



Universiteit
Leiden
The Netherlands

Emotional privacy: a philosophical investigation of emotion recognition technology

Prégent, A.P.

Citation

Prégent, A. P. (2026, September 2). *Emotional privacy: a philosophical investigation of emotion recognition technology*. Retrieved from <https://hdl.handle.net/1887/4309114>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4309114>

Note: To cite this publication please use the final published version (if applicable).

Part Two:
The Emotional Life

3. Emotions: On their nature, role, functions, and relation to expressions

3.1. Introduction

This thesis aims to identify the key ethical, social, and normative impacts involved in tracking affective life – moving beyond purely technical challenges. Given that affective life is a complex, multifaceted, and deeply social phenomenon, breaking down disciplinary silos in research on emotions and emotional expressions is essential. To gain a clearer understanding of their respective functions and the core elements of their interrelations, I draw on scholarship from multiple fields, including philosophy, psychology, affective computing, and affective neuroscience. This discussion establishes the necessary theoretical foundation regarding affective life, and lays the groundwork for the normative analysis presented in the last part of the thesis.

Accordingly, in this chapter and the next I examine the intricate relationships between emotion, emotional expression, and emotion recognition. This chapter focuses on the phenomenon of emotion, analysing its nature, its role in (moral) decision-making, judgements, and actions and its main functionalities at the intrapersonal level. The chapter seeks to combine normative insights from philosophy, especially work on human morality and the nature of emotions, with empirical findings from (social) psychology and affective neuroscience.

Section 3.2 briefly introduces the notion of emotions and disentangles the concept from other intentional mental states, namely desires, beliefs, and judgements. **Section 3.3** revisits canonical philosophical accounts of human reasoning and morality and analyses in particular their treatment of the relations between emotions and morality. This analysis underscores how the concepts of emotion, moral salience, and values are deeply intertwined and how emotions are an essential part of moral reasoning and moral actions – even within accounts that explicitly try to keep emotions separated from moral reasoning. **Section 3.4** expands the analysis by adopting an integrative perspective on emotions that include transdisciplinary insights structured along three broadly conceived perspectives: biological/evolutionary, cognitivist, and social constructionist. **Section 3.5** relies on key points of agreement between these perspectives, and then further turns to the relation between emotions and their expressions, focusing particularly on their action-tendencies, typically associated physiological changes, and performative potential to provide an integrative and function-first overview of emotions.

3.2. Emotions

In the effort to understand the nature, role, and functions of emotions, philosophy has, particularly over the past century, been increasingly joined by psychology, cognitive science, affective neuroscience, and, more recently, affective computing.⁴⁷ This has given rise to transdisciplinary contributions that have significantly expanded the scope of inquiry by shifting attention from the moral relevance of emotions to their broader significance in human life. This expanded lens allows for a richer appreciation of the complex roles emotions play not only in individual deliberation but also in social interaction, communication, and culture.

Emotions are highly complex phenomena that stand in close relation to other mental states, notably beliefs, judgements, and desires. Beliefs are typically recognised as cognitive states that represent the world as being a certain way and are evaluable for truth or falsity. Judgements are often taken as a subset of beliefs, marked by their normative commitment and evaluative character. Desires are conative states aimed at realising particular states of affairs and are often assessed in terms of prudential or moral coherence. Emotions, like these other states, possess intentional content – be it propositional, object-directed, or representational – and are typically evaluated in terms of their fittingness or appropriateness (i.e. whether the emotion suits its object in light of contextual and normative factors).

Emotions share significant features with beliefs, desires, and judgements, often blurring the boundaries between them. They are *belief-like* in their responsiveness to reasons and evidence, and in their potential revisability in light of new information (Greenspan 1988; Ben-Ze’ev 2000). They are *judgement-like* due to their evaluative character and their capacity to commit agents to certain stances or actions (Nussbaum 2001; Solomon 1976). Finally, they are *desire-like* in that they exert motivational force, prompting agents towards action (Goldie 2000; Hume [1739]2003). Some theorists go further, treating emotions as *subtypes* of beliefs, judgements, or desires (e.g. Hume [1739]2003), while others view them as hybrid states partly constituted by those elements (e.g. Greenspan 1988; Ben-Ze’ev 2000; Prinz 2004; Tappolet 2017).

Crucially, what helps disentangle emotions from their counterparts is that while they are all typically regarded as intentional states, emotions are characterised by a distinctive phenomenological dimension: there is something *it is like* to experience an emotion. This

⁴⁷ In this work, I will follow the terminology of Dukes et al. (2021) and refer to these various disciplines as ‘affective sciences’ when referring to them in the general sense.

experiential aspect distinguishes them from purely cognitive states and affirms their role in shaping our lived reality. Emotions are thus situated between affective, cognitive, and conative faculties. They are felt, evaluative, and motivational, capable of guiding perception, practical reasoning, and decision-making. Furthermore, emotions are profoundly interwoven with perceptual and social dimensions. They are central to interpersonal relationships and responsive to cultural and normative expectations, often shaped by socially embedded practices of expression, regulation, and interpretation. This theoretical diversity testifies to the conceptual richness of emotions and the tensions that arise in attempts to capture their nature.

3.3. Emotions and morality

Before turning to the transdisciplinary discussion on the complex topic that emotions represent, I first examine their relation to morality and the canonical works that first brought these concepts together. Much of the transdisciplinary research on emotions begins either by acknowledging the significant contributions and insights of these early accounts into emotional life or by critiquing their notable misconceptions of the phenomenon – both of which have profoundly shaped the ways in which emotions are conceptualised today. The nature and function of emotions, especially in relation to (moral) reasoning, have been central to philosophical inquiries into the human mind. Philosophers such as Plato, Aristotle, Hume, and Kant, among others, have laid the foundational work for exploring the role of emotions in morality. Primarily exploring this role within the broader context of developing ethical theories and conceptualising morality (Tappolet 2022), they interestingly took radically divergent positions on whether emotions are necessary for reasoning. While the work of Plato and Kant influenced later scholars – particularly within the ‘classical’ rationalist tradition – to view emotions as a threat to rationality and to reject them as potential core components of judgement (Solomon 1976), the work of Aristotle and Hume shaped competing traditions in virtue ethics and sentimentalism, emphasising the role of emotions in moral perception and as a key element of moral action.⁴⁸ Thus, engaging with both sides is necessary if we are to arrive at a more comprehensive understanding of the role and diverse functions of emotions in morality, and of how these traditions shaped later transdisciplinary studies.

⁴⁸ Interestingly, Plato’s tripartite division of the soul has been replicated in the early models of the mind in cognitive sciences, which continued to assume a strict division between cognition, conation, and affection (Hilgard 1980). It influenced early views in psychology and neurosciences, so much so that they rarely even questioned the division (Scherer 1984: 293). As noted by Frijda (1988), by Ellsworth and Scherer (2003), and by many other affective researchers, cognitive appraisal theories are also heavily influenced by philosophical reflections on emotions, particularly in the underlying role of pleasure and displeasure in different emotion-types such as happiness, anger, fear, and guilt.

3.3.1. Emotions as a threat to morality: Moral rationalism

Plato and Kant, who at least to some extent, appear to reject emotions from their conceptions of morality, particularly with regard to the possibility of making rational judgements, have at times conceded that emotions, or at least moral feelings, may play an instrumental role in moral reasoning. Their wariness of emotions seems to stem primarily from the belief that individuals have limited control over them, rendering emotions a potential obstacle to the attainment of rational judgement and moral action. Nonetheless, prominent interpretations of Plato's and Kant's theories have sought to reintegrate – at least partially – emotions into their respective frameworks of morality, arguing that their theories intrinsically require the endorsement of certain emotions to fulfil their aims. Precisely because of their (ultimately unsuccessful) ambitious attempt to demonstrate how motivations, moral actions, and moral reasoning might be achieved independently of emotions, their work reveals – perhaps more clearly than any other – the intrinsic interdependencies among emotions, motivations, actions, and (moral) reason.

Emotions in Plato's account of (moral) reasoning

Since the attainment of rational judgement in Plato's work depends heavily on the ability to control one's intellect, emotions were generally regarded as a threat to reason.⁴⁹ Yet, recent reinterpretations of Plato's work, such as those by Candiotto and Renaut (2020), Price (2009), Lorenz (2006), and Solomon (2008), argue that emotions may be integral to the philosophical journey and the development of one's moral reasoning.

