evolutionary past (Griffiths, 1996, pp. 519-521; Lewens, 2007).

For instance, bodily changes constitutive[113] of fear, e.g. eyes wide open, trembling and piloerection, originate in our evolutionary history as an adaptive response to danger, adaptive because fear was selected to promote a fight or flight response to dangerous situations. Similarly, bodily changes constitutive of disgust are an *adaptive* response to noxious stimuli, adaptive because disgust lowers the risk of ingesting infected food.

The psychoevolutionary approach has its roots in Charles Darwin's theory of evolution and in his work about the expression of emotions (Darwin, 1872). Darwin's studies influenced Paul Ekman's empirical research on the expression of emotions that has been incorporated into the psychoevolutionary theory (Ekman, 1972, 1977). Following Ekman, Griffiths argues in favour of the existence of so-called affect-programmes.[114] For Griffiths and other psychoevolutionary theories (DeLancey, 2004), affect-programmes are basic emotions[115] and each type of affect programme constitutes one type of basic emotion. When an affect-programme is activated, it displays complex, coordinated and automatic patterns of emotional responses. Affect-programmes are complex because they are composed of several features: facial and vocal changes, musculoskeletal response, changes of the autonomic nervous system and in the level of hormones. For instance, when the affect-programme of fear is activated, the heart rate increases, the body begins to sweat, etc. Those changes are coordinated because those responses occur in recognizable sequences and automatic because they do not need a conscious direction to unfold.

---

[113] Below I explain why bodily changes are constitutive and not the product of the basic emotions.

[114] Ekman owes the notion of affect programme to Tomkins (Tomkins, 1962). Griffiths's main argument for the existence of affect-programmes is an inference to the best explanation from the presence of emotional responses that depend upon an innate factor. For the related discussion see Griffiths (1990).

[115] This identification causes a conceptual revision: patterns of emotional responses are not caused by the emotion but *are* the emotion. See (Griffiths, 1997, p. 78).

According to the psychoevolutionary theory, affect-programmes can be triggered by two different systems, the high-level cognitive system and the automatic appraisal mechanism (henceforth, AAM). The high-level cognitive system is defined as the computational system "in which people use information of the sort they verbally assent to (traditional beliefs) and the goals they can be brought to recognize (traditional desires) to guide relatively long-term actions and to solve theoretical problems" (Griffiths, 1997, p. 92). When a subject starts thinking of possible dangerous outcomes of her action, her affect-programme of fear can be triggered by the high-level cognitive system as a result of her chain of thoughts. The AAM is the other system that triggers the emotions. It always operates independently from high-level cognitive processes. When it receives perceptual information, it evaluates rapidly and in an automatic way whether the information deserves an emotional response or not.

7.2. A Challenge from Emotional Learning

The notion of informational encapsulation was defended by Jerry Fodor, relative to input systems (Fodor, 1983). Fodor conceived the architecture of mind to have three components: transducers, input systems and central processors. Transducers are processes that have as an input the information that comes from the environment (*distal stimuli*) and as an output a representation (*proximal stimuli*). Proximal stimuli are then computed by the input system. Finally, the central processors have the function to compute the output of input systems to fixate beliefs.

According to Fodor, input systems are informationally encapsulated if and only if, given a set of inputs, processing within the input system cannot be *directly* causally affected by

information stored in other cognitive systems.[116] The property of informational encapsulation is considered the essential feature of modularity (Fodor, 1983, p. 71).[117] That means that if the notion of informational encapsulation fails, then a system would cease to be modular to an interesting extent. Griffiths argued that the AAM has some of the features of Fodor's input systems that characterize modularity (Griffiths, 1990, 1997).[118] In this chapter, I will concentrate on the claim that the AAM is informationally encapsulated.

Prior to a deeper analysis of the argument of the informational encapsulation of the AAM, I will offer a better account of what the AAM is. This is intended primarily to rule out uninteresting difficulties for the thesis that the AAM is informationally encapsulated. According to Griffiths, the AAM can learn new information about the kind of stimuli that merits an emotional response. The main motivation for introducing the learning capacity is to account for the fact that different people have different emotional reactions (or no reaction at all) towards the same stimulus, and that this difference cannot just be the product of an innate factor. In Griffiths's account, learning occurs through a conditioning process. Via conditioning,

---

[116] Indirect influences of the processing with the input system are compatible with informational encapsulation. Informational encapsulation is incompatible only with direct influences. If, for instance, I move my head towards you I am influencing my perceptual input system simply because I am changing the visual stimuli and consequently the input of the perceptual processing. Modifying the processing only in virtue of a change in the input of perceptual processing is not yet a direct cognitive influence. So, this possibility is consistent with the informational encapsulation of an input system. What needs to be excluded is a direct influence, namely an influence of a high-level cognitive state on the perceptual processing and not merely at the input of this processing.

[117] For a critical discussion of the relation between those properties see also Prinz (2006b).

[118] Other psychoevolutionary theorists argued for similar theses, e.g. Arne Öhman, Susan Mineka, Leda Cosmides and John Tooby assume that the "fear module", namely the process that triggers fear, is informationally encapsulated (Cosmides and Tooby, 1992; Öhman and Mineka, 2001).

stimuli with no special emotional significance can be paired through a conditioning process with another stimulus that elicits an emotional reaction. In this way, when the paired stimulus occurs, it elicits the emotional reaction as well. Now, following Ekman, he speculates that the AAM has a form of memory that stores this acquired information about classes of stimuli which merit an emotional response. Thanks to the memory of the AAM, the system can access information about the kind of stimuli that require activation of an emotional response (Griffiths, 1997, p. 92). The novel stimulus is thus compared with the previously assessed stimuli as requiring the emotional response. If it matches the previous ones, then the emotion activates. For instance, this model predicts that when Joe sees a gun, his AAM compares the visual stimuli of a gun to other gun-like memories. If gun-like features were previously assessed as dangerous, fear ensues.

The AAM explains our ability to learn which stimuli deserve an emotional response. However, this learning capacity creates an initial problem for the informational encapsulation of the AAM. If the AAM's evaluation relied on a memory which belongs to the high-level cognitive system, then the AAM would turn out to be cognitively penetrable. This is because we would have a causal influence of a previously stored memory, which belongs to the high-level cognitive system, on the working of the AAM. Consider the situation in which Mary believes that foreigners are dangerous, and this belief *b* is stored in her memory. If, by seeing a foreigner, the AAM triggered fear drawing on a piece of information, say *b*, that was stored in the high-level cognitive system, then Mary's AAM would not be informationally encapsulated. If the AAM's memory coincided with the memory of the high-level cognitive system, then the thesis of the informational encapsulation of the AAM would be trivially false. The defender of the informational encapsulation of the AAM needs to find a more appropriate explanation of emotional learning.

Griffiths does not recognise this as a serious problem for the informational encapsulation of the AAM and does not discuss this caveat. Drawing on Nicholas Shea's taxonomy (Shea, 2014), I will now offer a solution to this problem. A useful suggestion to tackle this issue is to assume that the AAM's memory does not coincide with the high-level cognitive system. This is because if the AAM's memory coincided with the high-level cognitive system, then the AAM would trivially be not informationally encapsulated. Conversely, if we assumed that the AAM's memory is different from the high-level cognitive system, the informational encapsulation of the AAM would not be trivially false. Let us assume, to examine a more interesting thesis, that the AAM's memory is not the same as the memory available to the high-level cognitive system. So, in this scenario, the AAM's memory would be stored within the AAM. If this were the case, then when it received an input, the AAM would compare it with previously stored information that does not come from the high-level cognitive system. In this case, we would still have a top-down influence on the AAM's process. However, as the background information is stored within the AAM, this case would not be cognitive penetration of the AAM.[119] Thus, the thesis of the informational encapsulation of the AAM could be maintained, at the cost of drawing a distinction between the memory of the AAM and the memory of the high-level cognitive system.

Shea's taxonomy (Shea, 2014) allows us to draw the distinction between the memory of the AAM and the memory of the high-level cognitive system. His taxonomy clarifies what is a top-down influence within the system, better characterising what it means to have a memory that is embodied in the AAM (Shea, 2014). The cornerstones of his taxonomy are the concepts of *implicit* and *explicit* mental representations. According to Shea, an *implicit representation* is defined as *"*a disposition to transition between two or more occurrent representations that can

[119] This is acknowledged by Fodor (1983) and Macpherson (2011).

have no influence on subsequent processing except via the representations between which the disposition subsists" (Shea, 2014, p. 81).[120] In other words, the information provided by implicit representations is embodied in the disposition to transition between two mental representations, and it is not represented anywhere else. An *explicit representation* is an "occurrent representation that is not implicit" (Shea, 2014, p. 81). Consider, for instance, some high-level vision processes, where previously stored information about the perceived object affects its depth assignments (Raichle, 1999). Previously stored information intervenes in our vision processes as cues, suggesting depth assignment. The information involved in depth assignment is implicit: it is a disposition to transition between certain cues e.g. curvature and depth assignments. Conversely, the representation that my blue pen is distinct from my black pen is explicit: it can act as input for many other psychological processes, such as remembering, recognition, imagination, etc.

Shea's framework can be applied to the discussion of the informational encapsulation of the AAM. Consider the case in which the AAM was disposed to make transitions from the stimulus of a snake to the presence of danger. This disposition implicitly encodes the information that if there is a snake, then there is a danger. However, this information would be an implicit representation. This is because it would influence subsequent processing only in virtue of causing the AAM's activation of fear, when the explicit representation of a present snake occurs. Conversely, the belief that snakes are dangerous is an explicit representation because it can act as an input to the processes of thinking, reasoning, planning, etc. A way of distinguishing top-down effects[121] that are embodied in the AAM from top-down effects that

[120] This implicit representation is a disposition.

[121] A top-down effect is, according to Nick Shea, the influence of previously stored information on a psychological process (Shea, 2014).

are outside the system is asking respectively (a) whether the causal impact of background information is implicit or (b) whether it is caused by the tokening of explicit representations outside the mechanism.[122]

The application of Shea's framework to the AAM's memory helps us to solve the challenge from emotional learning. This is because it gives a better framework to understand the nature of the memory embodied in the AAM, clarifying why it is not coincident with the memory in the high-level cognitive system. Shea's distinction between implicit and explicit representation elucidates exactly the kind of influence that defenders of the informational encapsulation of the AAM should exclude: an influence from an explicit representation on the implicit representations embodied in the AAM. Shea's taxonomy will be relevant to the discussion of the argument for the informational encapsulation of the AAM (sections 7.4-7.5).

7.3. The Argument for the Informational Encapsulation of the AAM

The main argument for the informational encapsulation of the AAM consists in: (i) presenting a plausible case in which a person has a recalcitrant emotion; and (ii) providing the best explanation of why recalcitrant emotion exists. Here I focus on Griffiths's argument for the informational encapsulation of the AAM (Griffiths, 1997).

In general, the occurrence of recalcitrant emotions is considered problematic because it is not clear why the emotional response *persists* against the judgement and the intention of the subject. Even if the subject puts great effort in controlling her emotional response, she cannot help but have the emotion. Theories of emotions have to be able to provide a compelling explanation of why recalcitrant emotions exist, but that is often a challenge, particularly for

[122] Note that Shea's taxonomy is not, by itself, incompatible with the possibility that implicit processing is modified by explicit representations or explicit thoughts.

cognitivist theories like *judgementalism*, which claim that emotions are evaluative judgments (Nussbaum, 2004; Solomon, 1988).[123] Consider a subject who has a recalcitrant fear of flying: she fears flying despite her belief that flying is safe. For judgementalists, fearing an object *o* requires also that the subject judges *o* as dangerous. Thus, to explain this recalcitrant fear, judgementalists seem to have to assume that the subject also judges that flying is not dangerous. However, this explanation is controversial. This is because when a subject has a recalcitrant emotion, she seems to lack the judgement that the object of her fear is dangerous.[124] Griffiths argued that the fact that recalcitrant emotions are possible is explained clearly by the informational encapsulation of the AAM. As he puts it:

> "The way in which certain kinds of emotional response, notably those dealt with by affect-program theory, can conflict with other cognitive processes would be neatly explained if there are processes that trigger affect-programs which are informationally encapsulated. In that case, there would not be a free flow of information between them and the rest of the mind. It is possible for a modular system to respond as if a certain state of affairs obtains although the organism as a whole believes that that state of affairs does not obtain." (Griffiths, 1990, p. 185)

Griffiths claims that, if the AAM were informationally encapsulated, then the AAM would trigger the emotion without being influenced by beliefs that are stored in the high-level

---

[123] Theories of emotions should also explain the alleged irrationality of recalcitrant emotions. This issue gave rise to a large debate, e.g. (Brady, 2011; Benbaji, 2013; Döring, 2015; Tappolet, 2016; Grzankowski, 2017; Naar, 2018). However, this chapter focusses only on the problem of the persistence of recalcitrant emotions.

[124] For a criticism of judgementalist's account of recalcitrant emotions, see, for instance, (D'Arms and Jacobson, 2003; Brady 2007). The judgementalist theorists of emotions have tried to modify the judgementalist account to explain recalcitrant emotions better, see, for instance (Roberts, 2003; Helm, 2001; Brady, 2009). However, the adequacy of these neojudgementalist explanations is still controversial.

cognitive system for evaluating the stimulus. So, the explanation of why it is possible for a subject to fear earthworms while thinking that they are actually safe is that the AAM activates the fear-affect-programme without taking into account the subject's beliefs about the safety of the situation. That would explain why recalcitrant emotions are possible without assuming that the subject has two contradictory beliefs. This seems in line with our everyday practice: we rarely think that a subject is falling into a contradiction when she has a recalcitrant emotion. Moreover, the explanation looks compelling since we know that, when we are in the grip of a recalcitrant emotion, e.g. a fear of a spider, we are not able to immediately alter our emotional response with our thinking. If we accept that this is the best explanation of these emotions that we currently have, it is possible to infer that the AAM is informationally encapsulated.

The argument in favour of the informational encapsulation of the AAM also supports a slightly different, but closely related thesis: the cognitive impenetrability of the AAM. The notion of cognitive impenetrability was introduced by Zenon W. Pylyshyn who claimed that a system is cognitively impenetrable if it cannot be altered by cognitive states. As he puts it, a system is cognitively impenetrable if it "cannot be influenced by such purely cognitive factors as goals, beliefs, inferences, tacit knowledge and so on" (Pylyshyn, 1980, p. 111). On the contrary, a system is cognitively penetrable if the system can be *directly* altered by cognitive states, or, in Pylyshyn's words, if "the function it computes is sensitive, in a semantically coherent way, to the organism's goals and beliefs, that is, it can be altered in a way that bears some logical relation to what the person knows" (Pylyshyn, 1999, p. 343).

