HOW DO WE KNOW OUR EMOTIONS?

Richard Moran's Agential Model Applied to Standing Emotions

By

Levan Sikharulidze

Submitted to Department of Philosophy, Central European University

In partial fulfillment of the requirements for the degree of Master of Arts

Supervisor: Professor Katalin Farkas

Vienna, Austria

2024

Acknowledgements

I would like to thank my very first philosophy teacher Mikheil Mirashvili and all others who helped me find the most wonderful way to live.

I would like to express my gratitude towards CEU, and especially, Department of Philosophy. For two years this has been a place where I felt like home, where I grew up as a scholar and as a person, got challenged and inspired, met some amazing people, and collected precious memories for life.

I would like to thank my supervisor, Kati Farkas, who has been very supportive. I got a lot from her as a philosopher, and her advice as a supervisor has been especially helpful for my work.

Finally, I want to thank my family, my brother Ila, and my friends in Vienna or in Georgia: Nano, Zura, Iva, Brylea, Stephen, Darren, Tika, Medea, Tiko, Alex, Gocha, and others. Without them my life wouldn't be that colorful.

Dedication

To my mother who is always there for me

Table of Contents

Introduction

"Do you believe that a person can change radically?" When I answer this question either positively or negatively, it is me who speaks his mind. I communicate what I do or don't believe. Other people take my words the same way – as expressing my thoughts, my beliefs. The way I speak my mind when I say what I believe, what I desire, what I intend, and, as I will aim to argue in this thesis, what I feel as a long-lasting, standing emotion, is taken by others to be furnished with a special kind of authority. This authority, as defenders of agential model of self-knowledge claim, comes from the fact that I am a responsible, rational agent, coming to know my mind by determining my mental states based on justifying reasons I have for or against them. In order for this to be possible, my mental states of believing, desiring, intending must be active in a sense of being reason-responsive.

What I aim to do in my thesis is to inquire about whether a version of agential model of self-knowledge – Richard Moran's agential account – can be applied to a particular type of mental states – standing emotions. I don't aim to defend this theory from various criticisms against it, or to argue for its general advantage over other models, all things considered. What I will do is to analyze whether it has the resources to account for self-knowledge of our standing emotions. In doing this, I cannot avoid discussing how it is applied to beliefs. However, I will not challenge the success of this application. Rather, I will assume that the approach works with beliefs, and I will describe how it does so. I will address the theory's applicability to desires and intentions as well. The comparison between these mental states will help me elucidate my point better. As a result of discussing all these, I will conclude that the model is applicable to standing emotions.

The topic of my thesis is very underexplored. Although one can find literature in modern analytic philosophy both on agential accounts of self-knowledge, on the one hand, and scarce, but still some material on standing emotions, on the other, to my knowledge, there is no account arguing for or against the applicability of Moran's agential model to standing emotions. Thus, I am expecting for my thesis to make an original contribution to the debate.

As I described above, there are two main topics for my thesis: Richard Moran's agential account of self-knowledge and the phenomenon of standing emotions. In order to argue for the applicability of the former to the latter, I will describe and analyze both in turn.

In Chapter One, I will start with accounting for the phenomenon of self-knowledge, in general, and the main approaches towards its explanation in the literature. I will show the motivation of agential view for diverting from other existing theories. I will describe the main principles of agential model, and, consequently, I will account in detail for Moran's version of the view. This chapter will provide readers with all the important knowledge about the account, the applicability of which to standing emotions is under consideration.

In Chapter Two, I will give a picture of the state of art in philosophy of emotions, describing several main trends and distinctions, relevant for my purposes. Then, I will account for the phenomenon of standing emotions, explaining its main features and characteristics, situating it within the broader context. This chapter will give readers the necessary knowledge about standing emotions, demonstrating that the nature of these mental states allows them to be a good candidate to be applied Moran's agential model.