Discussing “emotions” in Plato's work is always a delicate task, primarily because the term itself – of Latin origin – has no direct equivalent in ancient Greek. As Price notes, drawing on Liddell and Scott (1940), the Greeks instead used the term *pathos*, which broadly meant “that which happens to a person or thing” (Price 2009: 121). Consequently, scholars often prefer the terms “desires” or “motivations” when interpreting Plato's work, as these better capture the range of phenomena encompassed by *pathos*. This is particularly helpful, since Plato's work indeed covers experiences as diverse as thirst, fear, anger, and sexual arousal (Cooper 1999); phenomena that

⁴⁹ As noted by Solomon (1977) in Greece, *apatheia* (“apathy”) was a popular term that meant ‘freedom from the emotions’. *Apatheia* was believed to be a necessary step to attain *ataraxia*, a state where the individual achieves “peace of mind”. Moreover, it is interesting to note that this general idea, according to which it was preferable to distance oneself from one's own emotions, was also found in the non-Western views, as we can see in the Upanishads, Buddhists, Confucian, and Taoist theories of the mind (Solomon 1976).

contemporary affective researchers would not typically classify under a single (affective) category. Yet, setting aside these etymological and conceptual challenges, Plato's account nevertheless allows for a reading that anticipates some key elements of modern theories of emotion.

Candiotto and Renaut, for instance, argue for a reinterpretation of Plato's work in which some emotions should be recognised as having the potential to become "moral, just, and means for reason to rule over one's life" (Candiotto & Renaut 2020: 6). This concerns more particularly fear and love (*eros*). Plato's account attributes cognitive-like properties to certain desires and motivations that appear to arise from the lower parts of the soul, suggesting an early recognition of the role that affective states play in *shaping reasoning* and *guiding decision-making*. As Lorenz notes, "Plato [...] think[s] that the non-rational parts of the soul are able not only to give rise to fully formed motivating conditions, but also to be the bearers of expectations and of forward-looking emotions such as fear (Timaeus 69 C 5–D 4)" (Lorenz 2006: 58). Candiotto and Renaut support Lorenz's interpretation of the *Timaeus*, adding that Plato associates emotions such as fear with anticipatory judgements, implying that they can possess a cognitive dimension and influence reasoning about future events (Candiotto & Renaut 2020: 4-7).

By contrast, *eros* – the love of the Good – has long been seen as more than a mere perturbation arising from the lower parts of the soul. As Robert Solomon points out, "Plato's discussion of eros as the love of the Good in his dialogue *The Symposium* [demonstrates that] there are emotions involved in reason as well" (Solomon 2008: 4). *Eros* is not simply a disruptive force; it is integral to the philosophical journey, motivating individuals to aspire to higher ideals.

Viewed in this light, at least some emotions such as fear and love are not inherently antagonistic to reason, but can instead be essential to the pursuit of wisdom and moral virtue.

Emotions in Kant's account of moral reasoning

In his investigation of an *a priori* principle that guides rational agents' (moral) decisions and actions, Kant briefly considers, and then argues against, emotions as a potential foundation for providing an adequate basis for the objective laws of practical reason. Kant rejects the possibility of grounding practical reason in emotions on the basis that emotions are "empirical principles", with empirical weaknesses such as a lack of stability and apparent differences from one rational being to another (Tappolet 2022: 11). Emotions are neither seen as a good candidate to ground

morality nor do they seem to be compatible with the candidate that Kant chooses: the Categorical Imperative (Engstrom 2010). Moral actions are moral because they are done from duty to the moral laws. “From duty” is seen as one of the three possible *motivations* that guide human beings’ actions. As articulated by Korsgaard, you can perform an action from duty, from immediate inclinations, or “because you are “impelled to through another inclination,” that is, as a means to some further end (AK 4:397)”⁵⁰ (Korsgaard 1998: viii). From these three types of motivations, only actions performed from duty qualify as moral actions and are the instantiation of a free and autonomous will. Emotions (inclinations/desires), recognised as the second type of motivational states, can motivate an agent’s actions, but these actions *cannot* be moral. In addition to arguing that emotions cannot ground moral actions, Kant is clear that they *need not* participate in the decision-making process that leads to them: “an action from duty is to put aside entirely the influence of inclination and with it every object of the will ... there is left for the will nothing that could determine it except objectively the law and subjectively pure respect for this practical law” (G 400:01, in Guyer 2010: 130).⁵¹

Yet, Kant acknowledges the difficulty this requirement poses for moral agency and moral action, particularly through his concept of the “feeling of respect.” Although inclinations should be excluded from the reasoning process that leads to moral action, Kant repeatedly affirms the fundamental role of the subjective experience of a “pure respect” for the moral law (Herman 1993). What appears most troubling for Kant, therefore, is not the presence of emotions per se, but their potential *motivational force*. Barbara Herman, for instance, argues that moral emotions – such as respect – are not incidental but essential to Kantian moral agency (Herman 1993). This feeling of pure respect for the moral law leads Kant to recognise that, even if inclinations should play no role in the determination of duty and related actions, at the very least a feeling of respect remains necessary for the agent to grasp and *act* upon the demands of morality. Lawrence Blum (1982) also argues in favour of reconsidering the motivational power of some emotions in Kantian ethics. Specifically, altruistic emotions should be distinguished from the inclinations and changeable moods acknowledged by Kant, as they are “capable of the strength and reliability which the Kantian demands of moral motivation” (Blum 1982: 4) Blum argues for the recognition of the substantial role of sympathy, compassion, concern, and care in morality and moral motivation, claiming that these emotions possess the required capacity to counteract inclinations and personal feelings. In doing so, the original Kantian view of altruistic emotions

⁵⁰ By using “AK 4”, Korsgaard follows the German standard edition of Kant’s works, specifically referring to Kant’s *Grundlegung zur Metaphysik der Sitten* (Groundwork for the Metaphysics of Morals).

⁵¹ By using “G 400:01” Guyer uses a different abbreviation terminology but also refers to the Groundwork.

is significantly undermined and effectively replaced – though the formal structure of Kantian morality is retained (Blum 1982: 38-42).

This is evidence that even in some of the most rationalist philosophical psychologies of the mind, it has been extremely difficult, and, as I argue here, even impossible, to fully justify moral reasoning without sooner or later admitting to the intrinsic role of emotions in its fulfilment.

3.3.2. Emotions as essential to morality: Virtue ethics and sentimentalism

Plato's and Kant's wariness towards the potentially manipulative aspects of emotions has significantly constrained their accounts of emotions, particularly in terms of the nature, roles, and functions of emotions in morality. Thus, it is essential to turn to those ethical theories that have treated emotion as constitutive of moral understanding, in order to gain a richer and more comprehensive picture of the role and diverse functions of emotion in moral life.

Emotions in Aristotle's account of moral reasoning

As John M. Cooper (1999) noted, Aristotle's theory of virtue is a direct result of a coherent progression from Socratic rationalism – where emotions have little place – through the Platonist theory of mind – where emotions are discussed but mostly seen as non-rational and misleading, to Aristotle's account, where they are given a key role. Aristotle's theory of virtue is interesting because it embraces a view of emotions in which they are in an 'interfusion' with reason (Cooper 1999: 119). On the notion of control more particularly, Aristotle asserts that emotions are something that happens to rational agents, without any control or choice in the matter. Yet he argues that rational agents have control over their emotions once they have arisen. For him, the fact that one gets angry about x when x happens is indeed outside of one's control (Aristotle 1925: *NE* 1 105b28 and 1 06b9). However, one can *decide* whether one stays angry at x or not. In this sense, the emotions that are felt by the agent are as much of a choice as the actions performed in a situation. Thus, to be virtuous or not depends on the agent's actions and emotions. Feeling the right emotions and doing the right actions in a situation is what virtue is about (Stark 2001: 441). Exercising oneself to feel the right emotion, in its appropriate intensity and at the right moment, is essential to enable one to make the right decision. Appropriate emotions in a given situation guide the agent's decision by giving the appropriate degree of attention to the elements at play in that situation.

This is more complex than it might appear, as emotions can, and should in some contexts, be *multiple*. Hence, as noted by Stark, the courageous person who stands confident in battle where appropriate needs also to “fully and clearly appreciate the potential harm of proceeding into battle”, thus she needs to also feel fear (Stark 2001: 441). If the potential harm were staggering, the right decision would not be to stand and fight but to fear and flee, thus the appropriate virtue would be prudence or caution instead of courage (Stark 2001: 441).⁵²

Stark’s insistence on Aristotle’s recognition of the possibility – and the necessity – of experiencing a multiplicity of emotions in a given situation as a condition for appropriate decision-making highlights both the richness of affective life and the complexity of its relation to morality. Aristotle’s account treats emotions as subject to rational assessment; accordingly, emotions can be judged as appropriate or inappropriate, in terms of both their intensity and their object-directedness. The virtuous person can evaluate the appropriateness of her own emotions, and an external observer may likewise assess the fittingness of another’s emotional response.