The existence of recalcitrant emotions can be explained by the informational encapsulation of the AAM as much as by the cognitive impenetrability of the AAM. This is because, if the AAM were cognitively impenetrable, the subject's contrary belief would not be able to causally

affect the AAM.[125] Since the informational encapsulation of the AAM requires the cognitive impenetrability of the AAM but the cognitive impenetrability of the AAM does not imply its informational encapsulation, recalcitrant emotions could be more economically explained by abducing the cognitive impenetrability of the AAM rather than its informational encapsulation. It is, therefore, possible to formulate a better argument from recalcitrant emotions, as follows:

P1: There are cases of recalcitrant emotions.
P2: The best explanation of why it is possible to have recalcitrant emotions is that the AAM is cognitively impenetrable.

C: The AAM is cognitively impenetrable.

This argument is better than the previous one because it is more economical than the former: if the AAM were informationally encapsulated, then, necessarily, the AAM would be cognitively impenetrable. The reverse is not true.

7.4. Can We Overcome Recalcitrant Emotions?

If the AAM were cognitively impenetrable, then we would not be able to influence the basic emotions triggering system through our thinking. However, we know that recalcitrant emotions can be overcome. Consider yourself on a roller-coaster. You feel fear, even if you are secured to the coaster car by a resistant lap bar and safety belts, and you know that you are entirely out

[125] The notion of cognitive penetrability is less theoretically demanding because it merely implies that there cannot be any *direct causal influence of belief-like states* on the AAM, whereas the notion of informational encapsulation denies the possibility that *any state external to the system has a direct causal influence* on it (Robbins, 2015). A system could be cognitively impenetrable, for instance, but not informationally encapsulated if it were encapsulated only relative to beliefs but not to the causal influence of representations that come from different other systems, e.g. vision or auditory.

of danger. It is plausible to think that, through effort and learning, you could cease to be scared while being on the same roller-coaster. There is empirical evidence that cases of precisely this sort occur. In what follows, I focus on evidence involving people with a phobia because phobias are the instances of recalcitrant emotions that have been studied more thoroughly. Phobias are anxiety disorders that affect people who experience persistent and intense fear when confronted with stimuli that they can be brought to judge as not dangerous.

Some studies show that people who have arachnophobia, for instance, can overcome it by undergoing therapy, e.g. (Paquette et al., 2003; Hauner et al., 2012). The study by Vincent Paquette and colleagues shows that recalcitrant emotions can be overcome by cognitive behavioural therapy (CBT). CBT is a therapeutic technique, which aims to change the subject's beliefs and behaviour through a combination of cognitive and behavioural therapy (Öst and Clark, 2013). Cognitive therapy focusses mainly on changing what the subject thinks through discussion with the practitioner. During the conversation, the therapist asks questions to the subject, to help them better evaluate their situation and to help them to analyse their behaviour. Conversely, behavioural therapy (henceforth BT) consists in exposing the subject (in the laboratory or the real world) to the eliciting stimuli in order create a different association with it (Öst and Clark, 2013).

Paquette and colleagues' study was conducted with a group of twelve subjects that had phobia of spiders and on a control group of thirteen people without phobia. Their experiment shows that after CBT, all participants could touch a tarantula, without any detectable emotional reaction. Initially, all subjects were scanned with a functional magnetic resonance imaging technique (fMRI)[126] while viewing a film that depicted spiders and another one depicting butterflies, to compare the brain activity when viewing butterflies from the activity measured

[126] An fMRI is an imaging technique that measures the brain's activity by the changes in blood flow.

when seeing spiders. People with phobia displayed fear reactions while viewing the film, but people without it did not display any reaction. FMRI results show that the subject with phobia presented an activation of the right dorsolateral prefrontal cortex and a parahippocampal activation that was absent in people without phobia. The dorsolateral prefrontal cortex is a brain area normally associated with cognitive emotion regulation through reappraisal[127] (Golkar et al., 2012). Its activation was interpreted as the attempt of the subjects with phobia to use cognition to lower and control the emotional reaction. The parahippocampal cortex is usually associated with cognitive processes such as contextual associations and episodic memory (Aminoff, Kveraga, and Bar, 2013). The parahippocampal activation was considered to be an automatic activation of memories that generated spider phobia. During four consecutive weeks, participants with phobia met their therapist who used CBT to help them in overcoming the phobia.

In these sessions, subjects were initially exposed to exercise books containing pictures of spiders, then to film about living spiders, finally to a real spider which they were asked to touch. The exposure was conducted using "guided mastery"[128] for correcting misbeliefs about spiders. Paquette and colleagues considered CBT effective. This is because, during the fourth session, all subjects were able to touch a real spider without reporting a fear reaction and after the fourth session, the brain activation pattern of subjects with phobia was similar to the one noted in normal control subjects, without frontal or hippocampal activity.

[127] In chapter 8, I discuss reappraisal as a strategy for regulating one's own emotion.

[128] Guided mastery is a psychotherapeutic technique that helps the patient to develop coping skills. This objective is reached by using a modelling technique to change faulty beliefs and by enhancing the perceived self-efficacy to cope with the fearful object (Bandura, 1997).

Specific data about which beliefs were corrected or induced in patients during the therapy are not available because fMRI studies on CBT for treating spider phobia[129] do not offer a detailed description of how cognitive therapy was conducted. However, knowing that practitioners used CBT, we can infer what beliefs are likely to be corrected and how. Cognitive behavioural practitioners challenge patients' dysfunctional cognitions regarding, for instance, their ability to cope with the emotion employing a therapeutic technique, which is called *cognitive restructuring* (Leahy and A., 2012). Cognitive restructuring helps the patient to identify dysfunctional thinking. The patient then tries to correct false beliefs and assumptions. For instance, the therapist can ask subjects to evaluate their beliefs, providing a reason in favour of and against dysfunctional beliefs, defining their terms, identifying the evidence they have for or against a given belief and reflecting with them about how these beliefs can be tested. To accomplish this objective, practitioners can use different techniques, like asking patients to engage in role-playing games, or to write a daily record of their dysfunctional thoughts, to reflect on the evidence they have.[130]

This empirical evidence is not in conflict with P1 because it does not deny that recalcitrant emotions occur, but it denies only that they cannot be overcome. However, it may conflict with P2 depending on the reason why recalcitrant emotions cease to be persistent. On the one hand, if conscious or unconscious cognitive states do not have a causal influence on the system, then the AAM could still be considered cognitively impenetrable, but it remains to be explained why the AAM ceases to activate the emotional response with purely non-cognitive influences.

---

[129] For fMRI studies about CBT therapy with people with spider phobia, see (Schienle, Wabnegger, and Scharmüller, 2014; Schienle, Schäfer, Stark, and Vaitl, 2009; Straube, Glauer, Dilger, Mentzel, and Miltner, 2006; Paquette et al., 2003).

[130] More details on this practice can be found in Leahy (2012).

On the other hand, if beliefs directly causally affect the AAM's process, then this may be incompatible with the thesis of the cognitive impenetrability of the AAM.

Let us assume that the AAM is cognitively impenetrable. Under this assumption, guided mastery and cognitive restructuring are not causally efficacious in changing the AAM's disposition to activate the emotional response. However, we can overcome recalcitrant emotions. Thus, the possibility to overcome recalcitrant emotions needs to be explained by a new association within the AAM that gradually lowers the fear response to the phobogenic stimulus, or by a modulation of an existing association. If this new association is just a new disposition of the system to move from one representation to another, it does not necessarily follow that background information outside of the AAM is tokened to causally affect the processing of the phobogenic stimulus. To rephrase by using Shea's terminology, overcoming a recalcitrant emotion would require only an implicit representation. The defender of the cognitive impenetrability of the AAM can explain the overcoming of recalcitrant emotions only appealing to a change in the implicit representations that maintain the emotion.

However, this explanation is not fully convincing because it clashes with our everyday experience of regulating emotions. We do believe that changing what we think has some effect on the emotions we feel. If I am on a flight and I fear flying despite my better belief that flying is safe, I may remember that I have reached all my previous flight destinations safely, or I may think that flying is safer than travelling by other means of transport, or I may focus on how amazing it is to see the Alps from the sky. It is a common experience that these types of cognitive changes affect to some degree our recalcitrant emotions via indirect strategies. The cognitive impenetrability thesis seems to conflict with our common experience of emotional regulation.

Moreover, the thesis of the cognitive impenetrability of the AAM looks revisionary also when applied to clinical practice, suggesting that rational thinking is not directly causally

efficacious in overcoming the phobia. Interestingly, according to this explanation, all the verbal parts of the CBT treatment seem redundant. The discussion with the practitioners during cognitive restructuring and in the setting-up of the task and during and after the exposure seems not causally efficacious for overcoming the phobia.

A defender of the cognitive impenetrability thesis may argue, at this point, that the efficacy of CBT can be reduced to the efficacy of BT. However, the thesis that BT does not work cognitively has been recently considered a naïve hypothesis (Hofmann, 2008; LeDoux, 2015, p. 377). This is because verbal discussions are not only a prerogative of CBT. BT is performed with the aid of verbal instructions as well. In BT, verbal instructions may regard the setting-up of the task. The subject discusses with the practitioner their expectations about the phobogenic stimulus, and subjects receiving BT are also encouraged to develop different cognitive states about the phobogenic stimulus after their exposure (Hofmann, 2008). Moreover, verbal instructions are not necessary for cognitive influences. Subjects may change their beliefs about the phobogenic stimulus, or about the possibility to control the emotional response and predictions of future events even if they are not told to do so.

Again, someone who supports the cognitive impenetrability of the emotion-triggering system may argue that verbal instructions and cognitive states occurring during BT are not causally efficacious. However, not only does this position seem too extreme, but it also seems in tension with recent research regarding the influence on the subject's beliefs for the success of the therapy (Hofmann, 2004, 2008; Mattick, Peters, and Clarke, 1989). Richard P. Mattick and colleagues argue that cognitive restructuring is causally efficacious for overcoming social

phobia. They conducted an experiment with forty-three people with social phobia[131] (Mattick et al., 1989). They divided participants into four groups, assigned respectively to (1) exposure therapy, (2) cognitive restructuring without exposure, (3) exposure combined with cognitive restructuring and a (4) waiting-list control group. Participants received six sessions of therapy, except those in the control group that did not receive any therapy. People in the first group were exposed to the phobogenic stimuli, and no method for cognitive restructuring was mentioned. Conversely, people in the second group received no exposure, but they were treated to bring about belief-change by using cognitive restructuring.[132] As they describe, cognitive restructuring was done by firstly suggesting that the phobia was affected by false beliefs such as "concern about the opinions of others, (b) the feelings that others were watching, and (c) the belief that others could see signs of anxiety" (Mattick et al., 1989, p. 6). Secondly, participants were asked to identify and change their irrational thoughts, analysing past or hypothetical situations related to the phobia. Also, participants in the third group received cognitive restructuring but they combined it with exposure. This enabled the third group to practice cognitive restructuring on their exposure situation, instead of hypothetical ones. Results show that treated subjects improved significantly in comparison with the waiting list group. Cognitive restructuring paired with exposure performed better than exposure therapy alone, and cognitive restructuring alone was better than exposure alone at the follow-up measures.

[131] Social phobia is an anxiety disorder that occurs when the subject fears social interactions. Participants involved in the study had a phobia of eating, drinking, writing or singing in public, and being in public places (Mattick et al 1989). For further detail regarding social phobia, see Valente (2002).

[132] The procedure of the therapy is explained in Goldfried (1980) and Ellis (1962).

These results suggest that cognitive restructuring is causally efficacious [133] and that, consequently, P2 is false.

Cognitions are not only relevant for overcoming social phobia, but also for overcoming other types of recalcitrant emotions. For instance, lower expectations about future harm due to the phobogenic stimulus are related to better therapy outcomes and endorsing beliefs about the possibility to change emotions have been linked with more successful emotional regulation (Gutentag, 2017).[134] In general, cognitive states seem to have an important influence in the overcoming of recalcitrant emotions.

However, supporters of the cognitive impenetrability of the AAM may argue that this empirical evidence does not challenge the cognitive impenetrability thesis. They may insist that the cognitive penetrability of the AAM would require a direct tokening of an explicit representation in the AAM process. We could distinguish two possible ways according to which cognition may be causally efficacious for overcoming recalcitrant emotions: one is *indirect* and the other one *direct*. The cognitive influence would be indirect if cognitive restructuring alters the AAM processes by activating a new implicit association between the phobogenic situation and the absence of danger, or by weakening the connection between the relevant stimulus and the presence of danger. For instance, consider a subject with social phobia who thinks that getting onto a bus is relaxing. This thought might contribute to the

[133] This thesis is, to some extent, controversial. See (Feske and Chambless, 1995; Gil et al, 2001). However, other studies suggest that cognitive restructuring is superior to exposure in social phobia (see, for instance, Hofmann, 2004; Longmore and Worrell, 2006).

[134] Tony Gutentag and colleagues asked participants to rate their beliefs about the controllability of their emotional reaction. Then, participants were asked to watch a movie and to reduce any emotion of disgust felt in response to it. This experiment showed that people who believed that emotions were controllable were more successful in regulating their disgust.

formation of a new association between being in a crowded bus and feeling calm. However, this influence would not imply a tokening of an explicit representation in the AAM because the association might only be a disposition to the transition between bus-like features and a relaxing situation, which is not able to have a causal impact on other processes. More importantly, this would not be a case of cognitive penetrability of the AAM because the system would not receive any influence from outside the system. The explicit representation would merely have an indirect influence. Similarly, if the subject imagines getting onto a bus and not facing any dangerous situation, this imagination might induce fear extinction in her because she weakens the implicit association that she has between getting onto a bus and being in danger.

Conversely, we would have a case of a direct causal influence if one of the subject's cognitive states directly causally affected the AAM processes. The defender of the cognitive impenetrability of the AAM may argue that this empirical evidence only shows an *indirect* rather than a direct causal influence on the AAM.

It is difficult to draw a clear distinction between (1) a cognitive therapy that is causally efficacious in virtue of an indirect influence of explicit representations, due to the creation of a new implicit associations in the AAM and (2) a cognitive therapy that is causally efficacious because it receives not only indirect cognitive influences, but also direct influences of explicit representations. The possibility of direct cognitive influences on the AAM seems plausible: it can account for the efficacy of cognitive restructuring, and it fits with evidence showing that cognitive states can be causally efficacious for emotional regulation. However, the competing hypothesis, according to which recalcitrant emotions are never overcome by direct explicit representations but are only overcome by indirect explicit and implicit representations, cannot be easily ruled out.

There is currently a gap in the empirical literature: no empirical study investigates specifically whether the system that triggers the emotion may be influenced by *direct* cognitive influences. There is a need to design new experiments that are able to discriminate between hypotheses (1) and (2). And even if the impossibility of direct influences of cognitive representations on the AAM were granted to the defender of the cognitive impenetrability of the AAM, still, the cognitive impenetrability thesis would overlook the importance of indirect *cognitive* influences. Even if influences were only indirect, these influences are causally efficacious and the thesis of the cognitive penetrability of the AAM does not take them correctly into account. Therefore, the debate needs to be framed differently. We need to be able to do justice to the efficacy of cognition to regulate recalcitrant emotions, even if this regulation is only indirect, while at the same time retaining the intuition that recalcitrant emotions display some sort of resistance to cognitive influences.