In Chapter Three, after having accounted for both Moran's account of self-knowledge and the phenomenon of standing emotions, I will answer the main question of my thesis – whether Moran's agential model can be applied to standing emotions. In order to do that, first, I will

specify that the relevant kind of agency under consideration is rational agency. Then, I will discuss whether we can exercise rational agency with respect to standing emotions. Consequently, I will address a challenge concerning the Transparency Condition – one of the main constituents of Moran's account. By discussing the applicability of the Transparency Condition, I will thereby inquire what kind of reasons standing emotions respond to. As a result of discussing all these essential issues, I will conclude that Moran's agential model of self-knowledge can be applied to standing emotions.

Chapter One:

Self-Knowledge and First-Person Authority

1.1 Accounts of Self-Knowledge

Knowledge of one's own mental states is standardly called "self-knowledge" (Gertler, 2021). Since Descartes, many philosophers believed that self-knowledge was a special kind of knowledge, significantly different from knowledge of the external world, including the knowledge of other people's mental states. What exactly are the special characteristics of self-knowledge, however, is a matter of philosophical disagreement. There are a number of competing theories trying to account for allegedly essential features of self-knowledge.

Brie Gertler distinguishes four different ways in the philosophical literature to account for the distinctiveness of self-knowledge:

1. Self-knowledge is especially secure, epistemically.
2. Self-knowledge is (sometimes) acquired by use of an exclusively first-personal method.
3. Our capacity for self-knowledge reflects our cognitive agency.
4. One's pronouncements about one's own mental states carry a special authority or presumption of truth (Gertler, 2021, pp. 1-2).

As she claims, only statements (1) and (2) emphasize some kind of "privileged access" to our own mental states, and account for a distinctive feature of self-knowledge as purely epistemic. On the other hand, the defenders of the statement (3), although rejecting purely epistemic approach, don't deny the existence of special epistemic relation (Gertler, 2021). However, what

is important is that they ground the latter in an agential aspect, some of the authors introducing normative concepts, such as responsibility or rationality.

Statement (4) seems to express the most commonly shared belief, permeating our everyday practice, that in normal circumstances we grant each other a special kind of authority with respect to our self-ascriptions of mental states. For instance, when I say that I am sad, in normal circumstances, my word is taken to be true. Which terms are used to describe this practice or assumption often shows which aspects are emphasized. For example, some philosophers describe it as "a default hypothesis of correctness" (McGeer, 2008, p. 81) which shows that they are primarily interested in epistemic security of self-knowledge. Sometimes it is described as "epistemic privilege" or "first-person privilege" (Gertler, 2021, p. 5) that emphasizes the fact that we acknowledge the existence of epistemic asymmetry, or it may be referring to some kind of special method of access. In any case, what is important for our purposes is that the above-mentioned common assumption about a special kind of authority is a feature that most theories of self-knowledge aim to explain. Although, it is unclear what constitutes the first-person authority, it is not my aim here to answer the latter question.

Arguably, every account of self-knowledge handles and explains at least some (kind of) mental phenomena adequately. Brie Gertler groups a number of theories as empiricist[1] (Gertler, 2018) and argues that these theories take self-knowledge to be epistemically based in empirical justification or warrant. These theories, she notes, face an important challenge: they portray us "as mere observers of a passing cognitive show", thereby neglecting the fact that some mental episodes such as "believing and intending are things we *do*, for *reasons*" (Gertler, 2018, p. 1).

[1] Sometimes other concepts are used to refer to the same thing, such as "purely epistemic" (Moran, 2001, p. 38), "purely theoretical" (Moran, 2001, p. 27).

As Gertler claims, agentialist authors such as Burge, Moran, Bilgrami, and Boyle argue that our capacity for self-knowledge is grounded in our rational agency – in "our ability to conform our attitudes to our reasons, and to commit ourselves to those attitudes through avowals" (Gertler, 2018, p. 1). I will discuss the agential model in more detail later. For now, what has been said suffices for giving an idea about the motivation why agentialists might find some accounts deficient for elucidating the nature of authoritative, first-personal self-knowledge.