Emotions in Hume’s account of moral reasoning

The classical rationalist view of emotions has been challenged by the antithetical view, which portrays emotions as dominating reason. In his work, Hume addresses the motivational foundations of moral action and consequently rejects the rationalist claim that morality is grounded in reason alone, famously asserting that “reason is, and ought only to be the slave of the passions” (Hume [1739–1740] 2000, in Solomon 1977: 41). Hume thus makes a significant move by inverting the traditional hierarchy of reason and emotion established by the rationalists.

Although he acknowledges that emotions can be dangerous and mislead one in one’s judgement of a situation, the key problem of the rationalist view is its failure to recognise the complexity of the emotional life. Hume posits a clear distinction between two types of emotions: “violent passions and calm desires and tendencies” (Mackie 1980: 46). In his view, rationalists have confused reason with the latter type (calm desires and tendencies), because these emotions, unlike the violent ones, produce almost no immediate sensations in the mind. Hence, apparently unaware of their existence, rationalists have mistakenly opposed reason alone to violent passions, when they should have opposed calm desires and tendencies (alongside reason) to violent passions (Mackie 1980: 46–47). In doing so, rationalists have also erroneously claimed

⁵² This is what distinguishes the courageous person from the rash one on Stark’s account (Stark 2001: 441).

that reason is the source of morality, when in fact its source is to be found in the emotions. The main reason why reason cannot be the source of morality is that reason alone cannot motivate action. For Hume, *reason* is a cognitive faculty concerned with the discovery of truth, the comparison of ideas, and the inference of causal relations (Hume [1739-1740] in Solomon 1977). Hume denies that reason has any motivational power of its own. It can direct us by informing us of the means to achieve our ends, but it cannot supply those ends or move us to act. The idea that emotions are necessary because reason alone cannot motivate the will to action is interestingly very close to the reinterpretation that Barbara Herman (1993) offers of Kantian morality, with the exception that in the latter ‘emotions’ are reduced to one specifically, namely, the feeling of respect for the moral law. But this nonetheless echoes a key point in the study of emotions, which is their motivational powers, which seem to ultimately be necessary for (moral) actions.

Following Hume’s account, emotions of the second type (calm desires and tendencies) possess *evaluative properties*. Their evaluative properties are what guide people’s actions and decisions by expressing approval or disapproval of a particular situation or individual. Approval and disapproval are kinds of pleasure and displeasure which one feels when assessing people’s behaviours and/or one’s own actions (Tappolet 2022). As Tappolet notes, these feelings are mediated by what Hume refers to as the mechanism of sympathy: “Given this mechanism, what happens when we consider a trait that tends to benefit others, such as benevolence, is that we tend to feel pleasure, thereby approving that trait” (Tappolet 2022: 10-11). Hume’s interest in sympathy as a human ability to *understand* the feelings and beliefs of others is not derived from pure reasoning but from a more phenomenological form of apprehension, in which the senses provide *information* unavailable through *a priori* reasoning.⁵³ In addition to being the mechanism by which we form moral judgements – and thus the basis of his moral theory – sympathy also enables socialisation and, to some extent, civilisation. Indeed, Hume roots justice in these natural feelings of approval (pleasure) and disapproval (displeasure), though he grants reason an important role in connecting them to “the social effects of the system of justice” (Mackie 1980: 85). Through the mechanism of sympathy, for instance, we can feel approval for the interests of society, and through reason we can understand that justice tends towards the common good. Emotions, therefore, are a significant source of knowledge, in that they can illuminate one’s preferred ends and goals (Tappolet 2022: 9). As Collier (2011) points out, Hume’s account of

⁵³ This also creates an interesting parallel with Blume’s modified account of Kant’s morality.

emotions (or feelings) significantly influenced today's contemporary theories of emotion rooted in sentimentalism and the psychological feeling tradition, including appraisal theory.

3.3.3. Contemporary accounts of emotions in moral reasoning

Contemporary accounts of moral reasoning rarely deny the role of emotions, though some assign them a more implicit instrumental role while others regard them as intrinsic to moral judgement and moral actions. The general idea that emotions provide moral salience⁵⁴ has been endorsed in different philosophical traditions, including sentimentalism (Greenspan 1988; D'Arms & Jacobson 2006), cognitivism (de Sousa 1987; Nussbaum 2001), phenomenology (Scheler [1923]1970; Stein [1917]1964), virtue ethics (Hursthouse 1999; Vallor 2016) and hermeneutics (Ricoeur 1990). Hence, considerable changes have been observed in philosophical work, requiring us to draw a broader and more complex picture of the full range of emotions that appear to underpin our most rational judgements and guide us in the pursuit of long-term goals. I will only say a few words about them here, as I will discuss them further in the next section.

Within the broad view that emotions contribute to morality, we can distinguish between holistic accounts – which typically investigate how moral emotions play an intrinsic role in the formation of moral judgement and action – and emotion-specific accounts, which concentrate on one or a few individual emotions. Emotions such as guilt, shame, pride, compassion, empathy, remorse, fear, and anger are among the most frequently discussed in moral theories of emotion, as their presence is often a key indicator of the presence of genuine moral dilemmas or normatively complex situations. Below I focus mostly on holistic accounts as I first want to draw a general view of emotions.

On the holistic side, Martha Nussbaum is one of the most prominent contemporary philosophers regarding the study of emotions in moral reasoning and judgement. Her hybrid account, between neo-Aristotelian and cognitivist appraisal theory, describes emotions as evaluative judgements, which “shape the landscape of our mental and social lives” (Nussbaum

⁵⁴ I understand “moral salience” to refer to an agent's ability to recognise which features of a situation are morally relevant. In philosophy of emotion, that ability is intertwined with the nature of emotions and hence it is seen as emotions capacity to render salient the moral features of our lives in given situations (e.g. feeling anger when seeing injustice). On the concept of moral salience *tout court*, see notably Barbara Herman (1993). On the concept of salience as a property of emotions that enable them to appraise relevant moral features of a situation, see notably Martha Nussbaum (2001).

2001: 1). She claims that emotions are judgements of value, and as such, are deeply intertwined with practical reasoning. But her account goes further by attributing an eudaimonistic direction to these judgements: “emotions appear to be eudaimonistic, that is, concerned with the person’s flourishing” (Nussbaum 2001: 31). Christine Tappolet (2017) also argues for the necessary role of emotions in reasoning, claiming that emotions are epistemic touchstones that allow us to perceive values, hence they help us track practical reasons. Without denying that they may be defeatable, she argues that they nonetheless “confer *prima facie* justification to evaluative beliefs” (Tappolet 2017: 168). Hence, an agent who meets Tappolet’s conditions – namely, possessing appropriate epistemic and practical dispositions – may act on the basis of her emotions and legitimately claim to be acting in accordance with reason (Tappolet 2017: 194). Drawing heavily on Aristotle’s virtue ethics, Rosalind Hursthouse identifies emotions as central to moral agency (Hursthouse 2000). On her account, emotions are moral responses to situations, which will shape what the agent takes to be relevant in a given context. Finally, Ben-Ze’ev (2000) and Blum (1982)⁵⁵ have both addressed the important question of the ‘partial’ character of emotions,⁵⁶ and defended the idea that this property of emotions is, in fact, *essential* to distinguish morally relevant features in specific contexts. In interpersonal situations, being partial instead of impartial is necessary to communicate but also to perceive other people’s moral values and concerns.

As we will see below, philosophical accounts of emotions continue to exert a substantial influence on contemporary theories in psychology and affective neuroscience. Sentimentalist traditions, which locate the roots of moral agency in emotional experience, find clear echoes in research demonstrating the centrality of affect in moral judgement, empathy, affiliation, and pro-social motivation (Haidt 2001; Damasio 1994). Cognitivist theories, which conceptualise emotions as evaluative judgements or appraisals (e.g. Nussbaum 2001; Solomon 1976), directly inform modern appraisal theories in psychology (Lazarus 1991; Scherer 2005) and have shaped neuroscientific research showing that emotional responses depend on higher-order interpretation rather than mere physiological reaction (Scherer 2009). These approaches contribute to the now-standard view that emotions track meaning and are modulated by goals, beliefs, and context. At the same time, virtue-theoretic perspectives, which treat emotions as dispositions central to moral character and practical wisdom, anticipate contemporary findings in developmental and social psychology on emotional learning, habituation, and regulation

⁵⁵ See Blum (1982) for a more emotion-specific account focusing on altruistic emotions in morality.