7.5. Is the AAM Cognitively Impenetrable?

The idea that there may be causal influences of beliefs in the AAM's process is in tension with the thesis of the cognitive impenetrability of the AAM[135] because cognition may be directly causally efficacious in overcoming the phobia. And even if it were only *indirectly* causally efficacious, the cognitive impenetrability thesis significantly overlooks the causal role of such influences. Therefore, an alternative explanation of the persistence of recalcitrant emotion is needed.

Prior to presenting this alternative explanation P2, I will begin by considering that the above-mentioned empirical evidence does not show an *immediate* causal influence of beliefs on the process of triggering emotions. That is because subjects do not cease to have recalcitrant

[135] Note that those empirical findings are also in tension with the thesis of the cognitive encapsulation of the AAM.

emotions as soon as they acquire the contrasting belief, but they need, in general, repeated encounters with the emotion-eliciting stimulus. In clinical practice, subjects with phobia need to undergo a process of therapy that requires several sessions to be effective. So, there is a sense according to which the AAM is still cognitively impenetrable, but only for limited time.

This distinction between these two concepts was initially introduced by Paul Churchland in a philosophical discussion with Fodor about cognitive penetration in perception (Churchland, 1988). Churchland tried to undermine Fodor's view on cognitive impenetrability of perception, by focusing on evidence of perceptual learning. With perceptual learning, perception develops different capacities, and Churchland claimed that some of them would involve cases of cognitive penetration. Instead of discussing his rather complex argument, I will focus on his distinction made in the literature about perceptual vision between synchronic and diachronic cognitive penetration. In general, a perceptual process is synchronically penetrated if a belief penetrates the processes quickly and very easily while a perceptual process is diachronically penetrated if its perceptual processing is reconfigured through time.

I propose the introduction of a reference to time in the discussion of the AAM's possibility to overcome a recalcitrant emotion, saying that synchronic cognitive penetration of the AAM occurs when, given a certain input, the AAM *instantly* ceases to trigger the affect-programme because of a causal influence of a cognitive state.[136] On the other hand, diachronic penetration of the AAM occurs if the AAM ceases to trigger the affect-programme after a period of training *because of a causal influence of a cognitive state*.[137]

---

[136] Tappolet (2016) suggested that emotions are diachronically penetrable. She may also believe that the emotion-triggering system can be cognitively penetrable, although she does not explicitly argue for this thesis and she does not develop this point further.

[137] This influence may be direct or indirect.

The difference between these two kinds of cognitive penetration is not clear-cut because there is no agreed criterion about where to draw the line. In any case, the learning process of overcoming emotions could count as a clear case of diachronic cognitive penetration, since the AAM ceases to activate the fear response after some sessions of cognitive-behavioural therapy (Paquette et al., 2003), and, as I argued in the previous section, these cognitive influences may be causally efficacious.

Once this distinction is drawn, the empirical finding of Paquette and colleagues is incompatible with the claim that the AAM is diachronically impenetrable, but it is not incompatible with the claim that the AAM is synchronically impenetrable. Thus, P2 is not compelling. It is possible to find a better explanation of recalcitrant emotions, that is the following:

> P2* The best explanation of why recalcitrant emotions are possible is that the AAM is synchronically cognitively impenetrable.

This new explanation could provide a convincing explanation of why recalcitrant emotions occur. According to P2* those emotions happen because, when certain associations regarding the dangerousness of a certain stimulus are built into the AAM, this association cannot be eliminated instantaneously by the opposite judgment. However, this does not mean that the system must be in general cognitively impenetrable or informationally encapsulated. An instance of that is the possibility of overcoming a recalcitrant emotion thanks to, among other things, a cognitive influence. P2* is a better explanation than P2 because it is less theoretically demanding, and moreover it is consistent with the empirical data suggesting a causal influence of a belief in the process that triggers the emotion. The argument for the cognitive impenetrability of the AAM is not compelling because its second premise is not convincing. Our common practice and the empirical data are more easily explained by the diachronic penetrability of the AAM. P2* is a better explanation of recalcitrant emotions because it is less

theoretically demanding and better fits with the empirical evidence. Therefore, we should adopt a new view of the cognitive penetrability of emotions, according to which emotions are synchronically impenetrable, but diachronically penetrable.

Accepting the claim that the AAM is diachronically cognitively penetrable requires also re-thinking the concept of modularity of the AAM. That is because if the notion of informational encapsulation fails, then it also fails the essential feature of modularity. One possibility would be outlining a different notion of quasi-modularity along the lines that Jesse Prinz proposes (Prinz, 2006a). Recalcitrant emotions can be overcome, by a diverse set of cognitive influences. Even if all these influences were indirect and, therefore, no synchronic cognitive penetrability were put in place, the cognitive influences on the system would still be diachronically penetrable.

My account does not entail that recalcitrant emotions can always be overcome by CBT. The above-mentioned evidence only shows that it is possible that recalcitrant emotions can in principle be overcome by CBT. However, it does not establish that this is *always* the case. Consider the hypothesis that there are some emotion elicitors that are innate and suppose the loss of support is an innate elicitor of fear. Now, consider a situation in which subjects lose balance and experience fear while believing themselves to be safe (for instance while being in an amusement park attraction). It might be the case that these subjects undertake CBT to eliminate their recalcitrant emotion, and yet they cannot stop the activation of fear when they lose balance.[138] That could be due to the association being so deeply built into their AAM that it is almost impossible to change; thus, they can just learn how to lower the emotional reaction

[138] Mineka and Öhman argued that we exhibit greater difficulty to overcome fears triggered by stimuli related to our evolutionary history rather than fears about non-evolutionary related stimuli. See Mineka and Öhman (2002).

of fear. Even if that situation happens, my account will not be undermined because it does not entail that all recalcitrant emotions can in principle be overcome.

One of the virtues of my revised version of the argument for the cognitive impenetrability of the automatic appraisal mechanism is that it is fully compatible with the psychoevolutionary framework. The diachronic penetrability of the AAM is compatible with the possibility of receiving data that contrasts with our beliefs because the system remains synchronically impenetrable. The possibility that the AAM is diachronically penetrable, an option that P2* leaves open, would be adaptive. Imagine a situation in which you have a system that triggers fear when it faces a certain stimulus, even if you think that this stimulus is not dangerous. This fear response makes your life difficult because it arises against your will, and this hinders you; moreover, imagine that your conscious or unconscious belief cannot do anything to change that response. The only thing you can do is try to change your disposition to feel fear *indirectly*, for instance breathing slowly or undergoing a counterconditioning therapy because your high-level cognitive system cannot have a causal impact on those processes. And imagine another situation in which you also can learn how to change your disposition to feel fear by using your belief, in addition to other indirect ways. The latter situation would be better than the former because it would allow you to reduce more quickly the waste of energy involved in having repeated cases of recalcitrant fear. Therefore, the possibility that cognition has a diachronic causal impact on the process of the AAM is still consistent with the evolutionary framework.

7.6. Summary and Conclusion

In this chapter, I argued that a new account of cognitive penetrability of the emotion-triggering system is required to explain the persistence of recalcitrant emotions, and the possibility to overcome them. Psychoevolutionary theories of emotions explain this persistence with the informational encapsulation of the AAM. Although the informational encapsulation of the AAM successfully explains why beliefs are not immediately able to halt recalcitrant emotions,

the AAM's informational encapsulation thesis also rules out *a priori* the possibility of cognitive influences on the AAM. I showed that this is problematic because it is in tension with our everyday practice of emotional regulation, as well as with empirical evidence suggesting that cognitive states can affect the AAM. A new account of the cognitive penetrability of the AAM is required to explain both the persistence of recalcitrant emotions in spite of the subject's contrary belief, as well as the possibility to overcome recalcitrant emotions using reasoning strategies. Using the distinction between synchronic and diachronic cognitive penetrability, I defended the synchronic cognitive impenetrability of the AAM and the diachronic cognitive penetrability of the AAM. I argued that my proposed explanation is able to explain the persistence of recalcitrant emotions and it coheres with our everyday understanding of how emotion regulation works.

This chapter has shown that informational encapsulation and cognitive penetrability are subtler than we think. This new understanding of cognitive penetrability of the emotion-triggering system has the potential to influence other philosophical discussions too. For instance, it could inform the way in which we think about responsibility and control over our emotions, since it supports the possibility to diachronically control them by using rational strategies. I will develop this issue in the following chapter. Moreover, my account could shape the debate on the nature of the emotion, making us reassess (some) of the arguments for the perceptual account. For instance, the informational encapsulation of the emotion-triggering system has been used to support the *perceptual account* of emotions, which maintains that emotions are perceptions of values (see, for instance, Tappolet 2016; Prinz, 2004 and de Sousa, 1987). My account may be used to challenge the argument from informational encapsulation

for the perceptual account.[139] Finally, inspired by my new account, there could be new experimental research aimed at understanding the nature of emotions and this may contribute to distinguishing between two theoretical explanations of the efficacy of CT and CBT. This may be the first step to developing new therapies for people with anxiety disorders, claiming, for instance, the importance of diachronic cognitive influences for overcoming phobias.

[139] My argument may be used to challenge what Salmela called the "perceptual system argument": an argument for the perceptual theory of emotion (Salmela, 2011, p. 3).

CHAPTER 8: EMOTIONS AND EPISTEMIC RESPONSIBILITY

ABSTRACT

Are we epistemically responsible for our emotions? Both recent work in epistemology, e.g. (Moser, 2002; Steup, 2000) and on emotions (Brun, Doğuoğlu, and Kuenzle, 2008) has not given enough attention to this important question. We currently lack an account of the epistemic responsibility for our emotions that thoroughly addresses the interplay between types of epistemic norms and control we have over our emotions. In this chapter, I aim to present such an account. I argue that the type of epistemic responsibility for our emotions we have depends on the kind of control we have over our emotions. Drawing on the extant discussion on the normativity of beliefs (such as Alston, 1988; Feldman, 2000; Olson, 2015) I distinguish three types of control over emotional states: synchronic, diachronic and fine-tuning control. I also distinguish three different types of epistemic norms, and I present a new account of how the epistemic responsibility for our emotions relates to different types of epistemic norms and control.

INTRODUCTION

Consider again, Shakespeare's tragedy *King Lear* (Shakespeare, 1987 [1608]). Recall that at the beginning of the play, King Lear asks his daughters to express their love for him. Instead of falsely overstating her affection, as the other sisters just did, Cordelia answers sincerely that her love does not need to be supported by flattering words. Listening to this answer, King Lear angrily disclaims his parental care from Cordelia. In chapter 4, I have argued that King Lear's anger is epistemically rationally assessable. In chapter 6, I have argued that an emotion is epistemically irrational when it fails to be formed in accordance with the evidence. Given that there is nothing extremely offensive in Cordelia's speech, King Lear's anger is epistemically

irrational[140] because it is failing to be felt in accordance with his evidence. But is King Lear *epistemically responsible* for failing to form his emotion in accordance with the epistemic rationality norms? And are we, like King Lear, responsible for having emotions that infringe norms of epistemic rationality?[141]

The answer to this question depends at least in part on what type of control we have over our emotions. Control is something that we actively do, and emotions, in contrast, are often regarded as *passive* mental states (see, for instance Gordon, 1986; Helm, 2001). However, we seem to enjoy some degree of control over our emotions in the long term: we can diachronically control our emotions through training and habituation (Goldie, 2000; Helm, 2001), *via* a change in our evaluations and "educating" them (D'Arms, 2013). And in chapter 7, I have argued that we are able to diachronically control our emotions using rational strategies.

We currently lack an account of the epistemic responsibility for our emotions that thoroughly addresses the interplay between types of epistemic norms and control we have over our emotions. In this chapter, I attempt to fill this gap, developing a fully fleshed account of epistemic responsibility for our emotions. In a nutshell, I argue that we are epistemically responsible for an emotion *e* if, and only if, given an epistemic rationality norm that applies to an emotion *e* of a certain type *n*, we enjoy the type of control over *e* that is relevant to the type of epistemic norm.

---

[140] In this chapter, I focus mainly on epistemic rationality norms of emotions, which are norms that focus on the mental state's responsiveness to the evidence. See my chapter 1, and Bortolotti (2010) for an overview of different rationality norms. Epistemic norms of emotions have been put forward by Patricia Greenspan (2004), Michael Brady (2013), Deonna and Teroni (2012), Carolyn Price (2015), George Pitcher (1965), Gabriele Taylor (Taylor, 1975).

[141] I defend an account of the epistemic rationality of emotions in chapter 4.

I structure the chapter as follows. In section 8.1, I present some background issues to pave the way for my positive view of the epistemic responsibility for our emotions. In section 8.2, building upon the extant literature on epistemic responsibility for beliefs, I distinguish between three different types of control that subjects have over the emotions: direct, indirect and fine-tuning control. In section 8.3, I distinguish three different types of epistemic norms that apply to the emotions: synchronic, diachronic and structural norms and I present and defend my own account of epistemic responsibility for our emotions, relating different types of epistemic norm with each different type of control.

8.1. William Clifford and Epistemic Responsibility

Towards the end of the 19th century, William K. Clifford argued that we have responsibility for our beliefs (Clifford, 2011). He defended this point with the example of a shipowner that repeatedly ignores his evidence about the seaworthiness of his ship up to the point that he is fully convinced that the ship is solid and safe for the passengers. Being now certain about the ship's resilience, the shipowner refuses to refit the ship. However, shortly after the ship's departure, the vessel sinks. Clifford argues that the shipowner of this example is *epistemically responsible* for his belief that the ship was seaworthy. According to him, the shipowner is blameworthy because he breaks an epistemic norm: the norm that a belief ought not to be formed (or maintained) without sufficient evidence for it (Clifford, 2011, p. 186). Since the shipowner endorses his belief without having sufficient evidence for it, he is blameworthy. His blameworthiness does not result from the actual consequence of the action undertaken in light of this belief, e.g. allowing the ship to sail, nor from the truth or falsity of the belief that the ship was seaworthy. According to Clifford, even if the ship had not sunk, the shipowner would have still been blameworthy for his belief (Clifford, 2011). This is because an agent is responsible for the process according to which she acquires and maintains a belief, and this does not depend on the truth or falsity of the belief, but on the evidence-gathering and weighing

up process. In the case described by Clifford, the evidence available to the shipowner for his belief was not sufficient. Therefore, for Clifford, the shipowner was blameworthy.[142]

The idea that having some control over a certain mental state is required for being responsible for that mental state has been assumed by many accounts of the epistemic responsibility for beliefs (Alston, 1988; Audi, 2008; Feldman, 2000; Ginet, 2001; Weatherson, 2008).[143] This conforms to the ordinary practice of considering control a *sine qua non* condition for responsibility (Fischer and Ravizza, 1998). Clifford's scenario nicely illustrates such a view. The claim that the shipowner is responsible for his irrational belief is grounded in the shipowner's ability to control his belief that the ship is seaworthy. There is a sense according to which the shipowner controls his belief since he could have formed his belief otherwise: the shipowner could have paid attention to his doubts about the seaworthiness of the ship, and he could have gathered better evidence against his belief. In this scenario, the shipowner's responsibility for his belief depends at least in part on his ability to control its belief-formation process, regulating his reasons for having such a mental state. If the shipowner did not have any control over his belief that the ship was seaworthy, then he would have lacked

---

[142] Clifford considered blameworthiness to be coextensive with responsibility. However, in this chapter, I will consider blameworthiness to be conceptually distinct from responsibility, as has been extensively shown by Hanna Pickard, e.g. (Pickard, 2013, 2017). In a nutshell, Hanna Pickard argues that blameworthiness is conceptually distinct from responsibility because you can hold an agent responsible for an action, even if the agent is not blameworthy because she has excuses that lift the blame. In this chapter, I will not give a full account of the conditions according to which an agent is blameworthy for experiencing an emotion that infringes an epistemic norm. Such an enterprise falls beyond the scope of this chapter and I will leave this to future work.