Before moving to the agential model, considering the purposes of my thesis, I need to make clear one important distinction in philosophy of mind between conscious, occurrent mental episodes and standing mental states. On some understanding, the former are episodes that are part, if we use the term by William James, of "the stream […] of consciousness" (James, 1890, p. 233). These episodes may include bodily sensations, perceptual experiences, daydreams or imaginations, feelings and emotions, conscious thought, reasoning, etc. What is assumed by calling these mental phenomena conscious is the fact that they are directly, immediately present to our mind, in our conscious experience. We can direct our "mind's eye" to them, we can attend to and individuate them just by being directly, consciously aware of them. There are theories of self-knowledge, Gertler calls them "acquaintance accounts" (Gertler, 2021), according to which we have a direct access to our mental states in epistemological sense (I don't need to know anything else in order to know these states), as well as in metaphysical sense (there is nothing in-between my knowledge and the mental state itself) (Gertler, 2021, p. 6). These theories seem to be best suited to account for the self-knowledge of the above-mentioned conscious mental states/episodes.

However, the subject's mental life is not exhausted by episodes in the stream of consciousness. It has other, non-conscious items, including standing states: "those mental features like beliefs, desires or intentions, which a subject can have even if she is not conscious, or when her

consciousness is occupied with something else" (Farkas & Crane, 2022, p. 303). Thus, a standing state of, say, believing that p and the conscious episode of judging, assenting to or endorsing p are different kinds of entities.

Although not making this distinction explicit, Moran's agential model mostly concerns standing mental states/attitudes like beliefs, desires and intentions.[2] Considering this, standing emotions – the target phenomena of my thesis – seem to be better suited for Moran's model than their conscious counterpart – occurrent emotions. I will address this issue in more detail and explain its importance for our purposes later in my thesis.

1.2 Agential Accounts

Richard Moran's account belongs to agential theories of self-knowledge. Thus, before describing his model in detail, I will account for agential theories in general.

Brie Gertler (Gertler, 2022) points out a contrast in the direction of influence or causation between knowledge and agency, in order to show what the starting point for the agential view is. She notes that ordinary cases of knowledge usually involve some object or phenomenon affecting the subject. For example, I know that the keyboard in front of me is black, because this external object – the black keyboard – affects my visual perception; I see that it is black. By contrast, exercising agency involves the subject affecting objects or phenomena in the world. For example, I am exercising my agency now as I'm typing on this black keyboard. It is me who is doing this. The fact that the keyboard is being typed on depends on my agency.

[2] However, he says that his discussion covers "various emotional states" (Moran, 2001, p. xxxiii)

This distinction should not be controversial, since it is a very common way how we take things to be in our everyday life.

The distinction is important in order to see how agential accounts differ from other accounts of self-knowledge. As Gertler notes, the agential view claims that the standard model of knowledge described above (as being affected by objects or phenomena) cannot exhaustively explain self-knowledge of, at least some of, our mental states. The standard model can be useful to account for the self-knowledge of particular kind – passive – mental states, such as pain or other bodily sensations. However, it is implausible as an explanation of self-knowledge of active mental states, such as beliefs or intentions[3] (which Gertler classifies as "committal attitudes" – a sub-group of active states) (Gertler, 2022, p. 1).

Active states are things that we do. As Gertler notes (Gertler, 2022), the paradigmatic examples of committal attitudes – beliefs, intentions – are things that we do for reasons. For example, deciding to do a PhD or reflecting on the evidence and making a conclusion – that results in forming a belief that doing a PhD is hard – seem to be things that we do according to reasons. This kind of things, according to Gertler, are reason-responsive in other ways too: they are apt to change in response to the change in evidence or consideration. Agentialists assert that reason-responsiveness is a kind of agency (Gertler, 2022, p. 2).

Richard Moran's account can be grouped with several other agential accounts of Tyler Burge (1996), Akeel Bilgrami (2006), and Matthew Boyle (2009). As Brie Gertler notes, all of these accounts are characterized by a broadly Kantian approach to reason and agency:

> On this approach, a thinker's most basic self-conception (as an "I") is agential: we see ourselves as authors of our beliefs and intentions, rather than

[3] I will discuss the view why and in what sense beliefs are active mental states later in detail.