⁵⁶ The idea that there is no such thing as an ‘objective’ emotion.

(Gross 1998), as well as neurobiological accounts of emotional plasticity and (affective) social scaffolding (Singer & Klimecki 2014).

3.4. Integrative perspective on emotions

In this section, I analyse three general perspectives on emotions, which I consider to provide a broad overview of the work done on emotions, particularly in terms of the primary assumptions that have driven the field of affective computing.⁵⁷ These three general perspectives are 1) the biological/evolutionary view, 2) the cognitivist view, and 3) the social constructionist view.⁵⁸ I see these three general perspectives as three different ways to approach the same group of phenomena (viz. emotions). For the present work, the tensions among these perspectives are less interesting than the agreements they share. Each perspective brings to our attention a set of important features of emotions, which are not necessarily incompatible with the other perspectives if we do not endorse some of the fundamental assumptions of each individual perspectives. Perhaps more importantly, together, they provide us with an integrative understanding of the relations between emotions and their expressions. Below I present a brief summary of the core tenets of each perspective.

3.4.1. The biological/evolutionary perspective

The biological and evolutionary approach to emotion finds its classical articulation in Darwin's work *The Expression of the Emotions in Man and Animals* ([1872]1965).⁵⁹ On this view, emotions are understood as biological products shaped by millennia of evolutionary constraints, serving functions related to both survival and communication. Hence, emotions are biological responses to adaptive challenges across species.

Proponents of this perspective typically endorse the idea that emotions are independent of, and often precede, cognition, in contrast to the cognitive perspective (see Zajonc 1980, 1984). Accordingly, they posit a set of primary or basic emotions, conceived as automatic responses

⁵⁷ Other ways of grouping emotion accounts have been proposed. For example, the threefold division, which divides accounts into 1) feeling accounts, 2) appraisal accounts, and 3) motivational accounts, has received considerable support (see Solomon 1977; Scarantino 2020; Tappolet 2022). However, here I am following Cornelius (1996) division, as it is better adapted to also include accounts dedicated specifically to emotional expressions rather than only emotions, which will be useful for linking this section with the subsequent ones.

⁵⁸ Note that some of the authors mentioned in the different perspectives most of the time endorse a more complex and nuanced view than the general one defined here.

⁵⁹ On Darwin's view, emotional expressions have been finely tuned to meet our evolutionary needs for both expressing our inner affective states (see the *principle of serviceable habit*) and communicating with others (see the *principle of antithesis*) (Darwin [1872]1965).

elicited by situational changes (or absence thereof). This perspective received unparalleled popularity following Paul Ekman's claim (1992) that the basic emotions he identifies, namely fear, anger, sadness, joy, disgust, and surprise, are universally recognisable and innately expressed through distinctive physiological patterns and facial expressions (see Chapter 2, §2.2.1). Following the basic emotion theory, emotions are considered functionally linked to specific action tendencies (e.g. fear prepares for flight, anger for confrontation), and are often associated with characteristic patterns of autonomic activity such as increased heart rate, changes in skin conductance, or muscular tension, though with recognised variability across contexts and individuals (Ekman 1992; Izard 1971). These physiological changes are thus treated not as incidental by-products of emotions but as integral components of the emotional response system, preparing the body to act in ways that have been proven evolutionarily advantageous.

Proponents of this view differ in their commitments. Strong versions assert that emotions are innate mental states, universally expressed, and biologically predetermined. Within these strong variants, some theorists hold that physiological changes are inherent parts of the emotion itself (Ekman 1992), while others treat them as components of the emotional system rather than mere outward behavioural expressions (Izard 1971). Weaker variants maintain that while the underlying physiological mechanisms may be species-wide, their expressive manifestations are modulated by learning and context. Thus, the weaker accounts will typically state that, while emotions are grounded in biology, their expressions can be shaped by environmental and cultural variables in functionally adaptive ways (Griffiths 1997). To get around these social difficulties, stronger versions, such as Ekman's theory of emotion, will develop notions such as the concept of display rules, which are cultural norms that regulate the 'surface' of emotional expressions (Ekman & Friesen 1969). Display rules aim to account for the observation that emotional expressions are often suppressed, exaggerated, or altered depending on social context. Yet, Ekman (1992) maintains that one can always see the "genuine" emotion if one knows what type of micro-expressions to look for in a given prototypical emotion-type.

As mentioned in the preceding chapter, this biological and evolutionary framework has significantly influenced the field of affective computing, particularly in efforts to design systems that can recognise, model, or simulate human emotions. The FAC system, inspired by the work

of Carl-Herman Hjortsjö⁶⁰ and developed by Ekman and Friesen (1976, 1978), is still in use today in state-of-the-art scientific research in affective computing.⁶¹

In sum, the biological/evolutionary perspective often emphasises the innateness, universality, and (necessary) physiological embodiment of emotions. While its early formulations focused on observable expressions, contemporary variants integrate neurophysiological evidence to support the view that emotions are complex, biologically embedded states with deep roots in evolutionary history. Contemporary affective neuroscience has deepened our understanding of the biological basis of emotion by identifying specific brain structures, pathways, and bodily systems involved in emotional processing. LeDoux (1996), comparing the amygdala to the hub of a wheel, demonstrated its involvement in the appraisal of emotional meaning by showing how fear responses are coordinated by the amygdala and subcortical circuits. Panksepp (1998)⁶² identified primary affective systems in mammals that govern emotional behaviours across species, attempting to bridge the gap between the mammalian brain and the human mind.

But this perspective has been challenged, most notably by Barrett (2017) and Barrett and Satpute (2019) who, following a constructivist framework, argue that emotions are not discrete neural packages, but dynamic and context-sensitive events constructed by the brain through predictive processes, emotional regulation, and concept-based inference. This view shifts the emphasis of the biological/evolutionary perspective from hard-wired and fixed emotion circuits to flexible brain networks that support embodied inference.

Hence, despite sustained critiques of the basic emotion theory, the biological/evolutionary perspective as a whole, and particularly its contemporary developments, remains foundational, offering indispensable insights into the biological, embodied, and adaptive functions of emotion.

⁶⁰ See Wagner, H. (1997) for a detailed explanation of how the work of Ekman & Friesen took inspiration from Hjortsjö, C. H. (1969), including the development of FACS (Wagner 1997: 34-35). See also on this point Eibl-Eibesfeldt (1989).

⁶¹ See Chapter 2. iMotions, for instance, explicitly states its use of FACS. The assumption that emotional expressions have identifiable physiological patterns – such as facial configurations, vocal prosody, or biometric data – underpins many, if not most, contemporary computational approaches to emotion detection and human-machine interaction.

⁶² See also Montag & Panksepp (2016).

3.4.2. The cognitivist perspective

Cognitivist theories of emotion share the foundational view that emotions are intentional, meaning-laden responses to states of affairs in the world. Unlike biological/evolutionary perspectives, which conceive of emotions as hard-wired mechanisms (e.g. Ekman 1992; Panksepp 1998), cognitivist theories treat emotions as cognitive engagements with value – albeit in varied forms. Within the cognitivist perspective, scholars diverge on whether emotions are best understood as judgements, appraisals, perceptions, or narrative structures, but all agree that emotions are not reducible to physiological states or behavioural reflexes. Instead, they are intentional phenomena, responsive to reasons and assessable in terms of rationality or fittingness.

The judgement theories have played a major role in overturning the traditional philosophical idea that emotions were irrational (or arrational). Philosophers in the classical rationalist tradition were notably wary of the negative effects that emotions could have on reason, and generally looked for ways to suppress them from the identified components that led to rational judgement (Solomon 1976). The judgement theories may be conceived as the most classical (or radical) form of cognitivism, directly opposing the classical rationalist assumptions. They hold that emotions have evaluative properties through which it is possible to process certain types of information and yield value-laden judgements (Solomon 2003; Green 1992; Goldie 2000; Nussbaum 2001). According to Solomon (1976, 2003), an emotion such as anger is not simply a feeling but a complex evaluative stance: a judgement that one has been wronged and that this matters. Although not all judgements are emotions, emotions themselves are evaluative judgements; they enable individuals to interpret and navigate morally salient situations.