[143] The validity of this principle has been challenged by Sharon Ryan (Ryan, 2003). Her strategy consists in considering four different versions of the claim that "ought" implies "can" and to show that they do not work because for each of them we can construct a scenario where the agent lacks control about an action but nevertheless she is responsible for the action. However, we can always revise the OIC principle to account for these situations.

responsibility for his belief about the seaworthiness of the ship.[144] In the next section, I distinguish three different types of control over the emotion that are plausible candidates for the epistemic responsibility of emotions.

8.2. CAN WE CONTROL OUR EMOTIONS?

Drawing on the literature on doxastic control over beliefs, I introduce different types of control over emotions and I discuss their relevance for the epistemic responsibility for our emotions. It is possible to distinguish between three types of control over beliefs: (1) direct doxastic voluntarism, e.g. (Alston, 1988; Ginet, 2001; Weatherson, 2008); (2) indirect doxastic voluntarism, e.g. (Feldman, 2000); and fine-tuning control, e.g. (Olson, 2015). In what follows I illustrate these types of control and I present an analogous distinction between the types of control over emotions.

8.2.1. Direct Control

Direct doxastic voluntarism is the thesis that we can enjoy direct control over at least some classes of beliefs. Direct control consists in the ability to believe *b*, or cease to believe *b,* as a result of immediately executing an intention to do so (Alston, 1988). This type of control is modelled on the kind of control an agent enjoys over actions. Suppose that I want to control the movement of my finger, to press the door-phone of my friend's apartment. I perform this action by immediately executing my intention to move the finger in a certain way as to put pressure on the door-phone button. In this example, my action of pressing the button follows

[144] One might suppose that even in the absence of control, we might still be epistemically responsible for putting ourselves in a situation where we know that we are likely to infringe an epistemic norm. This is an interesting idea. However, I am not interested in exploring this option here. I will only focus on the epistemic responsibility for our emotion.

immediately the execution of my intention to do so. Crucially, I do not need to perform another action to press the button. Direct doxastic voluntarism understands at least some instances of believing as a particular type of action that we can immediately perform simply by executing our intention to do so.

We can present a thesis concerning emotions which is analogous to direct doxastic voluntarism. I call this thesis direct emotional voluntarism: the thesis that we enjoy direct immediate control of (at least some types) of emotions. Emotional voluntarism requires that we control our emotions as if we had in our possession a switch that allows us to switch on and off the emotion. Although the intensity of beliefs is not widely discussed, the intensity of emotion is an important feature of the emotional experience, as it has been argued that the emotion's intensity represents the intensity according to which the object of the emotion has the evaluative property that figures in the content of the emotion (Nussbaum, 2001). So, direct control may then also apply to the intensity of the emotion, and it may also require that we are able to control our emotion as if we had in our possession a dimmer switch that could increase and decrease the emotional intensity. More precisely, we enjoy direct control over emotions if and only if the following condition holds:

> Direct control$_{emotion}$: An agent *S* can experience (or regulate, or not experience) the emotion *e* as a result of immediately executing her intention to experience (or regulate the emotion's intensity or not experience) *e*.

In general, we seem to lack the ability to immediately and directly experience (or not to experience) a particular emotion simply by executing our intention to do so. Suppose that Susan wants to feel proud of her ability as a PhD researcher in maths. She knows that she is a good mathematician. After all, she has received much positive feedback from her supervisors, from other professors and colleagues about her work and she has published five papers in well-respected journals. However, she does not feel pride for her accomplishments. No matter how

much she wants to feel proud of her accomplishments, she seems to lack the ability to immediately and directly feel proud for her academic successes. Like her, we also seem to enjoy very little ability to control the intensity of our emotions.[145] The idea that emotions are states about which we enjoy a very limited direct control is grounded in the observation that emotions, like perceptions, exhibit a sort of resistance. It is difficult to generate or suppress them at will. This characteristic of emotions is the source of the general understanding of emotions as passions, that is, as mental states that happen to us (Helm, 2009).

Exceptions to this trend are Jean Paul Sartre and Robert C. Solomon, who challenged the thesis that emotions are passive mental states and argued for the controversial thesis that emotions are – very often – like actions (Sartre, 1962; Solomon, 2007). According to scholars such as these there is a sense according to which we *choose* our emotions, even unconsciously. We could interpret this as an argument for the ubiquity of the possibility to enjoy direct control over our emotions. Although Solomon does not deny that our choice to have an emotion may be constrained by the circumstances, he insists that experiencing an emotion is still a choice and we are the "co-author" of our emotions (Solomon, 2007, p. 192). This claim is defended by presenting a scenario in which a subject who is at the bar with his friend is insulted by a passer-by. This scenario is supposed to show that experiencing an emotion is a choice. As he puts it:

> I feel myself getting tense (…), but now, do I just get angry? I have a number of choices, though none of them very pleasant. I can walk away, and shoulder what abuse and ridicule may follow. I can answer him in kind, an unwise choice given his size and

[145] Suppose that a subject enjoys direct control over the intensity of the emotion but does not enjoy direct control over the experience of the emotion. If this is so, she lacks full direct control over her emotion, but enjoys *some degree* of direct control.

evident disposition. I can pretend to ignore him. I can really try to ignore him, for instance, by engaging my friend in a lively discussion of university politics. I can get angry at my friend for bringing me to such place. (Solomon, 2007, p. 192)

Solomon presents a situation in which the agent chooses whether to *act* on the basis of the emotion. Solomon's scenario may support the thesis that the choices we make when we experience an emotion – and prior to the emotional experience – may then affect future emotional episodes, but it does not support the thesis that we have direct control over emotions. However, the claim that the choices we make when we experience an emotion affect future emotional episodes is not sufficient to support the possibility – and even less the ubiquity – of direct emotional voluntarism. Direct emotional voluntarism requires that the agent chooses *the emotion* and not merely the actions following the emotion. Therefore, the situation presented by Solomon does not fully support the claim that we often directly control our emotions.[146]

Research on emotion regulation has shown that emotions cannot be regulated directly by volitional effort (Kappas, 2008). Moreover, controlling a mental state is not symmetric to controlling of an action. Pamela Hieronymi observes that we can do an action by deciding to do it because such action is (practically) good, but we cannot believe *x* by deciding to believe *x* for entirely practical reasons (Hieronymi, 2008). Similarly, emotions cannot be held simply because we want to have them, or because they are good. We do not seem to enjoy the ability to experience an emotion simply as a result of immediately executing our intention to do so: it seems that we need to be able to present the right kind of consideration to the emotion system

[146] Solomon's argument may support the possibility of using indirect and long-term strategies to control the emotion. Perhaps *this* is the sense according to which emotions are like actions. I will discuss indirect and long-term strategies to control our emotions in the following two subsections.

first. We seem to lack direct control, therefore the ubiquity of direct emotional voluntarism is false. Note that direct emotional voluntarism is not rejected here on a conceptual basis, but only by empirical observations. However, from the observation that something is never done, we cannot infer that it is not possible to do it (Weatherson, 2008). Even if in some exceptional cases it would be possible to enjoy direct emotional voluntarism, this phenomenon would be rare, and consequently, this will not be a particularly interesting candidate for the kind of control which grounds our epistemic responsibility for our emotions.

8.2.2. Indirect Control

Indirect doxastic voluntarism is the thesis that we enjoy indirect doxastic control over at least some classes of beliefs. We enjoy indirect doxastic control of a belief *b* when the agent intends to form the belief *b* and the agent successfully forms the belief *b* via performing (an)other action(s). This type of control is modelled on the type of control that we enjoy about certain actions, like opening doors or writing a particular word on a piece of paper. We can perform these actions only if we do (an)other action(s), e.g. pulling the door and moving the hand, holding the pen in such a way that we write the desired word. According to Alston, an agent enjoys indirect doxastic control over her belief *b* when she believes *b* as a result of her intention of believing *b* and undertaking a series of actions *over time* (Alston, 1988). This type of control is modelled on the kind of control we enjoy about characteristics of our body, such as our fitness level, or our lung capacity. For instance, in normal circumstances, by exercising regularly, we can keep the body fit and by regularly training with a spirometer we can improve, to some extent, our lung capacity. Similarly, we enjoy indirect doxastic control over our belief when we repeat a certain action over time with the aim of obtaining a certain belief. For instance, suppose that somebody wants to believe that God exists. She does a series of actions, e.g. going to mass, hanging out with theists, avoiding atheists with the intention to form the belief that God exists. And suppose that, as a result of this, she really believes that God exists.

This would be a case of indirect control over the belief that God exists. Analogously, we can call indirect emotional voluntarism the thesis that we enjoy indirect control over at least some type of emotions. We can define indirect emotional control as follows:

Indirect control$_{emotion}$: An agent *S* can experience (or regulate, or not experience) the emotion *e* as a result of executing their intention to experience (or regulate, or not experience) *e* by undertaking (a) certain action(s).

For instance, suppose that Jane is really bold and does not experience the fear of walking on the surface of a frozen pond even if she has evidence that walking there is extremely dangerous. Jane enjoys indirect control over her fear of walking on the frozen pond if and only if she can regulate the intensity of her fear as a result of undertaking an action or a series of actions, e.g. directing her attention to signs about the danger of the frozen pond, thinking about people who fall into the icy water.

Situation selection may also be a type of indirect control: suppose that I intend to believe that I am wearing a red coat. If I believe that I am dressed in a red coat as a result of my action of wearing a red coat, I enjoy indirect control of such belief. This is an odd-type of control, and perhaps it is questionable to consider it a genuine example of indirect control over beliefs. This is because the control over the belief, in this case, is merely the result of the agent's modification of the world, that then changes the basis for the belief (and not the processes that, given a certain input, deliver the belief). However, in the emotion case, there is an interesting parallel of this regulatory strategy of emotions: situation selection, which consists in regulating the emotion by choosing to avoid or look for a situation that we expect to have (or lack) emotionally salient input (Gross, 2015).

There are plenty of indirect strategies that help us to suppress or generate emotions, or to reduce or increase their intensity. Psychologists used the umbrella term "emotional regulation" to refer to a number of strategies that can be undertaken to control the emotions, both in their

activation and in their intensity. Emotional regulation is defined as the "goal directed processes functioning to influence the intensity, duration and type of emotion experienced" (Gyurak, Gross, and Ektin, 2011, p. 401). It applies to all the stages of an emotional experience. Emotional regulation can be in place (i) when we select the emotionally relevant stimulus; (ii) when we decide to be in a situation in which we expect to respond with a certain type of desirable emotional response; (iii) when we refuse to be in a situation where we expect to experience an undesirable emotion;[147] (iv) when we direct our attention away from or object;[148] (v) when we influence our appraisal by cognitive change or "reappraisal"; (vi) when we modulate the behavioural and expressive responses after the emotion has been triggered (Gross, 2015). It can also occur when we modify our facial feedback to generate or suppress a particular emotion (Levenson et al., 1990).

Indirect control requires us to have the intention to regulate a particular occurrence of emotion *e*. This type of control may be of two kinds: the generation of an emotion that has not been triggered yet, and the regulation of an emotion that has already been triggered. We may regulate our emotion by undertaking the following actions: (a) situation selection, (b) attentional deployment, (c) cognitive change, or reappraisal, and (d) modulating behavioural and bodily changes. Consider an agent who is indifferent to rationalist architecture but wants to learn to admire it. Suppose that she has evidence that rationalist architecture is admirable: she knows that it is widely appreciated by her favourite art critics and that it has a great reputation among world-leading architects. She may try to generate the emotion of admiration

[147] Emotional regulation can be in place also when we put ourselves in a situation where we expect to experience an emotion.

[148] Emotional regulation can be in place also when we direct our attention on a particular object, attempting to arouse an emotion.

by (a) selecting a situation where there is rationalist architecture, (b) directing her attention to signs of the architect's ability to design rationalist buildings and to aspects of the building that she believes to be beautiful, (c) thinking about the exceptional value of rationalist architecture in comparison with other forms of architecture, and (d) trying to display the facial and bodily expression typical of the emotion of admiration. In some cases, when our emotion is resistant to generation, we may need to repeat these actions over a long period of time. One afternoon spent trying to learn to admire Giuseppe Terragni's rationalist architecture, for instance, may not be enough to succeed in admiring it. The same strategy may be at play when we regulate an existing occurrence of an emotion *e*. We often do emotional regulation, and not just to fulfil epistemic norms about emotions, but also for practical or moral reasons, when we may enjoy in "emotional camouflage", that is, when we try to hide our emotion (Ekman, 1992). We do this, for instance, when we try to hide tension while speaking in public, when we mask sadness when we meet a person we love who we know is going to die or when we try to halt our anger at somebody. To sum up, we seem to be able to indirectly regulate our emotion by using situation selection, attentional deployment, cognitive change, and response modulation.

8.2.3. Fine-Tuning Control

In this section, I develop a third type of control over emotions called "fine-tuning", drawing on Dustin Olson's work on epistemic agency (Olson, 2015). He suggests that we can wilfully modify how we form beliefs by changing our belief-forming disposition. As Olson puts it:

> "Epistemic agency denotes the motivation and ability to refine and alter one's belief-forming methods and subsequent belief-forming practices – these methods and practices can collectively be understood as a doxastic disposition of belief-forming abilities, which can also be characterised as one's propensity to form true or false or coarse-grained or fine-grained beliefs within different domains" (Olson, 2015, p. 450).