> as an inert thing in which attitudes merely occur. This authorial agency is essentially rational: it is exercised when we believe or intend on the basis of reasons" (Gertler, 2021, p. 13).

Like Moran, Burge and Bilgrami explain self-knowledge by appealing to the concept of rational agency. For example, Bilgrami claims that the fact that we are rational agents, necessarily implies that we can know our attitudes. Thus, we are necessarily aware of our beliefs just because they are commitments (Gertler, 2021, p. 14). As for Burge, our rational agency is a source of the obligation to meet some rational norms. Thus, by being responsible to satisfy these rational norms, we are epistemically entitled to the judgments regarding our mental attitudes without which those norms cannot be satisfied (Gertler, 2021, p. 14). Boyle too, shares the importance of responsibility in explaining our awareness of our conscious beliefs (Gertler, 2021, p. 14).

I will not discuss these accounts in detail. I will focus on Richard Moran's account, since I believe that it provides us with the strongest theoretical framework that could explain the best the target phenomenon of my thesis – standing emotions.

1.3 Moran's Version of Agential Model

Now, it is time to account for Moran's agential model in detail. Richard Moran, perhaps the most influential author in the agential theories of self-knowledge, grounds his account in agency that is constituted by our capacity to rationally shape our attitudes (Moran, 2001). According to him, it is our *agential authority* with respect to our attitudes that results in our *epistemic authority*. Since Moran holds that any philosophical account of self-knowledge must be an account of the first-person perspective (Moran, 2001, p. 91), the aim of his agential approach is to explain a distinctively first-personal relation (Moran, 2001, p. 27). In order to do this, he claims, we cannot draw upon exclusively on privileged, special kind of epistemic

access, and, more generally, on purely theoretical awareness. Instead, we should understand the phenomenon under consideration in relation to the agential responsibility:

> The special features of first-person awareness cannot be understood by thinking of it purely in terms of epistemic access (whether quasi-perceptual or not) to a special realm to which only one person has entry. Rather, we must think of it in terms of the special responsibilities the person has in virtue of the mental life in question being his own. In much the same way that his actions cannot be for him just part of the passing show, so his beliefs and other attitudes must be seen by him as expressive of his various and evolving relations to his environment, and not as a mere succession of representations (to which, for some reason, he is the only witness) […] It is modeling self-consciousness on the theoretical awareness of objects that obscures the specifically first-person character of the phenomenon, whether or not this theoretical perspective takes the specific form of the perceptual model of introspection […] A more complete characterization of the first-person perspective will require bringing the agent more explicitly into the picture (Moran, 2001, pp. 32-33).

It's important to note that Moran does not deny that there are essential aspects of self-knowledge that must be explained in epistemological terms. What he denies is that the purely epistemic questions and/or approaches exhaust the phenomenon of self-knowledge and challenges associated with it. In his words, "the difficulties in properly characterizing the first-person position are not merely epistemological ones" (Moran, 2001, p. 3). There are wider self-other asymmetries which oblige us to "ground the discussion as much in moral psychology as in epistemology" (Moran, 2001, p. 2).

Theoretical, purely epistemic accounts mentioned above would explain the special character of first-personal authoritative self-knowledge by appealing to epistemic authority regarding our own mental states or attitudes. Epistemic authority is mainly described by epistemically privileged position, which in turn, results from privileged, first-personal access to one's own mental states. The privileged position, according to Gertler, implies that "compared with others' beliefs about my attitudes, my own self-attributions are more strongly justified, or open

to fewer sources of error, or produced by processes that are more reliable (etc.)" (Gertler, 2022, p. 3).

The main problem with a purely epistemic approach is that it leaves out something crucially important about the nature of first-personal, authoritative self-knowledge – its intimate connection to "a capacity for rational, responsible self-directed agency" (McGeer, 2008, p. 83).