My anger *is* that set of judgements. Similarly, my embarrassment is my judgement to the effect that I am in an exceedingly awkward situation. My shame *is* my judgement to the effect that I am responsible for an untoward situation or incident. My sadness, my sorrow, and my grief *are* judgements of various severity to the effect that I have suffered a loss. An emotion is an evaluative (or a “normative”) judgement, a judgement about my situation and about myself and/or about all other people (Solomon 1976: 126, author’s original italics).

Hence, for Solomon, the judgement constitutes the emotion itself. Similarly, Nussbaum (2001) argues that emotions are evaluative judgements that reveal the importance of external events to

one's flourishing. Thus, emotions are more precisely *judgements of value*, making them a type of (evaluative) belief.⁶³

Alongside judgement theories, we can find appraisal theories as another subtype of cognitivism, although slightly more moderate than their counterparts in their claim of emotions as evaluative phenomena. Appraisal theories do not hold that emotions may yield *explicit* judgement, but rather, that their nature is evaluative. Klaus Scherer's Component Process Model (1984, 2001), for example, formalised this idea by considering emotions as complex processes rather than events. Through a combination of cognitive appraisal and physiological responses, Scherer's model conceptualises emotions as "the interface between an organism and its environment mediating between constantly changing situations and events and the individual's behavioral responses" (Scherer 1984: 295). His theory emphasises the dynamic and interactive nature of emotional experience. Phoebe Ellsworth and Klaus Scherer (2003) have further refined the appraisal model by proposing to recognise *dimensional appraisal variables*, such as control, agency, and certainty, which differentiate emotions like fear, anger, and guilt. Under the appraisal model, appraisals may be considered as either antecedents of emotions or components, with the latter being retained as the most probable (Ellsworth & Scherer 2003: 575). Richard Lazarus (1991) proposes that emotions need not involve full-blown conceptual thought but rather involve appraisals of events in terms of their relevance to the individual's goals and coping capacities. One of Lazarus' notable contributions to the field is his distinction between primary appraisal and secondary appraisal. Emotions evaluate the significance of a situation for our well-being (primary appraisal), but they are also an evaluation of our potential to cope with the new situation (secondary appraisal). Thus, in the secondary appraisal, there is an element of anticipation, where the appraisal is actively projecting the individual into assessing the probability of potential outcomes. In his secondary appraisal component for compassion, Lazarus notes

If there is self-directed blame, then the probability of guilt is increased; if blame is directed at the victim, then the probability of anger is increased; if there is no blame, then compassion is likely, depending on the dispositional characteristics of the person. (Lazarus 1991: 290).

This not only shows the forward-looking component in secondary appraisal but also the interrelations between emotions and formal types of events, and between a person's dispositional characteristics and their emotions. These models integrate physiological, cognitive, and motivational components. While appraisal theorists share with judgement theorists the idea

⁶³ Even though it should be noted that in Nussbaum's account emotions also have an affective (feeling) component and thus are not *purely* a type of belief.

that emotions are about value, they are generally more open to non-conscious forms of evaluation, including embodied affective processing.

Within the cognitivist perspective, narrative and perceptual theories complete the picture by offering further refinements. On the narrative side, Keith Oatley (1992, 2004) suggests that emotions are simulated narratives that help individuals understand and navigate social interactions. Emotions, on this view, are like cognitive “mini-dramas” that reveal what matters to the agent in a given situation, guiding attention, inference, and behaviour. Ronald de Sousa (1987) offers a closely related model in his theory of “paradigm scenarios”. These socially and culturally learned scenarios shape emotional perception and response by linking types of situations with patterns of evaluation and action. For de Sousa, emotions are rational insofar as they are formed within coherent repertoires and respond fittingly to evaluative features of the world.⁶⁴ Emotions are thus embedded in the narrative structure of consciousness, linking past experience, present meaning, and future action.

A yet different group of theories is proposed by the perceptual account. Tappolet’s theory of emotions, for instance, postulates that emotions *represent* the world as being a certain way. However, unlike judgement or appraisal theories, perceptual theories argue that the content of emotions is not conceptual but *representational*: emotions do not evaluate or appraise events, people, or states of affairs, but rather perceive *value properties* in a way analogous to sense perception. Emotions, like perceptual experiences, are non-conceptual but representational: they present the world as dangerous, offensive, admirable, or shameful. Emotions, on this view, are not conclusions of reasoning but rather quasi-perceptual modes of awareness, capable of being accurate or inaccurate in how they track evaluative facts.

Despite their differences, they all defend the view that emotions are structured engagements with meaning. Whether as judgements, appraisals, narratives, or perceptual tools, emotions are treated as cognitively structured, rationally assessable, and responsive to normative considerations. On this perspective, emotions process information from situations or events and attribute meaning to it. This ascribed meaning is seen as essential for navigating daily situations, by guiding one’s decisions and actions towards goals that matter to them. Together,

⁶⁴ The “evaluative features of the world” are those aspects of a situation that *matter* to an agent, in terms of personal significance, practical relevance, or moral or social meaning.

these cognitivist theories provide a rich and flexible account of emotions that accommodates their intentionality, complexity, social embeddedness, and evaluative character.

3.4.3. The social constructionist perspective

The social constructionist perspective holds the idea that emotions are designed to serve social roles and purposes (Averill 1980, 1982; Armon-Jones 1985; Parkinson, Fischer & Manstead 2005). One of its central points is that emotions (and emotional expressions) are learned rather than innate (Harré 1986; Lively & Weed 2016). Catherine Lutz, for instance, argues that the very existence of emotions could be situated at the cultural level (Lutz 1988).⁶⁵ This claim stems from the observation that emotions are subject to fundamental cultural variability, which results in some emotion-types being present in some cultures and not in others (Averill 1982; Wierzbicka 1999; Tcherkassof 2018). Anthropological studies conducted among the Utkuhikhalingmiut in northern Canada have revealed that some communities lack a concept corresponding to one of the so-called “basic emotions” – namely, anger (Briggs 1971). As a result, they are allegedly unfamiliar with the emotion itself. This directly challenges the claims of the biological and evolutionary accounts of emotion. Moreover, the social constructionist perspective has assumed a powerful interplay between emotions, cultures, and languages. For instance, the English language does not have an equivalent for some emotions that are present in other cultures such as “*dépaysement*” in French or “*Schadenfreude*” in German (Russell 1991). Kundera ([1978]1987) supposes that the word (and corresponding emotion) “*Litost*” in Czech, has “no exact translation in any other language” (Kundera [1978]1987 in Russell 1991: 426).

Extending this perspective through a neurobiological foundation, Singer and Klimecki’s work (2014) on empathy and compassion has supported the view that social emotions are not only evolutionarily adaptive but also deeply embedded in brain systems that facilitate social cohesion and moral concern. In particular, their work brings empirical evidence that it is possible to transform social emotions through training: “This finding underlines the malleability of social emotions as it shows that a short-term compassion training of several days can foster positive feelings and related brain activations, even when persons are exposed to the distress of others.” (Singer & Klimecki 2014: 877). Hence, it seems that societies, through culture and social norms, at least partially shape the landscape of our emotions.

⁶⁵ Additional claims have been made in this view besides the central role of culture, one particularly interesting being gender variability (Fischer 2000).

Set out in this way, these views may appear irreconcilable.⁶⁶ However, if we relax some of their stronger claims, they need not be. In the next section, I adopt a function-first integrative perspective on emotions, grounded in core agreements between these different perspectives, to further investigate ERTs' potential reach into the affective life.

3.5. Adoption of a function-first integrative perspective on emotions

The preceding sections show that no single philosophical or scientific perspective can, on its own, account for the complexity of emotional life. Each, however, isolates a genuine dimension that ought to be retained. I therefore adopt an integrative and functional perspective on emotions: a methodological approach that explains emotions in terms of the functions they fulfil in lived practice (particularly in practical reasoning, social coordination, and normative life) while integrating transdisciplinary insights from their embodied, evaluative, and social dimensions. This approach aligns with contemporary function-first and hybrid models in philosophy and affective science (e.g. de Sousa 1987; Frijda 1986; Griffiths 1997; Prinz 2004; Scarantino 2017), which likewise combine biological, cognitive-evaluative, and social analyses to render emotions intelligible in terms of action and interaction.⁶⁷ In this dissertation, however, theories of emotion function as explanatory lenses, assessed by how well they illuminate what emotions do in practice. The function-first integrative perspective adopted here is thus selective in the following sense: it retains only those theoretical commitments needed to understand emotions as forces that orient attention, confer salience, guide action, and shape social meaning.