Fine-tuning control does not require having the intention to form a specific belief, but it requires having the intention to modify the belief-forming *disposition* (Olson, 2015, p. 452). The main difference between indirect-doxastic voluntarism and fine-tuning control is that fine-tuning control applies to the belief-forming disposition, and not to the formation of a specific belief *occurrence*. We enjoy this type of control over beliefs when we use cognitive abilities, e.g. attention, reflection and deliberation, to improve our disposition to form beliefs. Olson illustrates this with the following example:

> "Consider Campbell's desire to form accurate beliefs about scotch so as to be able to converse with other aficionados. Currently, Campbell's beliefs are not fine-grained concerning the different qualities he experiences when drinking scotch. His desire to form these more fine-grained beliefs motivates him to undertake certain acts that will refine his palate – e.g. he reads the taster's notes for each scotch he drinks, familiarizes himself with the jargon associated with the different scotch-drinking steps, and is phenomenologically sensitive to the experiences he has while drinking scotch. Eventually Campbell's beliefs concerning the different scotches he samples become fine-grained. Thus, rather than simply forming coarse-grained beliefs, like merely having the ability to distinguish between a scotch whisky and a rye whisky, Campbell can now form more fine-grained beliefs, such as the age, region, and casking processes of the different scotches he samples" (Olson, 2015, p. 453).

As we see from this example, Campbell does not want to acquire a particular belief, nor does he employ an indirect strategy to form a particular belief about whisky. What he wants to gain is the ability to form fine-grained beliefs about a certain domain, in this case, about scotch and other whiskies. He is similar to a musician, who practices not only to perform one particular piece but because she wants to improve her general ability as a performer, e.g., a cellist who wants to practice refining the quality of the sound, to avoid scratching the sound when playing loud, and to improve the agility of the hands and bow control.

The notion of fine-tuning control for emotions can be modelled on Olson's notion of epistemic agency. Emotions are activated by appraisals that are often stable dispositions. In many cases, it is not only the occurrent emotion that is misguided but also the disposition of responding with an emotional reaction to a type of stimulus. Consider a person with a phobia of spiders. She may be able to regulate her fear of spiders in one particular case, by altering the intensity of her fear. This may be done by employing emotional regulation, as described in the previous section. But when another spider comes, the fear of that spider occurs again. We might consider her emotion as irrational *over time*. This is because her fear seems grounded in a defective disposition. Fine-tuning control is not directed to regulate or generate a particular emotion occurrence, but it aims to generate *stable patterns* of generation or suppression in a given domain. In this case, fine-tuning control can be used to regulate the fear response to situations related to spiders. Fine-tuning control is not only used to overcome misguided emotional dispositions, as in the case of a person with a phobia. It can also be used to create better emotional dispositions. Fine-tuning control occurs when emotional regulative strategies repeated over time target one or more *emotional dispositions* and not a particular occurrence of emotion. We can define fine-tuning control over emotions as follows:

Fine-tuning$_{\text{emotion}}$: An agent *S* possesses fine-tuning control of her emotion in a certain domain *D* when she is able to control her emotion-formation practices in *D*.

We seem to be able to fine-tune our emotions to an interesting extent. Philosophers and psychologists suggest that we enjoy the ability to acquire new emotional dispositions. Ronald de Sousa, for instance, argues that during our development we learn the typical scenarios in which each emotion is appropriate, and we learn to experience the emotion accordingly (de Sousa, 1987, p. 184). In cognitive psychology, it is a platitude that we enjoy the ability to learn

new associations between stimuli and an emotional response.[149] This platitude is the result of the observation that we are not innately endowed with all emotional dispositions, but we acquire them during our development (Thompson, 1991). Adults teach children to control the emotions they undergo by guiding them when they are not successful. Through development, children acquire new associations, as well as the ability to regulate emotions. Even if some dispositions exhibit learning preparedness, e.g. the acquisition of the disposition to be scared by spiders or the acquisition of the disposition to be scared of heights (Mineka and Öhman, 2002; Öman and Mineka 2001), we are born with only a small amount of associations (if any) and we acquire most of them. Psychotherapy shows that through therapy like cognitive behavioural therapy (Jokić-Begić, 2010) or exposure therapy (Hauner, Mineka, Voss, and Paller, 2012) we can overcome a previous emotional disposition to transition between a certain situation to certain emotional responses, e.g. (Paquette et al., 2003). This empirical research shows that we can have emotional learning. However, to make a case for fine-tuning control, we need to ensure that we enjoy the possibility to control the way in which we form a certain emotion in a particular domain, and that this is not the mere product of automatic processes. If emotional regulation were entirely automatic, with no possibility to control the emotion-formation process using our reasoning abilities, then we would not enjoy this type of control. However, we can use regulative strategies by executing our intention to do so. Many regulative strategies involve features that are already under the agent's control, e.g. deliberative, attentional and reasoning capacities. These regulative strategies may be used to fine-tune our emotional system. Regulative strategies, when repeated over time, can become automatic (Gyurak et al., 2011). Suppose, for instance, that an agent wants to reduce her fear of talking in public. When she is about to talk in public, she starts to control her breath and posture to

[149] See also Tappolet (2016c) for a discussion about the plasticity of the emotional system.

reduce the intensity of the fear. She breathes diaphragmatically, stands still and maintains an assertive posture, making eye-contact with the audience. After giving some talks, she starts to breathe automatically with her diaphragm. The strategy in this case becomes automatic. However, the subject still enjoys control of the adopted strategy: the regulative strategy has been initiated by executing her intention to do so.

Consider the following example: Joanne aspires to be an influential politician. She wants to train her ability to feel anger in the political context because she needs to detect any unkindness and insult addressed to her or to a member of her party. To do this, she observes how more experienced members of her party react to comments they receive, and she imitates their behaviour. She uses them as a model to shape her emotional dispositions. She may also employ Solomon's "as if" strategy to experience an emotion, which consists in acting as if one has a certain emotion, to experience the emotion as a result (Solomon, 2007, p. 198). For instance, when Joanne realises that there is an offence, she may try to feel the appropriate way by acting as if she was angry. If Joanne realises that she is overdoing it, she may use strategies to decrease the intensity of her anger. When she feels angry at something for no reason, she may try to re-describe her situation to regulate her anger. After a while, she acquires an excellent ability to detect offence that helps her in her political activity. Joanne may train both her ability to feel anger in the right circumstances, and to the right extent, and also her ability to experience other emotions such as hope and fear in the context of her political activism. Thus, fine-tuning control of emotions may involve multi-factorial emotional expertise, that are not limited to a single type of emotion.

8.2.4. Intention to Control

When displaying emotional regulation, agents have an intention to control their emotion, or their disposition to experience emotions, as the examples considered in the previous section show. In particular, given that the focus of this chapter is *epistemic* responsibility, we need to

qualify the nature of the intention to regulate the emotion. When an agent regulates an emotion, she requires an ideal that sets the target of the regulation (D'Arms, 2013). To regulate an emotion in the epistemic sense, agents need to have the intention to regulate their emotion using epistemic rationality norms as a regulative ideal.

In some cases, agents are aware that there is something epistemically faulty with their emotions and may form an intention to control their own faulty emotion. For instance, a person with arachnophobia knows that there is something epistemically faulty with her fear of spiders and may want to overcome her fear so as to align it with her better belief that spiders are safe. In other situations, agents may lack a specific intention to control a particular emotion occurrence, but they may try to improve their emotion-forming disposition, adopting fine-tuning control as the example of Joanne shows. However, in other situations, such as in King Lear's scenario, agents seem to lack the intention to control their emotions (either because of ignorance of undergoing faulty emotions, or because of stubbornness). Without having the intention to control one's emotional state, the agent may fail to actually control her emotion using direct and indirect control. Similarly, without having the intention to control one's disposition to experience emotion, we may actually fail to fine-tune our emotions. However, this does not imply that the agent lacks the ability to (directly/indirectly or fine-tune) control her emotion. What establishes whether an agent has the ability to control her emotional experience, in the epistemic sense, is whether she is able to control it, under the condition that she has the intention to control it using epistemic rationality norms as a regulative ideal.

8.3. Towards a Diachronic Notion of Epistemic Responsibility

Consider a subject like King Lear, who is experiencing a certain emotion. What types of epistemic rationality norms may apply to their emotion? Here I argue that we can distinguish

three types of epistemic rationality norms[150] that may apply to emotions: synchronic, diachronic and structural epistemic norms.[151] I then defend a pluralist and graded account of the epistemic responsibility for emotions. In a nutshell, I argue that depending on the type of epistemic norm we are considering, a different type of control is relevant, and a different conclusion about responsibility follows.[152]

Synchronic epistemic norms are norms that concern the activation, suppression, or the regulation of the intensity of a given occurrence of emotion *e* in an immediate, or very short period. More generally, synchronic norms of emotions have the following schematic form: at time *t*, you ought (or are permitted[153]) to experience *e* (or not to experience/ regulate *e*), under conditions *C*.[154] Consider, for instance, Patricia Greenspan's account of epistemic rationality: according to her, an emotion is epistemically rational if and only if the evaluation that represents the content of the emotion is warranted by the subject's relevant evidence

[150] In what follows, when I will talk about epistemic norms, I will only refer to epistemic *rationality* norms.

[151] I do not have space here to argue for a specific account of the epistemic rationality norms that apply to the emotions. Here I only argue that there may be three different types of epistemic norms that apply to the emotions.

[152] One possible way to generate epistemic norms is to think about what norms a subject would need to follow, if she wanted to be likely to experience *correct* emotions. For an account of the correctness conditions of emotions, see, for instance (D'Arms and Jacobson, 2000b). Epistemic rationality norms may apply to the emotion's intensity and to the emotion's type. Epistemic rationality norms may apply to a particular emotion in the longer term. Epistemic norms may also apply to the way the emotions are generated.

[153] I am aware that epistemic rationality norms that feature a "permissibility" operator may be thought to be more plausible than norms that feature an "ought" operator. This is because somebody may argue that if we were to fulfil epistemic norms, our mind would be cluttered with emotions. Giving a full explanation of the epistemic norms of emotions falls beyond the scope of this chapter. For the sake of this chapter, I do not need to commit to any particular account of epistemic norms for emotions.

[154] This norm might be anchored to a specific situation.

(Greenspan, 2004, pp. 128-129). We could formulate synchronic epistemic norms inspired by her account as follows: we might consider the epistemic norm that an agent is not permitted to experience anger when the evaluation that represents the content of the emotion is not warranted by the subject's relevant evidence.[155] This epistemic norm can be given a synchronic interpretation, according to which the norm ought to be immediately satisfied.

There are two types of diachronic norms: norms that focus on the occurrence of an emotion, that I will call *diachronic epistemic norms*, and norms that focus on our disposition to generate our emotions, that I will call *structural epistemic norms*. I call *diachronic epistemic norms* the norms that apply to the emotion *e* over an extended period. These norms require the subject to feel (or not to feel) a particular emotion *e* over a certain time; for instance, an agent may be subject to the obligation to experience happiness for having obtained an academic job, which she deeply craved. This obligation may be diachronic: she ought to take the steps to experience happiness for her success, over an extended amount of time. Indirect epistemic norms have the following structure: during an interval of time Δt,[156] an agent *S* epistemically ought (or is permitted) to take steps to experience, or regulate or not experience the emotion *e*.

Structural epistemic norms do not apply to a particular occurrence of an emotion *e*, but the way in which we generate our emotions. In general, structural epistemic diachronic norms have the following structure: during an interval of time Δt an agent *S* epistemically ought (or is permitted) to take steps to epistemically improve her emotion-formation practices by doing ψ,

---

[155] Here I use this norm as an illustration of a synchronic epistemic norm. In this chapter, I do not intend to offer a defence of such a norm.

[156] The specific amount of time available for an agent to regulate her emotion before she has flouted the norm may depend on the epistemic norm considered, and the context in which the agent is in. The time might also depend on the duration of the emotional experience. A full account of this issue falls beyond the scope of this chapter.

in context *C*. Structural norms are built on the nature of that to which the emotion is supposed to be sensitive. Given that, in this case, I focus on epistemic norms, structural norms are concerned with responsiveness to the evidence.

Clifford argued that it is "wrong always, everywhere and for anyone, to believe anything upon insufficient evidence" (Clifford, 2011, p. 186). This norm may be interpreted as targeting the evidence-gathering process for a specific belief. The evidence-gathering process is not instantaneous but requires the agent to take steps in advance to ensure that she forms her beliefs upon sufficient evidence, and this procedure requires time. Clifford's norm could be understood as a *structural diachronic* norm, as follows: in context *C* during the time interval Δt an agent *S* epistemically[157] ought not to believe that *p* without gathering sufficient evidence for *p*. An example of a similar diachronic norm of emotion is the following: during the time Δt, in context *C* an agent *S* ought not to experience the emotion *e* without acquiring sufficient evidence for the evaluation giving rise to *e*. Structural diachronic norms may prescribe that the subject forms different epistemic habits. For instance, an agent may be required to improve their anger-formation practices by ensuring that the object of the emotion is related to the emotion's cause, by checking consistency between anger and her extant beliefs about the negative value of the object, by seeking further evidence for the alleged offensiveness of the object.

As I argued in section 8.2, agents may enjoy different types of control over their emotions, and control may come in degrees. Since control over emotions may come in degrees, so also the responsibility for emotions requires a graded notion of responsibility. Graded accounts of

[157] It is controversial whether Clifford intended the "ought" to be epistemic. Here I do not aim to interpret Clifford's theory. I merely aim to show an example of a structural diachronic norm, which is inspired by Clifford's argument.

responsibility have been defended by Alex Madva (Madva, 2018). Madva argues against a binary notion of responsibility for implicit bias. He presents empirical evidence showing that agents may enjoy different degrees of control. Although it is often challenging to control one's own implicit bias, it is not impossible to do so. And if control is graded, his argument follows, it would be mysterious why we should adopt a binary notion of responsibility. Similarly, I argue that control over emotions comes in degrees. We may not be able to cease to feel angry with someone instantaneously, but we can undertake certain steps to reduce the anger's intensity. Someone may not be able to immediately change her disposition to be intensely angered by a crying child, but she may be able to take steps to reduce the intensity of that emotion, or she may take actions to change her disposition to feel anger in response to crying children. She may also develop a disposition to ensure that the emotion of anger she feels at crying children is experienced in accordance with her evidence on that subject.

In what follows, I distinguish between three types of responsibility for emotions: synchronic, diachronic and structural responsibility, as follows:

Synchronic responsibility condition: At time t, the agent S is synchronically responsible for fulfilling the synchronic epistemic norm n about her emotion e if and only if S enjoys direct or indirect control over her emotion e in such a way as to be able to satisfy the synchronic epistemic norm n.

Diachronic responsibility condition: During the time interval Δt, an agent S is diachronically responsible for fulfilling the diachronic epistemic norm n if and only if S enjoys indirect control over her emotion e in such a way as to be able to satisfy the diachronic epistemic norm n.

Structural responsibility condition: During the time interval Δt, an agent S is structurally responsible for fulfilling the structural epistemic norm n if and only if S enjoys fine-tuning control of her disposition to trigger emotions in such a way as to be able to satisfy the structural norm n.