Although this perspective could serve as a general approach to emotions, it is adopted here for a more specific reason: to better grasp the distinctive ethical challenges raised by fully functional ERTs. Because ERTs aim to render affective life externally readable, interpretable, and classifiable, a proper ethical evaluation of such technologies requires an understanding of what

⁶⁶ As is the case almost everywhere else in research, specific research foci and ideologies inevitably lead to the rejection or deliberate neglect of certain facts or possibilities. As Dukes et al. (2021) rightly point out, affective sciences as a whole have been rejected (or deliberately ignored) for a long time because scientists “typically neglect what they cannot measure”. This is true of emotions but also of the different perspectives on emotions. Hence, we can observe that, typically, the biological/evolutionary views tend to ignore cognitive-related aspects and reciprocally, cognitivist views typically ignore the biological/evolutionary related components – which does not necessarily mean that they reject them but rather that they *overlook* them (Dukes et al. 2021). This may be seen from two different points of view, either that the result of this is that fundamental theories of emotions seem to have irreconcilable core tenets, or that, by focusing on things so drastically different, they may complement each other. I endorse the latter. (see Green (2007: 108-109), for a comparable stance.)

⁶⁷ As already discussed, other hybrid approaches may be found in the work of Solomon (1976), Nussbaum (2001), Goldie (2000), Ben-Ze’ev (2000), and Armon-Jones (1985), to name only a few.

is at stake when emotions are treated in this way. An integrative perspective on emotions provides us with such understanding.

3.5.1. Transdisciplinary agreements on their nature, role, and functions

Despite ongoing disagreements among biological, cognitivist, and social constructionist perspectives on both the definition and ontological status of emotions (Fehr & Russell 1984; Scarantino & de Sousa 2018), recent work across philosophy, psychology, affective neuroscience, and anthropology has generated many hybrid approaches that try to reconcile these different views. These hybrid approaches have been very helpful to pinpoint where major agreements may be found. These convergences, while framed within different methodological and theoretical commitments, offer a productive basis to better understand both the capacities and contexts in which emotions operate.

There have been many attempts at complete or partial reconciliation between some of the deepest existing tensions between those perspectives. Fridlund's work (1991) notably attempts to reconcile both the biological/evolutionary view with the social constructionist view. De Sousa's account (1987), despite being cognitively oriented, treats emotions as culturally shaped but biologically grounded. Frijda (1986), Griffiths (1997), and Barrett (2017) also developed hybrid approaches that bring together components of the biological/evolutionary perspective with the cognitivist perspective. Frijda's work (1986, 1988), on emotions as states of action readiness, blends both the cognitivist and biological perspective, while also accounting for some key aspects of the social constructionist perspective. Solomon's work, although classically cognitivist at first (1976), later embraces the social constructionist perspective (2003). And Ben-Ze'ev's work (2000) notably brings together aspects of the cognitivist view with the social constructionist one by recognising emotions as grounded in cognitive evaluations while also being deeply shaped by social, cultural, and interpersonal contexts.

Hence, I now turn to some of the main transdisciplinary agreements on the nature, role, and functionalities of emotions, with the function-first integrative perspective and the hybrid approaches in mind.

Agreements on the nature of emotions

Contemporary agreements can be noted across the different disciplines concerned with the study of emotions. Regarding the nature of emotions, most affective researchers agree that emotions are multifaceted, object-directed, and context-sensitive, and that a comprehensive theory of emotions needs to consider both their psychological and social dimensions.

Emotions are complex, multifaceted phenomena that cannot be reduced to a single component or causal mechanism. Most affective researchers agree that emotions involve a constellation of physiological, cognitive, behavioural, and expressive elements. Key components are typically recognised, even though the role, importance, and relations of these components are still debated. In general, affective researchers recognise that emotions have a 1) physiological (e.g. increased heartbeat), 2) phenomenological (e.g. unpleasant feeling), 3) evaluative (e.g. appraising the bear coming your way), 4) expressive (e.g. upper eyelids raised, jaw dropped open, lips stretched horizontally), and 5) behavioural component (e.g. an impulse to flee). From an integrative perspective, emotions are thus best viewed as biologically *grounded*, cognitively *structured*, and socially intelligible modes of affective orientation, *shaped by* and shaping the (cultural) contexts in which they emerge.

Additionally, it is generally agreed that emotions are intentional, object-directed phenomena,⁶⁸ which makes them potentially assessable based on their fittingness and/or appropriateness. In contemporary research, “object-directed” is broadened to also allow emotions to be directed towards people, situations, events, or state of affairs. This object-directedness is often analysed in terms of both particular objects (e.g. the bear, the sister, the rug) and formal objects (e.g. danger, loss, slights) (Kenny 1963; Teroni 2007). Assessment of particular objects is dependent on their relation to a formal object; in other words, the appraisal necessitates some conception of the given object in the formal form to evaluate the emotion’s fittingness. For instance, the reason one feels fear when seeing a bear is because one appraises the bear as something dangerous (Scarantino 2020: 122; Naar 2021). Relatedly, if the fitting objects of fear are dangers, the fitting objects of anger are slights, and those of sadness are losses, and so on (Scarantino 2020: 122).

⁶⁸ This is one particularity that distinguishes emotions from other affective states, such as moods for instance, which have no particular object.

Parallel to this is the agreement on the idea that people may (and likely will) have multiple, simultaneous emotions at a given time (Greenspan 1980; Nussbaum 2001; Fischer, Pauw, & Manstead 2019). Within virtue ethics, emotions are often understood as multiple and co-occurrent in the agent, because this multiplicity is required for the full appraisal of complex situations. Some virtues are attained only through this navigation of the emotional life. Recall the courageous person, who following Stark's analysis, can only be courageous if she feels both the fear of the fearful and the confidence, hope, or feeling of duty (in battle) (Stark 2001: 441). Fischer, Pauw, and Manstead (2019) give another example, where someone receives a diagnosis of a terminal disease. In their scenario, the affected person feels sad about the situation but also angry and guilty, all for different reasons. After analysis, it appears that these emotions seem to be all appropriately directed, and in that sense, fitting, yet they are simultaneously occurring.⁶⁹ The co-occurrence of multiple emotions usually arises when the emoter is hesitant in the face of, or in recognition of, several valuable possible points of view. This complexity is also well aligned with folk intuitions about inner emotions, especially in complex contexts or social interactions.

Agreements on the role of emotions

Among the major agreements on their role is their central part in reasoning processes and decision-making. Contemporary affective researchers consider emotions as the *dominant driver* of most meaningful decisions in life (Frijda 1986; Lazarus 1991; Nussbaum 2001; Keltner et al. 2014), but they are also seen as playing an active role in the formation of judgements and choices on a daily basis (Kahneman & Tversky 1979; Lerner et al. 2015). Emotions are recognised as providing useful information to the self about what is valuable and what should be pursued (or avoided) (Solomon 2003; Frijda 1986; Nussbaum 2001). Hence, there is a broad consensus that emotions play a central role in human life, not merely as internal states but as processes that mediate between agents and their environments. Emotions are recognised as orienting devices: they direct attention, structure perception, and highlight what is salient in a given context. They inform practical reasoning, especially in conditions of uncertainty, where explicit deliberation may be insufficient to resolve value-laden dilemmas. In this respect, emotions function as both affective responses and evaluative guides, enabling agents to navigate complex social and moral landscapes (Frijda 1986; Nussbaum 2001; Tappolet 2017). Crucially, emotions also mediate social interaction. They signal evaluative stances to others, communicate expectations, and invite

⁶⁹ See Fischer, Pauw, & Manstead (2019: 9) for more details on the example.

particular forms of recognition, resistance, or alignment. They are thus implicated not only in individual inner experience but also in the regulation and negotiation of social life.

Agreements on the functions of emotions

In terms of their functions, emotions are increasingly understood as adaptive, epistemic, normative, and social. Studies in neurobiology tell us that, biologically, at least some emotions appear to (have evolved to) serve to regulate the body's responses to significant stimuli, coordinating actions such as fight, flight, or care, and to cope with new (social) situations. Studies in cognitive science tell us that emotions foreground what matters to the agent in a given situation (de Sousa 1987; Tappolet 2017). They also perform an essential normative function, reinforcing social expectations and enabling the affective transmission of values and norms. Guilt, shame, indignation, and pride, for instance, do not merely reflect internal states but operate as mechanisms for social cohesion and moral accountability (Armon-Jones 1985; Ben-Ze'ev 2000; Nussbaum 2001). In addition, emotions are culturally and socially sensitive (Averill 1982; Wierzbicka 1999; Parkinson, Fischer, & Manstead 2005), and have a plastic, developmental dimension: they can be shaped, refined, or redirected through cultural training, interpersonal interaction, and reflective practice (Singer & Klimecki 2014; Barrett 2017). This shows that emotions are not static properties of the mind but complex, multifaceted, dynamic, and to some degree malleable.

The biological, cognitive, and social dimensions of emotional experience are not competing explanations but interdependent lenses for understanding emotions. Taken together, their agreements support a conception of emotion that sees them not as self-contained episodes, or states, but as dynamic processes that mediate between the individual and their world.

3.5.2. Emotions' action tendencies

So far, I have focused on emotions themselves, emphasising agreements on their nature, role, and functions. I now briefly turn to the *relation* between emotions and their related expressions, focusing in particular on the following relations: 1) emotions and action tendencies, 2) emotions and physiological changes, and 3) emotions and *performative* acts (in the sense that they iterate something).

The claim that emotions are action-oriented is not a new one. It is already well known that, for instance, anger is typically linked with aggression, fear with avoidance, happiness with prosocial

behaviour, sadness with reflection or withdrawal, disgust with avoidance or purification, guilt with reparative actions or distancing, and love with nurturing or commitment. These action tendencies of emotions are highly valuable for any emotion recognition task that wants to expand the ‘recognition’ task into a ‘prediction’ task. Understanding that anger commonly entails a disposition towards confrontation or aggression is useful for purposes of both self-preservation and security. Likewise, understanding that sadness often tends to yield social withdrawal, and that love tends to manifest into nurturing or commitment-type action is also useful. Frijda’s theory of action-readiness (1986, 1988) already discusses the ways in which emotions are action-oriented. For him, it is clear that the “primary” or “basic” emotions typically identified by the biological/evolutionary views, namely fear, anger, joy, surprise, disgust, and sadness, yield clearer correlations between specific emotion-type and specific action-type, whereas ‘higher-order emotions’ (Frijda 1988: 351), such as remorse, guilt, or jealousy, may keep the door open for more complex and/or uncertain action-type. In the case of guilt, for instance, narrowing the action-type to “reparative actions” and “distanciation” still leaves open two widely different type of actions. Andrea Scarantino (2017, 2019) similarly makes the case that different emotion types correspond to different forms of action tendencies, however, on his account, these associations are not deterministic as they only represent *probabilistic patterns* of action readiness that are highly context-dependent, and subject to modulation by individual knowledge and other social factors. Barrett’s theory of constructed emotion (Barrett 2017) offers a more nuanced alternative, where emotions function as context-based anticipatory predictions. Emotions infer the meaning of a situation and make sense of it based on prior experience, and from these prepare the body for appropriate action. On this account, emotions proactively navigate uncertain situations.

As Scarantino and Barrett’s accounts highlight, human emotional life rarely conforms to a neat one-to-one mapping between stimulus and response, as context, individual’s social situation, and culture greatly complicate the picture.⁷⁰ Additionally, as the moral theories of emotions have pointed out, a defining feature of emotions in lived experience is their simultaneity and potential for conflict. Emotions are frequently multiple, co-occurring processes that thus generate *competing* action tendencies. For instance, one may feel both guilt and anger towards a loved one, or experience fear and compassion in the face of another’s suffering. Nussbaum (2001), de Sousa (1987), and Greenspan (1980) have also acknowledged this emotional multiplicity, arguing that

⁷⁰ This is to say that, for instance, you may know that you tend to smile when nervous, which would not follow a strict one-to-one “universal” mapping in which smiles correspond to joy (or happiness).

it reflects the layered, value-laden nature of our concerns. Greenspan, in particular, emphasises the idea of emotional ambivalence as both a challenge and a resource for practical reasoning. De Sousa's concept of paradigm scenarios shows how emotional training in culturally salient contexts can generate patterns of response that pull in different directions, especially when new or hybrid situations emerge. These conflicting tendencies reflect the moral complexity of emotional life, especially where multiple values or duties intersect. As a result, emotions often present different tensions, constraints, but also possibilities through which agents must navigate. Emotional multiplicity does not undermine their action-oriented function, but it certainly makes it more complex.

Because emotions are processes of anticipating social outcomes, they are inherently other-responsive and norm-guided: they contribute to the structuring of social life through their expressions. Ben-Ze'ev (2000) argues along those lines when stating that emotions arise from and feed social interactions. Similarly, Fischer, Pauw, and Manstead (2019) have shown how emotions like shame, guilt, and anger function as moral regulators in group dynamics. These emotions not only orient the agent towards a response (e.g. reparation, protest, withdrawal) but also signal value violations or commitments to others once expressed, thereby supporting (or hindering) moral expectations within communities. Hence, emotions and their expressions are viewed as reactions to changes in situations or state of affairs but also as the cause of these changes for others. This means that their related action tendencies should not be assessed in isolation but rather understood as constituting the social context that they help maintain or change.

3.5.3. Emotions and physiological changes

Knowing that emotions are closely tied to what individuals value in the world – and that they shape both every day and life-defining decisions – is crucial for assessing the potential impact of ERTs. If fully functional, such technologies do more than simply detect emotional states; they also reveal a person's underlying values and their sense of what is morally salient.

As previously discussed, emotions also indicate the most probable action tendencies of emoters, at least in some cases. However, such recognition is possible only if there are some ways of *accessing* this type of information. On the question of the relevant indicators that reveal this type of information, we find on one end of the spectrum evolutionary and biological theories which claim that emotions are directly embodied and therefore detectable through physiological

changes. According to this view, specific autonomic changes – such as elevated heart rate or pupil dilation – are not merely correlated with some emotions but constitute an essential part of the emotion itself. One cannot feel a particular emotion type without *exhibiting* the associated physiological changes.

At the other end of the spectrum, theorists like Barrett (2006, 2017) reject the idea of fixed “fingerprints” for emotion types. While she acknowledges that certain prototypical physiological changes may commonly occur with specific prototypical emotions, she emphasises that such correlations are context-sensitive and highly variable. This means that in opposition to essentialist claims, ‘anger’ does not always involve a heightened heart rate, ‘fear’ does not necessarily involve sweaty palms, and ‘embarrassment’ does not always involve blushing.

What does it mean that four meta-analyses, summarizing hundreds of experiments, revealed no consistent, specific fingerprints in the autonomic nervous system for different emotions? It doesn’t mean that emotions are an illusion, or that bodily responses are random. It means that on different occasions, in different contexts, in different studies, within the same individual and across different individuals, *the same emotion category involves different bodily responses*. Variation, not uniformity, is the norm. (Barrett 2017: 15, italics from original author).

Between these poles lies the cluster-based approach. This view maintains that while no single physiological marker maps uniquely onto a particular emotion type, some patterns of changes tend to cluster around specific emotions. Isolating one variable – say, increased heart rate – and interpreting it as a marker of a single emotion type is methodologically flawed, as Barrett’s critique makes clear. Yet researchers have not exhausted the possibility that identifiable clusters of physiological responses may track emotional experiences, simply because not every possible physiological response have been measured in every possible way at the same time. (Scherer 2005; Larsen et al. 2008; Mauss & Robinson 2009).

This hypothesis has motivated a range of empirical efforts to identify which combinations of physiological changes might form meaningful emotion-specific clusters. Most research has, for practical reasons, focused on a specific set of physiological changes that typically include cardiovascular response, respiration rate, skin conductance, and pulse (Kreibig 2010; Zhang et al. 2020); attempting to find relevant clusters among them. More recent studies in affective computing have expanded the range of signals under investigation, including pupil dilation (Kinner et al. 2017), eye movement (Lim, Mountstephen, & Teo 2020), and brain activity patterns such as EEG signals (Q. Wang et al. 2022; Zhang et al. 2020).