In the previous section, I argued that we almost always lack direct control over our emotions. It follows that agents almost always lack the ability to fulfil the synchronic responsibility condition in virtue of their ability to directly control their emotions.

The diachronic responsibility condition is fulfilled more often than the synchronic one. This is because, as suggested in section 8.2.2, we often enjoy indirect control in the long term. Consider, for instance, Michael Brady's account of the irrationality of recalcitrant emotions (Brady, 2013; 2009), according to which a recalcitrant emotion is epistemically irrational when it wastes cognitive resources inclining a subject to believe what they have already rejected (Brady, 2009, p. 413).[158] We could use his account to formulate the following diachronic norm: during the time interval Δt, an agent S, who has already established the belief b, should not experience a recalcitrant emotion e that inclines S to accept b, when this emotion is wasting their cognitive resources. Consider Elena, who believes that Lila is not offending her. Despite this, Elena is angry at Lila and engages in continuous rumination about the alleged offensiveness of Lila's behaviour. According to the diachronic epistemic norm inspired by Brady's account, Elena is infringing an epistemic norm. She is diachronically responsible for infringing this norm insofar as she is able to not experience the anger at Lila over time. As in the previous case, during the time interval dictated by the norm, Elena might lack the ability to

[158] Giving a full presentation of Brady's account of the irrationality of recalcitrant emotions falls beyond the scope of the chapter. Here, I am using his theory only as an illustration for a diachronic norm.

halt her anger completely. However, she might be able to take steps to not experience her anger. If this is the case, she will fulfil the responsibility condition only partially. She will enjoy a limited degree of diachronic responsibility insomuch as she is able to partially satisfy her diachronic norm.[159]

The structural responsibility condition also seems to be possible to fulfil. As I have argued in section 8.3, there are many strategies available to control the dispositions to have an emotional experience, or different emotional experiences in a certain domain. Consider the following structural norm inspired by Clifford's account: in the political context, during the time interval an agent *S* ought not to experience the emotion *e* without gathering sufficient evidence for *e*.[160] As discussed in section 8.3, agents enjoy fine-tuning control. They have many strategies available that can help them to epistemically train their disposition to emotion. This norm, in particular, requires that the agent fine-tunes her emotion-triggering disposition to gather sufficient evidence for the emotion *e* and to align her disposition to trigger emotion in accordance with the evidence. As in the previous cases, the agent may lack the ability to fully fulfil this structural norm in the period of time required. However, the subject often enjoys the ability to take steps to improve her disposition to trigger emotions as required by the above-mentioned norm. For instance, she might improve her disposition by exercising emotional regulation to align her disposition to the evidence, or by gathering evidence for the emotion, e.g. by looking at the type of emotions experienced by others. As long as agents are able to take steps to fulfil the structural epistemic norm, they enjoy some non-trivial degree of

[159] The fact that somebody is able to satisfy the epistemic norm does not guarantee that they will actually satisfy the norm.

[160] It is worth stressing here that I do not want to argue for this epistemic norm. I am merely using this norm for illustrating the structural responsibility condition.

responsibility proportional to the extent to which they can take steps to fulfil the structural epistemic norm.

It might be the case that some emotional reactions to certain objects are so deeply rooted that they are almost impossible to change, even in the long term. This may be the case for the innate disposition to experience fear when we lose support (Morgan, 2003, p. 4), or when we hear a strong and unexpected noise, e.g. (Ekman et al., 1985). In these cases, agents will lack epistemic responsibility for having this emotion. Similarly, there may be cases where the regulatory capacity of our emotion-triggering system is impaired. In these situations, agents are not able to control their emotion to an interesting extent. It follows that in those cases, agents are not responsible for the emotions they have. Under the assumption that the agent lacks any sort of control whatsoever over this emotion, they have no kind of responsibility for it.

8.4. Summary and Conclusion

This chapter offered a novel account of epistemic responsibility for the emotions. I argued that epistemic responsibility for our emotions depends on the type of epistemic rationality norm we consider, and on the type of control we have over that particular emotion occurrence. I suggested that we are almost never responsible for infringing synchronic norms, but we can be responsible for infringing diachronic and structural norms.[161] Using a similar strategy, it might be possible to distinguish between synchronic, diachronic and structural norms in other domains (e.g. moral/ practical/ epistemic/ procedural/ aesthetic norms). Moreover, depending

[161] This holds under the assumption that your emotion can be subject to a norm when you are not able to control that emotion to fulfil the norm. I do not have the space to argue for such an account here. But what matters for this chapter, is that *if* your emotion is subject to that norm, and you lack synchronic control, then you are not synchronically responsible for it.

on which kind of control for the emotion we enjoy, we might be able to give an account of the responsibility for our emotions in other domains too.

We can finally examine whether King Lear is epistemically responsible for his emotion. Suppose that King Lear's anger ought to be experienced in accordance with the evidence. This norm, in the synchronic sense, is not fulfilled: King Lear's emotion is too intense, and the intensity of his emotion is not experienced in accordance with his evidence. However, Lear is not synchronically responsible for failing to comply with this synchronic norm because the level of direct control he enjoys over his emotion is limited. He is epistemically responsible for having the disposition to form his emotion of anger over time. He could have paid attention to evidence suggesting that Cordelia does love him, and he could have paid attention to the Earl of Kent's advice. He could have realised that he has epistemic reasons to feel less intense anger and he could have regulated his emotion indirectly over time to fulfil this diachronic norm. Therefore, he is diachronically responsible for his anger to a significant degree. Moreover, King Lear is also epistemically responsible for how he fine-tunes his emotion. His evidence-gathering process for his emotion is ignoring the significance of important pieces of evidence. He is failing to respect the structural norm that requires him to pay attention to the content of Cordelia's speech and to past instances of Cordelia's loving behaviour. Therefore, we can conclude that King Lear is also structurally responsible to a significant degree.

CONCLUSION

1. SUMMARY

In this section, I offer a brief summary of this thesis and in the following section, I indicate future research directions.

Overall, this thesis has shown that emotions and epistemic rationality are more intertwined than we previously thought. I have argued that most emotions are epistemically rationally assessable, and I have considered several issues arising from the possibility to epistemically rationally assessing the emotions.

Firstly, I have examined the aetiology of emotion. I focussed on the relationship between the evaluative content of emotion, and the content of the causal evaluation. I made a case for the claim that the evaluative content of emotion and the content of the causal evaluation can differ. I showed that the evaluative content of emotion could bring about an element of originality that is not inherited from the evaluation that triggers the emotion. I then considered the claim that emotions are always triggered by causal evaluations, and I offered a defence of this claim from the challenge from emotional contagion.

In the central part of the thesis, I examined the epistemic rationality of emotions. I presented my account of epistemic rational assessability of mental states, and I argued that most emotion occurrences meet these conditions. I then turned to identifying the epistemic rationality norms of emotions, showing that emotions and objectivity are not in opposition: I presented an account of evaluative properties that can be used to objectively establish whether an emotion is correct. Building on such an account, I offered a sketch of how epistemic norms should be generated and I argued that a complete epistemic evaluation of emotions needs to take into account not only the epistemic rationality status of the emotion, but also important contextual factors, such as the impact of the emotion on the subject's epistemic functionality.

In the final part of the thesis, I supported the claim that we can control and shape our emotions up to an interesting extent. I maintained that, although we mostly lack the ability to control our emotions in the short term, in the longer term we can use indirect and, in some cases, direct strategies to shape our emotional experiences. I then examined the implications this has for the epistemic responsibility we have for our emotions, defending a picture of emotions according to which we are often diachronically epistemically responsible for how we shape our emotions.

2. Implications for Future Research

The claims developed in this thesis open up new avenues of research in philosophy and in related disciplines, such as psychology and computer science.

I have put considerable emphasis on the relevance of contextual factors for a complete assessment of the epistemic status of emotions. Focusing mainly on the emotion of fear, I have considered real-life cases in which epistemically irrational emotions may deliver important epistemic benefits. Similarly, focusing on the emotion of worry, I have argued that epistemically rational emotions may, in some cases, be epistemically problematic. Future research could investigate whether other types of emotions exhibit the same pattern. Perhaps this will involve the interplay between motivational and attentional factors involved in emotional experiences. Future research may investigate whether we could be excused for having epistemically irrational emotions that deliver important epistemic or practical benefits.

The discussion of the control we enjoy for our emotions opens up new questions regarding the type of norms that regulate our emotions. In this thesis, I emphasised the need to distinguish between synchronic, diachronic, and structural epistemic norms. Further work could discuss a specific proposal of norms of epistemic rationality with a similar structure, and a similar distinction may be individuated in other emotion types.

My account of the epistemic rational assessability of mental states could be used to evaluate the epistemic rational assessability of different emotions. Moreover, it could be used not only to evaluate whether an emotion occurrence is epistemically rationally assessable, but also for evaluating whether occurrences of other types of mental states are so. Moreover, my claims that emotions are open to epistemic rational assessment could inform new fascinating projects at the intersection between philosophy and computer science. Recent developments in affective computing have brought to light the possibility that robots experience "robot-like emotions".[162] If there is some plausibility to the claim that robots experience emotions, new research could use this account of the epistemic rational assessability of emotions to discuss whether robots' emotions are epistemically rationally assessable. Epistemic norms of emotions could then be used as a regulative ideal not only by humans but also by tech developers for creating robots able to train their emotions in a way that is conducive to improving the epistemic rationality status of their emotions.

Considering the emotion-triggering system as cognitively impenetrable prevents us from appreciating that emotions may be causally influenced by our thoughts. In chapter 7, I argued that we need to employ a new framework for understanding whether the emotion-triggering system could be cognitively penetrable. Further research can be done to unveil the extent to which cognitive influence could directly affect the emotion-triggering system.

The arguments developed here support the claim that irrational emotions are neither necessarily pathological, nor rare phenomena. Once this claim is accepted, then new questions arise, connected with the possibility to train and develop our emotions. Future research may establish theoretical and educational advantages of shaping our emotions following epistemic rationality norms as suggested in the thesis.

---

[162] See, for instance, Mark Coeckelbergh (2010).

REFERENCES

Abramson, K. (2010). A sentimentalist's defense of contempt, shame and disdain. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion*. Oxford: Oxford University Press. 189-214

Alston, W. (1988). The deontological conception of epistemic justification. *Philosophical Perspectives, 2*, 257.

Alston, W. P. (1969). Feelings. *The Philosophical Review*. 78(1). 3-34.

Aminoff, E. M., Kveraga, K., and Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences, 17*(8), 379-390.

American Psychiatric Association, (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Arlington, Va.: American Psychiatric Association.

Arango-Muñoz, S., and Michaelian, K. (2014). Epistemic feelings, epistemic emotions: review and introduction to the focus section. *Philosophical Inquiries, 2*(1), 97-117.

Audi, R. (2008). The ethics of belief: doxastic self- control and intellectual virtue. *Synthese, 161*(3), 403-418.

Bandura, A. (1997). Clinical functioning. In A. Bandura (Ed.), *Self-efficacy: the exercise of control*. New York: W. H. Freeman & Company. 319-368.

Barrett, F. S., Robins, R. W., Wildschut, T., Sedikides, C., and Janata, P. (2010). Music-evoked nostalgia: affect, memory, and personality. *Emotion, 10*(3), 390-403.

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition and emotion, 23(7),* 1284-1306.

Barrett, L. F. (2017). *How emotions are made: the secret life of the brain*. New York: Macmillan.

Ben-Ze'ev, A. (2000). *The subtlety of emotions*. Cambridge, MA: MIT Press.

Benbaji, H. (2013). How is recalcitrant emotion possible? *Australasian Journal of Philosophy, 91*(3), 577-599.

Bermúdez, J. L. (1995). Nonconconceptual content: from perceptual experience to subpersonal computational states. *Mind and Language, 10*(4), 333-369.

Bermúdez, J. L. (2007). What is at stake in the debate on non conceptual content? *Philosophical Perspectives, 21*, 55-72.

Blackman, R. D. (2013). Intentionality and compound accounts of the emotions. *Southern Journal of Philosophy, 51*(1), 67-90.

Briesen, J. (2015). Perceptual justification and assertively representing the world. *Philosophical studies*, (172). 2239-2245.

Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford: Oxford University Press.

Bortolotti, L. (2015). The epistemic innocence of motivated delusions. *Consciousness and Cognition, 33*(C), 490-499.

Bortolotti, L. (2016). Epistemic benefits of elaborated and systematized delusions in schizophrenia. *The British Journal for the Philosophy of Science, 67*(3), 879-900.

Bortolotti, L. (2018). Delusions and three myths of irrational belief. In L. Bortolotti (Ed.), *Delusions in Context*. Cham: Palgrave Macmillan.

Bortolotti, L., and Sullivan-Bissett, E. (2018). The epistemic innocence of clinical memory distortions. *Mind and Language, 33*(3), 263-279.

Brady, M. (2009). The irrationality of recalcitrant emotions. *Philosophical Studies, 145*(3), 413-430.

Brady, M. (2013). *Emotional insight: the epistemic role of emotional experience*. Oxford: Oxford University Press.

Brentano, F. C. (1995). *Descriptive psychology*. London: Routledge.

Briefer, E. F. (2018). Vocal contagion of emotions in non-human animals. *Proceedings. Biological sciences, 285*(1873).

Calabi, C. (2013). Emozioni ambivalenti e contraddizioni (D. Bordini, Trans.). In C. Tappolet, F. Teroni, and A. Konselmann ziv (Eds.), *Le ombre dell'anima: pensare le emozioni negative*. Milano: Raffaello Cortina Editore.

Chosewood, L. C., and Wilson, D. E. (2009). *Biosafety in microbiological and biomedical laboratories*. Online resource: Centers for Disease Control and Prevention National Institutes of Health

Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: a reply to Jerry Fodor. *Philosophy of Science, 55*(2), 167-187.

Clifford, W. K. (2011). The ethics of belief. In L. Stephen, W. K. Clifford, and F. Pollock (Eds.), *Lectures and Essays* (Vol. 2). Cambridge: Cambridge University Press.

Clinton, J., and Roush, C. (2016). Poll: persistent partisan divide over "birther" question. *NBC News*.

Clore, G. L., and Ortony, A. (2000). Cognition in emotion: always, sometimes, or never? In R. D. Lane, L. Nadel, and G. Ahern (Eds.), *Cognitive neuroscience of emotion*. New York: Oxford University Press.

Coeckelbergh, M. (2010). Moral apperances: emotions, robots and human morality. *Ethics and Information Technology*. 12(3), 235-241.

Colombetti, G. (2014). *The feeling body: affective science meets the enactive mind*. London, England: The MIT Press.

Coltheart, M. (2007). Cognitive neuropsychiatry and delusional belief. *Quarterly journal of experimental psychology (2006), 60*(8), 1041.

Coplan, A. (2011). Understanding Empathy: Its Features and Effects. In A. Coplan and P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives*. Oxford: Oxford University Press.