However, a central challenge remains: if fear is typically – but not always – associated with increased heart rate, pupil dilation, elevated respiration, and heightened skin conductance, but anger can also be associated with the same cluster of physiological changes, then physiological overlap threatens the discriminability of emotional categories. Addressing this issue, Scherer (2005) argues that the real challenge here is that a comprehensive approach to emotion measurement must assess a constellation of factors:

[...] there is no single gold-standard method for [emotion] measurement. Rather, given the component process nature of the phenomenon, only convergent measurement via assessment of all component changes involved can provide a comprehensive measure of an emotion. In other words, in an ideal world of science, we would need to measure (1) the continuous changes in appraisal processes at all levels of central nervous system processing [...], (2) the response patterns generated in the neuroendocrine, autonomic, and somatic nervous systems, (3) the motivational changes produced by the appraisal results, in particular action tendencies [...], (4) the patterns of facial and vocal expression as well as body movements, and (5) the nature of the subjectively experienced feeling state that reflects all of these component changes. Needless to say, such comprehensive measurement of emotion has never been performed and is unlikely to become standard procedure in the near future. (Scherer 2005: 709).

Without accommodating all of these points, recent advances in affective computing, particularly in multimodal emotion recognition using deep learning techniques, have significantly expanded the range of physiological data that can be gathered and analysed within a single experimental setting. Of particular promise is recent work on physiological signals, which may allow researchers to detect more nuanced physiological clusters previously inaccessible through traditional methods (see Yin et al. 2017). New ML methods, coupled with robust multimodal datasets, might open the door to a real possibility of discovering new ways to classify (and understand) emotions. For instance, it may become possible to differentiate ‘primary-level’ emotions – such as fear in response to an immediate physical threat, as emphasised by biological/evolutionary accounts⁷¹ – from more complex, ‘high-level’ emotions, such as fear of professional failure or of disappointing one’s family through personal choices (Q. Wang et al. 2022).

The capacity to distinguish between these types of emotions could enhance the predictive power of ERTs. Returning to the case of fear, primary-level fear is typically associated with strong avoidance tendencies and immediate action-readiness in the physical domain. In contrast, high-level fears often do not translate into such action tendencies and may even inhibit physical withdrawal. Recognising this distinction in between levels of a given emotion type is thus crucial for any account of emotion that aims to track the relation between felt emotion, action tendency, and social context.

⁷¹ These are often compared to almost ‘reflex-like’ types of emotions.

Moving forward, I keep open the possibility of relying on cluster-based, or other promising, approaches for ERTs. I take emotions to exhibit variability in their physiological expressions, depending on context, individual history, and social meaning. Nevertheless, I maintain that reliable patterns – especially at the level of clustered physiological changes – are likely to exist and can be meaningfully identified. These patterns, while neither universal nor exceptionless, may support robust predictions about the kinds of behaviours and concerns emotions are tied to in particular contexts.⁷² This makes evident the necessity of subjecting ERTs to critical ethical scrutiny, particularly, as I will argue in Chapters 5 and 6, regarding privacy. Tracking the emotional life does not merely amount to registering surface affective states but rather also implies the scrutiny of intimate information about a person’s evaluative stance, motivational tendencies, emotional vulnerabilities, and more, in a way and at a scale that was not even conceivable before. By tracking the affective life, ERTs risk exposing these dimensions of the self that are typically protected from unsolicited access.

3.5.4. Emotions as performative acts

Finally, a last noteworthy relation is found in emotions and performative acts. Some have argued that emotions *are* performative acts in the sense that they *do* things. Solomon argues that “[a]s a general rule, we may say that an emotional expression is action demanded by the logic and ideology of the emotion and particular circumstances” (Solomon 1976: 165). On this view, to express an emotion is to engage in a form of action shaped by the evaluative structure of the emotion itself and the specific context in which it arises. De Sousa advances a complementary account, emphasising the communicative and interpersonal functions of emotional expressions. He maintains that such emotional expressions act as communicative gestures that disclose one’s evaluative stance towards a situation or agent, thereby enabling socially coordinated action and mutual understanding (de Sousa 1987: 191).

External manifestations of emotions are thus embedded in practices of social regulation, negotiation, and recognition. As the next chapter will explore emotional expressions in greater depth, I will only note for now that emotions are not only intrapersonal phenomena but also, and maybe even mostly, interpersonal ones. In close relationships, emotions can appear strikingly transparent, so much so that they often impose themselves with such clarity that denial

⁷² I return to this point in Chapter 4.

or misinterpretation seems implausible. At times, expressions of emotion communicate more powerfully than any concurrent verbal content. When a person's emotion contradicts their spoken words, it can feel as though their verbal message is being actively undermined or invalidated. It is as if they were simultaneously affirming and denying the same claim.

In this sense, emotions, in their expressive form, function as performative acts: their occurrence enacts a specific social meaning or stance, independent of (and sometimes in conflict with) explicit linguistic content. When in contradiction, the asymmetry between emotional and verbal communication introduces a power dynamic that can significantly alter the meaning of the information conveyed and shape the trajectory of an interaction.

As shown throughout this chapter, emotions are best understood as biologically grounded, cognitively structured, and socially intelligible. Yet they do not all serve the same functions, nor can they be analysed at a single level of explanation. Some emotions, such as guilt, shame, indignation, or pride are central to sustaining social cohesion and moral accountability. Others, such as anger, fear, or compassion coordinate more immediate behavioural tendencies. Even within a single emotion-type, such as fear, fine-grained distinctions must sometimes be drawn. This variability reflects that emotions are multifaceted, context-sensitive, and plastic, meaning that they are subject to modulation, training, and strategic social management.⁷³

3.6. Conclusion

In this chapter, I have outlined a function-first and integrative perspective on emotions to better understand the relations between the highly complex and meaning-laden concepts of emotions, values, moral salience, judgements, decisions, and actions. Given the complexity of emotions – rich, multifaceted, and still only partially understood, with no consensus on their definition – I adopted such a function-first integrative view on emotions to prioritise an overview of what emotions do and what relations they maintain over what they are.⁷⁴ My focus has been on examining (i) their foundational role in morality and reasoning processes through diverse readings of classical accounts of morality and reason (Plato, Aristotle, Hume, Kant), and

⁷³ However, this also means that reliable identification will be especially difficult, even with new methods.

⁷⁴ The function-first integrative perspective is methodologically agnostic about the nature of emotions, focusing on functions instead. That said, recognising certain functions may imply that I reject some ontological views on the nature of emotions (for example, the classical rationalist account).

contemporary approaches (Greenspan, Nussbaum, de Sousa, Tappolet, Ben-Ze'ev), (ii) their role in intrapersonal life, and (iii) the functions they serve for both the self and others.

I began by analysing canonical philosophical accounts of emotion and examined their treatment of the relations between emotions and morality. This analysis underscored how the concepts of emotion, moral salience, and values are deeply intertwined and how emotions are an essential part of moral reasoning and moral actions – even within the accounts that explicitly tried to separate emotions from moral reasoning. Building on this, I expanded the discussion to include transdisciplinary insights from contemporary work in philosophy, psychology, affective neuroscience, and anthropology. To structure this broad field, I organised these contributions into three broadly conceived perspectives: biological/evolutionary, cognitivist, and social constructionist. While these perspectives are not always fully reconcilable, many recent accounts adopt hybrid or moderated positions that allow for a more compelling understanding of emotions in functional terms.

I then examined some key characteristics of emotions in relation to their expressions: action tendencies, physiological changes, and performative potential. Emotions, particularly through their anticipatory evaluations of future courses of events, are closely linked to specific action-types. However, identifying or predicting these action types is notoriously difficult, as they vary across individuals, contexts, and other situational variables. Complicating matters further, emotions rarely occur in isolation: they often appear in dynamic, overlapping forms, where the individual experiences and reflects upon multiple emotions at once, thereby blurring potential connections between a single emotion and a clear behavioural outcome. Regarding their related physiological changes, I argued that patterns of clustered physiological changes may exist, at least for certain emotion types – though these will also inevitably vary across individuals and contexts. While current methods fall short of capturing such patterns reliably, future developments may yield robust predictions by relying on them. This possibility makes it all the more urgent to subject ERTs to rigorous epistemological, ethical, and normative scrutiny. Finally, I briefly addressed the relation between emotions and performative acts. Emotions, when expressed, do not merely reveal internal states; they also often iterate something through their expressive correlate, carrying interpersonal meaning, and shaping social reality.

This chapter has focused primarily on the intrapersonal level, examining expressions of emotions only insofar as they relate to emotions themselves. I have thus not yet addressed one

of the most crucial dimensions of expression: their social functions. The relationship between expression and emotion is only the beginning of the story. In the next chapter, I will shift focus to emotional expression proper – moving from the intrapersonal to the interpersonal domain.