Cosmides, L., and Tooby, J. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, and J. Tooby (Eds.), *The adapted mind: evolutionary psychology and the generation of culture.* Oxford: Oxford University Press.

Crane, T. (1988). The waterfall illusion. *Analysis, 48*(3), 142-147.

Crane, T. (1992). The nonconceptual content of experience. In T. Crane (Ed.), *The contents of experience: essays on perception*. Cambridge: Cambridge University Press. 136-157.

D'Arms, J. (2013). Value and the regulation of the sentiments. *Philosophical Studies, 163*(1), 3-13.

D'Arms, J., and Jacobson, D. (2000a). Sentiment and value. *Ethics, 110*(4), 722-748.

D'Arms, J., and Jacobson, D. (2000b). The moralistic fallacy: on the "appropriateness" of emotions. *Philosophy and Phenomenological Research, 61*(1), 65-90.

D'Arms, J., and Jacobson, D. (2003). VIII. The significance of recalcitrant emotion (or, anti-quasijudgmentalism). *Royal Institute of Philosophy Supplement, 52*, 127-145.

Damasio, A. R. (2004). *Looking for Spinoza: joy, sorrow, and the feeling brain*. London: Vintage.

Damasio, A. R. (2006). *Descartes' error: emotion, reason and the human brain.* London: Vintage.

Darwin, C. R. (1872). *The expression of emotions in man and animals*. New York: Philosophical Library.

Davies, S. (2011). *Emotional contagion from music to listener*: Oxford University Press.

de Sousa, R. (1987). *The rationality of emotion*. London: MIT Press.

de Sousa, R. (2008). Epistemic feelings. In G. Brun, U. Doguoglu, and D. Kuenzle (Eds.), *Epistemology and emotions*. Hampshire: Ashgate.

DeLancey, C. (2004). *Passionate engines: what emotions reveal about the mind and artificial intelligence*. Cary, NC, USA: Oxford University Press.

Deonna, J., and Teroni, F. (2012). *Emotions: a philosophical Introduction*. Florence: Taylor and Francis.

Deonna, J., and Scherer, K. R. (2010). The case of the disappearing intentional object: constraints on a definition of emotion. *Emotion Review, 2*(1), 44-52.

Dezecache, G., Jacob, P., and Grèzes, J. (2015). Emotional contagion: its scope and limits. *Trends in Cognitive Sciences, 19*(6), 297-299.

Döring, S. A. (2007). Seeing what to do: affective perception and rational motivation. *Dialectica, 61*(3), 363-394.

Döring, S. A. (2009). The logic of emotional experience: noninferentiality and the problem of conflict without contradiction. *Emotion Review, 1*(3), 240-247.

Döring, S. A. (2015). What's wrong with recalcitrant wmotions? From irrationality to challenge of agential identity. *Dialectica, 69*(3), 381-402.

Dôkic, J. (2018). The uncanny and other negative existential feelings. In Tappolet, C; Teroni, F., Konzelmann-Ziv, A. *The shadows of the soul*. Cambridge: Cambridge University Press.

Echeverri, S. (2017). Emotional justification. *Philosophy and Phenomenological Research*.

Ekman, P. (1972). *Emotions in the human face*. New York: Pergamon Press.

Ekman, P. (1977). Biological and cultural contributions to body and facial movement. In P. Ekman (Ed.), *The anthropology of the body*. London: Academic Press.

Ekman, P. (1992). *Telling lies: cues to deceit in the marketplace, politics and marriage*. New York: Norton.

Ekman, P., Friesen, W. V., and Simons, R. C. (1985). Is the startle reaction an emotion? *Journal of personality and social psychology, 49*(5), 1416-1426.

Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221*(4616), 1208-1210.

Elgin, C. Z. (2008). Emotion and understanding. In G. Brun, U. Dogluoglu, and D. Kuenzle (Eds.), *Epistemology and emotions*. Hampshire: Ashgate.

Ellis, A. (1962). *Reason and emotion in psychotherapy*. Secaucus NJ: Citadel Press.

Ellsworth, P. C., and Scherer, K. R. (2001). Appraisal Process in Emotion. In R. J. Davidson, Scherer, Klaus R., and Goldsmith, H. Hill (Ed.), *Handbook of affective sciences*. Cary, USA: Oxford University Press. 572-595.

Ersner-Hershfield, H., Mikels, J. A., Sullivan, S. J., and Carstensen, L. L. (2008). Poignancy: Mixed emotional experience in the face of meaningful endings. *Journal of Personality and Social Psychology, 94*(1), 158-167.

Evans, G. (1982). *The varieties of reference*. Oxford: Clarendon Press.

Faucher, L., and Tappolet, C. (2007). Introduction: modularity and the nature of emotions. *Canadian Journal of Philosophy, 36*(32), vii-xxxi.

Faucher, L. T., Christine. (2002). Fear and the focus of attention. *Consciousness and Emotions, 3*(2), 105-144.

Feldman, R. (2000). The ethics of belief. *Philosophy and Phenomenological Research, 60*(3), 667-695.

Fischer, J. M., and Ravizza, M. (1998). *Responsibility and control: a theory of moral responsibility*: Cambridge University Press.

Fletcher, G. (2016). Objective list theories. In G. Fletcher (Ed.), *The Routledge handbook of philosophy of well-being*. London: Routledge.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*: MIT Press.

Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *The Journal of Philosophy, 66*(23), 829-839.

Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.

Gendler, T. (2016). Imagination. In Zalta, E. (Ed.), *The Stanford Encyclopedia of Philosophy*, Retrieved from: http://plato.stanford.edu/archives/sum2018/entries/imagination/.

Ginet, C. (2001). *Deciding to believe*: Oxford University Press.

Goldfried, M. P., and Goldfried, A. P. (1980). Cognitive change methods. In A. P. Goldstein and F. H. Kanfer (Eds.), *Helping people change*: Elsevier.

Goldie, P. (2000). *The emotions: a philosophical exploration*. Oxford: Clarendon Press.

Goldman, A. I. (2012). *What is justified belief?*. In *Reliabilism and contemporary epistemology: essays*. Oxford Scholarship Online.

Golkar, A., Lonsdorf, T. B., Olsson, A., Lindstrom, K. M., et al. (2012). Distinct contributions of the dorsolateral prefrontal and orbitofrontal cortex during emotion regulation (distinct cortical loci for emotion regulation).PLOS ONE, *7*(11), e48107.

Goodman, N. (1954). *Fact, fiction and forecast*. London: Athlone Press.

Gordon, R. M. (1986). The passivity of emotions. In *The structure of emotions*. Cambridge: Cambridge University Press.

Grzankowski, A. (2017). The real trouble with recalcitrant emotions. *Erkenntnis*. 3, 1-11.

Greenspan, P. (1980). A case of mixed feelings: ambivalence and the logic of emotion. In R. A. O. (Ed.), *Explaining emotions*. London: University of California Press. 223-250

Greenspan, P. (2004). Emotions, rationality, and mind/body. In R. C. Solomon (Ed.), *Thinking about feeling: contemporary philosophers on emotions*. Oxford: Oxford University Press.

Greenspan, P. (1988). *Emotions and reasons: an inquiry into emotional justification*. London: Routledge.

Griffiths, P. E. (1996). The historical turn in the study of adaptation. *The British Journal for the Philosophy of Science, 47*(4), 511-532.

Griffiths, P. E. (1990). Modularity, and the psychoevolutionary theory of emotion. *Biology and Philosophy, 5*(2), 175-196.

Griffiths, P. E. (1997). *What emotions really are: the problem of psychological categories*. Chicago: University of Chicago Press.

Gross, J. J. (2014). Emotional regulation: conceptual and empirical foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (2nd ed.). London: The Guilford Press.

Gross, J. J. (2015). Emotion regulation: current status and future prospects. *Psychological Inquiry, 26*(1), 1-27.

Gunther, Y. H. (2003). *Essays on nonconceptual content*. London: MIT.

Gutentag, T. (2017). Successful emotion regulation requires both conviction and skill: beliefs about the controllability of emotions, reappraisal, and regulation success. *Cognition and Emotion, 31*(6), 1225-1234.

Gyurak, A., Gross, J., and Ektin, A. (2011). Explicit and implicit emotion regulation: A dual-process framework. *Cognition and Emotion, 25*(3), 400-413.

Hatfield, E., Bensman, L., Thornton, P. D., and Rapson, R. L. (2014). *New perspectives on emotional contagion: a review of classic and recent research on facial mimicry and contagion* (Vol. 8).

Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1994). *Emotional contagion* (Kindle version ed.). Cambridge: Cambridge University Press.

Hatzimoysis, A. (2009). Emotions in Heidegger and Sartre. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion*: Oxford University Press.

Hauner, K. K., Mineka, S., Voss, J. L., and Paller, K. A. (2012). Exposure therapy triggers lasting reorganization of neural fear processing. *Proceedings of the National Academy of Sciences U S A, 109*(23), 9203-9208.

Hayes, S., Hirsch, C., and Mathews, A. (2008). Restriction of working memory capacity during worry. *Journal of Abnormal Psychology, 117*(3), 712-717.

Helm, B. (2001). *Emotional reason*. Cambridge: Cambridge University Press.

Helm, B. W. (2001). *Emotional reason: deliberation, motivation, and the nature of value*. Cambridge: Cambridge University Press.

Herzberg, L. A. (2009). Direction, causation, and appraisal theories of emotion. *Philosophical Psychology, 22*(2), 167-186.

Hieronymi, P. (2008). Responsibility for believing. *An International Journal for Epistemology, Methodology and Philosophy of Science, 161*(3), 357-373.

Hofmann, S. G. (2004). Cognitive mediation of treatment change in social phobia. *Journal of Consulting and Clinical Psychology, 72*(3), 392-399.

Hofmann, S. G. (2008). Cognitive processes during fear acquisition and extinction in animals and humans: implications for exposure therapy of anxiety disorders. *Clinical Psychology Review, 28*(2), 199-210.

Huemer, M. (2001). *Skepticism and the veil of perception*. Oxford: Rowman and Littlefield.

Hume, D. (2000 [1738]). *A treatise of human nature*. Oxford: Oxford University Press.

Husserl, E. (1989). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy* (R. A. S. Rojcewicz, André Trans.). Dordrecht: Kluwer.

Irvine, E. (2014). Problems and possibilities for empirically informed philosophy of mind. In M. Sprevak and J. Kallestrup (Eds.), *New waves in philosophy of mind*. London: Palgrave Macmillan UK. 185-207.

Jacobson, D. (2011). Fitting attitude theories of value. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2011 ed., Vol. 2011): Metaphysics Research Lab, Standford University.

Jaggar, A. M. (1989). Love and knowledge: emotion in feminist epistemology. *Inquiry, 32*(2), 151-176.

James, W. (1884). What is an emotion? *Mind, os-IX*(34), 188-205.

Johnston, M. (2001). The authority of affect. *Philosophy and Phenomenological Research, 63*(1), 181-214.

Jokić-Begić, N. (2010). Cognitive-behavioral therapy and neuroscience: towards closer integration. *Psychological Topics, 19*(2), 235-254.

Justlin, P., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behavioural and Brain Sciences, 31*, 559-621.

Kappas, A. (2008). Psssst! Dr. Jekyll and Mr. Hyde are actually the same person! A tale of regulation and emotion. In M. Vandekerckhove, C. von Scheve, S. Ismer, S. Jung, and S. Kronast (Eds.), *Regulating emotions: culture, social necessity, and biological inheritance*. Malden, MA: Blackwell Publishing. 15-38.

Keefe, R., and Smith, P. (1996). *Vagueness: a reader*. London: MIT Press.

Kenny, A. (2013). *Action, emotion and will* (2nd ed.). Hoboken: Taylor and Francis.

Kivy, P. (1990). *Music alone: philosophical reflections on the purely musical experience*. New York: Cornell University Press.

Kivy, P. (2002). *Introduction to a philosophy of music*. Oxford: Clarendon Press.

Kristjánsson, K. (2010). The trouble with ambivalent emotions. *Philosophy, 85*(334), 485-510.

Larsen, J. T., McGraw, A. P., and Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology, 81*(4), 684-696.

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

Leahy, R. L., and Rego, S. A., (2012). Cognitive restructuring. In W. T. O'Donohue (Ed.), *Cognitive behavior therapy: core principles for practice*. Hoboken, NJ, USA: John Wiley and Sons.

LeDoux, J. (1989). Cognitive-emotional interactions in the brain. *Cognition and Emotion, 3*(4), 267-289.

LeDoux, J. (2015). *Ansia: come il cervello ci aiuta a capirla* (G. Gianbruno, Trans.). Varese: Raffaello Cortina Editore.

Levenson, R. W., Ekman, P., and Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27*(4). 363-384.

Lewens, T. (2007). *Darwin*. London: Routledge.

Lewis, D. (1989). Dispositional theories of value. *Proceedings of the Aristotelian Society, 63*, 113.

Lormand, E. (1985). Toward a theory of moods. *An International Journal for Philosophy in the Analytic Tradition, 47*(3), 385-407.

Lyons, W. (1980). *Emotion*. Cambridge: Cambridge University Press.

Machperson, F. (2011). Cognitive penetration of colour experience: rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research, 84*(1), 24-62.

Madva, A. (2018). Implicit bias, moods, and moral responsibility. *Pacific Philosophical Quarterly, 99*(S1), 53-78.

Mandelbaum, E. (2016). Attitude, inference, association: on the propositional structure of implicit bias. *Noûs, 50*(3), 629-658.

Mattick, R. P., Peters, L., and Clarke, J. C. (1989). Exposure and cognitive restructuring for social phobia: A controlled study. *Behavior Therapy, 20*(1), 3-23.

Mauss, I. B., and Robinson, M. D. (2009). Measures of emotion: a review. *Cognition and Emotion, 23*(2), 209-238.

McDowell, J. H. (1985). Values and secondary qualities. In T. Honderich (Ed.), *Morality and objectivity: a tribute to J.L. Mackie*. London: Routledge.

McHugo, G. J., Lanzetta, J. T., Sullivan, D. G., Masters, R. D., et al. (1985). Emotional reactions to a political leader's expressive displays. *Journal of Personality and Social Psychology, 49*(6), 1513-1529.

Meiselman, H. L. (2016). *Emotion measurement*. Amsterdam, Netherlands: Woodhead Publishing.

Meltzoff, A. N., and Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*(4312), 75-78.

Meylan, A. (2014). Epistemic emotions: a natural kind? *Philosophical Inquiries, 2*(1), 173-191.

Millikan, R. G. (1995). Pushmi-pullyu representations. *Philosophical Perspectives, 9*, 185-200.

Mineka, S., and Öhman, A. (2002). Phobias and preparedness: the selective, automatic, and encapsulated nature of fear. *Biological Psychiatry, 52*(10), 927-937.

Monteleone, J. (2017). Attention, emotion, and evaluative understanding. *Philosophical Quarterly of Israel, 45*(4), 1749-1764.

Moors, A. (2010). Automatic constructive appraisal as a candidate cause of emotion. *Emotion Review, 2*(2), 139-156.

Moors, A., and Scherer, K. R. (2013). The role of appraisal in emotion. In M. Robinson, E. Watkins, and E. Harmon-Jones (Eds.), *Handbook of cognition and emotion*. New York: Guilford Press.

Morgan, H. (2003). *Real learning: a bridge to cognitive neuroscience*. Oxford: Scarecrow Education.

Morton, A. (2002). Emotional truth: II emotional accuracy. *Aristotelian Society Supplementary Volume, 76*(1), 265-275.

Morton, A. (2010). Epistemic emotions. In Peter Goldie (Ed.) *The Oxford handbook of philosophy of emotion*. Oxford University Press. 385-399.

Naar, H., (2018). The real issue with recalcitrant emotions. Reply to Grzankowski. *Erkenntnis*. 1-6.

Nichols, S., and Stich, S. (2000). A Cognitive theory of pretense. *Cognition, 74*(2), 115-147.

Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259.

Nolfi, K. (2015). Which mental states are rationally evaluable, and why? *Philosophical Issues, 25*(1), 41-63.

Nummenmaa, L., Glerean, E., Hari, R., and Hietanen, J. K. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences of the United States of America, 111*(2), 646.

Nussbaum, M. (2001). *Upheavals of thought: the intelligence of the emotions*. Cambridge: Cambridge University Press. 183-199.

Nussbaum, M. (2004). Emotions as judgments of value and importance. In R. C. Solomon (Ed), *Thinking about Feeling: Contemporary Philosophers on Emotions*. Cary, NC, USA: Oxford University Press.

Ofri, D. (2013). *What doctors feel: how emotions affect the practice of medicine* (B. Press Ed.). Boston.

Olson, D. (2015). A case for epistemic agency. *Logos and Episteme, 6*(4), 449-474.

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, et al. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology, 26*(10), 1334-1338.

Oswald, A. J., Proto, E., and Sgroi, D. (2015). Happiness and productivity. *Journal of Labor Economics, 33*(4), 789-823.

Öhman, A., and Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483-522.

Öst, L.-G., and Clark, D. M. (2013). Cognitive behaviour therapy: principles, procedures and evidence base. In E. Skaret (Ed.), *Cognitive behavioral therapy for dental phobia and anxiety* (2nd ed.). Somerset, NJ, USA: John Wiley and Sons.

Panksepp, J. (1998). *Affective neuroscience: the foundations of human and animal emotions*. New York: Oxford University Press.

Paquette, V., Levesque, J., Mensour, B., Leroux, J. M., et al. (2003). "Change the mind and you change the brain": effects of cognitive-behavioral therapy on the neural correlates of spider phobia. *Neuroimage, 18*(2), 401-409.

Parfit, D. (1984). *Reasons and persons*. Oxford: Clarendon Press.

Parkinson, B., Totterdell, P., Briner, R., and Reynolds, S. (1996). *Changing moods: the psychology of mood and mood regulation*. London: Longman.

Peacocke, C. (1992). *A study of concepts*. London: MIT Press.

Pelser, A. (2014). Emotion, evaluative perception and epistemic justification. In C. Roeser Sabine; Todd (Ed.), *Emotion and value*. New York: Oxford University Press. 107-123.

Picard, R. W. (1997). *Affective computing*. London: MIT.

Pickard, H. (2013). Responsibility without blame. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, et al. (Eds.), *The Oxford handbook of philosophy and psychiatry* (1st ed.). Oxford: Oxford University Press. 1168-1185.

Pickard, H. (2017). Responsibility without blame for addiction. *Neuroethics, 10*(1), 169-180.

Pinker, S. (1998). *How the mind works*. London: Allen Lane The Penguin Press.

Pitcher, G. (1965). Emotion. *Mind, 74*, 326.

Pollock, J. (1987). Epistemic norms. *An International Journal for Epistemology, Methodology and Philosophy of Science, 71*(1), 61-95.

Preston, S. D., and de Waal, F. B. M. (2002). Empathy: its ultimate and proximate bases. *Behavioural and Brain Sciences., 25*(1), 1-20.

Price, C. (2006). Affect without object: moods and objectless emotions. *European Journal of Analytic Philosophy, 2*(1), 49-68.

Price, C. (2013). The problem of emotional significance. *Acta Analytica, 28*(2), 189-206.

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

Prinz, J. (2004). *Gut reactions: a perceptual theory of emotion*. Oxford: Oxford University Press.

Prinz, J. (2010). The moral emotions. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion*. Oxford: Oxford University Press.

Prinz, J. (2004). *Gut reactions: a perceptual theory of emotion*. Oxford: Oxford University Press.

Prinz, J. (2006a). Is emotion a form of perception? *Canadian Journal of Philosophy, 32*(Supplement).

Prinz, J. (2006b). Is the mind really modular? In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (Vol. xiv): Blackwell Publishing. 22-36.

Prochazkova, E., and Kret, M. E. (2017). Connecting minds and sharing emotions through mimicry: a neurocognitive model of emotional contagion. *Neuroscience and Biobehavioral Reviews, 80*, 99-114.

Pryor, J. (2000). The skeptic and the dogmatist. *Noûs, 34*(4), 517-549.

Puddifoot, K., and Bortolotti, L. (2018). Epistemic innocence and the production of false memory beliefs. *Philosophical Studies*, 1-26.

Pugmire, D. (2005). *Sound sentiments: integrity in the emotions*. Oxford: Clarendon Press.

Putnam, H. (1967). Psychological predicates. In W. H. Capitan and D. D. Merrill (Eds.), *Art, mind, and religion*. Liverpool: University of Pittsburgh Press. 37-48.

Pylyshyn, Z. (1980). Computation and cognition. *Behavioral and Brain Sciences, 3*(1), 111-132.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*(3), 341-365; discussion. 366-423.

Raichle, M. (1999). Positron emission tomography. In R. A. Wilson and F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences*. Cambridge: The MIT Press. 656-658.

Railton, P. A. (2003). Facts and values. In *Facts, values, and norms: essays toward a morality of consequence*. Cambridge: Cambridge University Press.

Rawls, J. (1999). *A theory of justice* (Rev. ed.). Oxford: Oxford University Press.

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology, 67*(3), 525-539.

Rice, C. M. (2013). Defending the objective list theory of well-being. *Ratio, 26*(2), 196-211.

Roberts, R. (2003). *Emotions: an essay in aid of moral psychology*. Cambridge: Cambridge University Press.

Robinson, J. (2005). *Deeper than reason: emotion and its role in literature, music, and art*. Oxford: Oxford: Clarendon Press.

Robinson, J. (2009). Emotional responses to music: what are they? How do they work? And are they relevant to aesthetic appreciation? In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion*. Oxford. 651-680.

Roseman, I. (1991). Appraisal determinants of discrete emotions. *Cognition and Emotion, 5*(3), 161-200.

Ross, L., Lepper, M. R., and Hubbard, M. (1975). Perseverance in self- perception and social perception: biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*(5), 880-892.

Ryan, S. (2003). Doxastic compatibilism and the ethics of belief. *Philosophical Studies, 114*(1), 47-79.

Salmela, M. (2006). True emotions. *Philosophical Quarterly, 56*(224), 382-405.

Salmela, M. (2014). *True emotions*. Amsterdam: John Benjamins Publishing Company.

Sartre, J.-P. (1962). *Sketch for a theory of the emotions* (P. Mairet and M. Warnock, Trans.). London: Routledge.

Scarantino, A. (2010). Insights and blindspots of the cognitivist theory of emotions. *The British Journal for the Philosophy of Science, 61*(4), 729-768.

Scarantino, A., and de Sousa, R. (2018). Emotion. *The Standford Encyclopedia of Philosophy, Winter 2018 Edition*. Retrieved from https://plato.stanford.edu/entries/emotion/.

Schachter, S., and Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review, 69*(5), 379-399.

Scherer, K. R., Schorr, A., and Johnstone, T. (1984). *Appraisal processes in emotion: Theory, methods, research*. Oxford: OUP.

Schienle, A., Schäfer, A., Stark, R., and Vaitl, D. (2009). Long-term effects of cognitive behavior therapy on brain activation in spider phobia. *Psychiatry Research: Neuroimaging, 172*(2), 99-102.

Schienle, A., Wabnegger, A., and Scharmüller, W. (2014). Effects of cognitive behavior therapy on regional brain volume in spider-phobic patients: Preliminary results. *Journal of Anxiety Disorders, 28*(2), 276-279.

Seidman, J. (2016). The unity of caring and the rationality of emotion. *An International Journal for Philosophy in the Analytic Tradition, 173*(10), 2785-2801.

Shakespeare, W. (2016 [1605]). *A midsummer night's dream* (M. Pagnini, Trans. Kindle ed.). Milano: Garzanti.

Shakespeare, W. (1987 [1608]). *King Lear*. Bristol: Bristol Classical Press.

Shakespeare, W. (2015 [1623]). *Macbeth*. (V. Gassman, Trans. Kindle ed.). Milano: Mondadori.

Shakespeare, W. (2005 [1622]). *Othello*. Cambridge: Cambridge University Press.

Shakespeare, W. (2008 [1623]). *The tempest*. Auckland: The Floating Press.

Shargel, D., and Prinz, J. (2018). An enactivist theory of emotional content. In H. Naar and F. Teroni (Eds.), *The ontology of emotions*. Cambridge: Cambridge University Press. 110-129.

Shea, N. (2014). Distinguishing top-down from bottom-up effects. In M. S. Biggs, D. Stokes (Ed.), *Perception and its modalities*: Oxford University Press.

Shelley, M. W. (2012 [1818]). *Frankenstein, or, the modern Prometheus*. Oxford: Oxford University Press.

Siegel, S. (2017). *The rationality of perception*. Oxford: Oxford University Press.

Simard, S., Savard, J., and Ivers, H. (2010). Fear of cancer recurrence: specific profiles and nature of intrusive thoughts. *Research and Practice, 4*(4), 361-371.

Simard, S., Thewes, B., Humphris, G., Dixon, M., et. al. (2013). Fear of cancer recurrence in adult cancer survivors: a systematic review of quantitative studies. *Research and Practice, 7*(3), 300-322.

Simons, R. C. (1996). *Boo!: culture, experience, and the startle reflex*. In. New York: Oxford University Press.

Smith, C. A. (2009). Putting appraisal in context: Toward a relational model of appraisal and emotion. *Cognition and Emotion, 23*(7), 1352-1373.

Smith, C. A., and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*(4), 813-838.

Smith, M. (1989). Dispositional theories of value: I - Michael Smith. *Proceedings of the Aristotelian Society, 63*, 89-111.

Solomon, R. C. (1977). *The passions*. New York: Anchor.

Solomon, R. C. (1988). On emotions as judgments. *American Philosophical Quarterly, 25*(2), 183-191.

Solomon, R. C. (2007). *True to our feelings: what our emotions are really telling us*. Oxford: Oxford University Press.

Spinoza, B. d. (2000). *Ethics*. Oxford: Oxford University Press.

Stich, S. (2011). *Beliefs and subdoxastic states*: Oxford University Press.

Stocker, M. (2013). *Intellectual and other nonstandard emotions*: Oxford University Press.

Straube, T., Glauer, M., Dilger, S., Mentzel, H.-J., et al. (2006). Effects of cognitive-behavioral therapy on brain activation in specific phobia. *Neuroimage, 29*(1), 125-135.

Sweeny, K., and Dooley, M. D. (2017). The surprising upsides of worry. *Social and Personality Psychology Compass, 11*(4).

Tappolet, C. (2000). *Émotions et valeurs*. Paris: Presses Universitaires de France.

Tappolet, C. (2004). Through thick and thin: good and its determinates. *Dialectica, 58*(2), 207-221.

Tappolet, C. (2005). Ambivalent emotions and the perceptual account of emotions. *Analysis, 65*(287), 229-233.

Tappolet, C. (2012). Emotions, perceptions, and emotional illusions. In C. Clotilde (Ed.), *Perceptual Illusions. Philosophical and psychological essays,* Palgrave-Macmillan. 207-224.

Tappolet, C. (2012). Emotions, perceptions, and emotional illusions. In C. Calabi (Ed.), *Perceptual Illusions. Philosophical and Psychological Essays.* London: Palgrave-Macmillan. 207-224.

Tappolet, C. (2016a). Emotion and motivation. In C. Tappolet (Ed.), *Emotions, values and agency*. Oxford: Oxford University Press.

Tappolet, C. (2016b). *Emotion and perception*. In *Emotions, values and agency*. Oxford: Oxford University Press.

Tappolet, C. (2016c). Emotions, values, and agency. Oxford: Oxford University Press.

Taylor, G. (1975). Justifying the emotions. *Mind, 84*(335), 390-402.

Teroni, F. (2007). Emotions and formal objects. *Dialectica, 61*(3), 395-415.

Teroni, F. (2017). *In pursuit of emotional modes: the philosophy of emotion after James*: Oxford University Press.

Thompson, R. (1991). Emotional regulation and emotional development. *Education Psychological Review, 3*(4), 269-307.

Tinbergen, N. (1974). *The study of instinct: with a new introduction*. Oxford: Oxford University Press.

Tomkins, S. S. (1962). Affect imagery consciousness. The positive affects. In *Positive affects* (Vol. 1). New York: Springer.

Toribio, J. (2007). Nonconceptual content. *Philosophy compass, 2*(3), 445-460.

Tucker, C. (2010). Why open-minded people should endorse dogmatism?. *Philosophical Perspectives* (24). p. 529-545.

Tye, M. (2000). *Consciousness, color, and content*. London: MIT Press.

Ugurbil, K. (1999). Magnetic resonance imaging. In R. A. Wilson and F. C. Kail (Eds.), *The MIT encyclopedia of the cognitive sciences*. Cambridge: The MIT Press. 505-507.

Valente, S. M. (2002). Social Phobia. *Journal of the American Psychiatric Nurses Association, 8*(3), 61-75.

Vineet, A., Sertan Kutal, G., Sangjun, L., Bo, et al. (2018). Hierarchical neural architecture underlying thirst regulation. *Nature*. *555*, 204-209.

Weatherson, B. (2008). Deontology and Descartes's demon. *Journal of Philosophy, 105*(9), 540-569.

Wegner, D. M. (2002). *The illusion of conscious will*. London: MIT Press.

Wiggins, D. (1998). A sensible subjectivism? In *Needs, values, truth: essays in the philosophy of value* (3rd ed.). Oxford: Clarendon Press.

Williams, B. (2011). *Ethics and the limits of philosophy*. Florence: Taylor and Francis.

Williamson, T. (1994). *Vagueness*. London: Routledge.

Wright, C. (1992). *Truth and objectivity*. London: Harvard University Press.

Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist, 39*(2), 117-123.