Thus, for agentialists, a more significant kind of authority is agential authority, which is capable of explaining epistemic authority, but not vice versa. While epistemic authority would be exhausted by reliably reporting one's mental states, such as feeling pain, agential authority requires a kind of commitment and taking responsibility for the mental state or attitude in question. Moran calls this act of expressing a commitment "avowal". Thus, as Carlos J. Moya notes, "avowal" is an expression of one's attitudes from a committed, endorsing, first-personal stance (Moya, 2006). Here, the *Transparency Condition* plays a crucial role. Since according to Moran, our agency and responsibility in relation to our mental states is tied to our capacity of avowal in accordance with the Transparency Condition. The latter, in Moran's words, with respect to belief, "is that from within the first-person perspective, I treat the question of my belief about P as equivalent to the question of the truth of P" (Moran, 2001, pp. 62-63).

In introducing Transparency Condition, Moran cites Gareth Evans[4]:

> "In making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward —upon the world. If someone asks me "Do you think there is going to be a third world war?", I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?" I get myself in a position to answer the question whether I believe that p by putting

[4] The paragraph below by Gareth Evans is not fully cited by Moran

> into operation whatever procedure I have for answering the question whether p. (Evans, 1982, p. 225).

Moran explains that, usually, when I ask myself a question "Do I believe that P?", I treat this question as equivalent to the question "Is P true?". What is important to note is that this kind of transparent relation between these two kinds of questions is uniquely first-personal. This is not how I normally relate myself to the question of someone else's beliefs (Moran, 2001, p. 60). This is why it would be absurd to judge "P is true but I don't believe P". While it makes perfect sense to say "P is true, but he does not believe P".

Thus, according to Moran, this commitment to a mental state (e.g. belief), in conformity to the transparency condition, which is a case of exercising my rational agency, is how I possess first-person authority regarding the self-ascription of this belief.

Moran says this exercise of rational agency is *immediate* in that a judgment is not inferred from observational, behavioral evidence. Even though these first-personal judgments are not based on any evidence, they don't lack authority. As Moran claims, they are not trivial, but substantial - representing a genuine cognitive achievement (Moran, 2001, p. 3).

According to Moran, *epistemic immediacy* is one of two distinct aspects of the first-person authority that normally come together but can in principle be dissociated. Epistemic immediacy asserts that, unlike other people who must rely on observation, evidence, or inference, I can say what is "in my mind" without any of this (Moran, 2001, p. 91). Another aspect is already mentioned *transparency* – 'looking through' one's attitudes, directly to the world. Transparency is grounded in endorsement: "when endorsement fails, then so does transparency, for without endorsement the person cannot declare his belief through avowal of it. He might still, however, retain a kind of immediate epistemic access to it" (Moran, 2001, p. 92). Therefore, mere immediacy is not sufficient to explain first-person authority. Since the latter

for Moran, as mentioned earlier, is the kind of authority of the subject to make up his mind, "tied to the presuppositions of rational agency" and "different in kind from the more purely epistemic authority that may attach to the special immediacy of the person's access to his mental life" (Moran, 2001, p. 92).

Thus, to summarize Moran's points so far, we can say that immediacy, the subject's special access to his mental states, is only one aspect of first-person authority (an aspect corresponding to purely epistemic authority). But it is not sufficient to account for the essence of the phenomenon of first-person authority. Since the latter involves something more. Namely, the kind of agential authority of the person making up his mind, which comes with avowals in conformity with the Transparency Condition, resulting in determining and knowing one's own mental state.

1.3.1 Immediate Access for Standing States?

There is one aspect about "immediacy" that needs to be addressed to avoid confusion. As I made clear in the first section of this chapter, conscious/occurrent mental episodes, that is, episodes that are part of the stream of consciousness, such as passing thoughts, are said to be immediately, directly presented to the subject, in a sense that our awareness of them is not mediated by anything else. For example, when we feel pain, we have immediate access to this experience. On the other hand, as I mentioned before, there are standing mental states that "a subject can have even if she is not conscious, or when her consciousness is occupied with something else" (Farkas & Crane, 2022, p. 303). These are the kind of state that Moran focuses on in his book. He thinks that we can have immediate access to standing states as well. This immediacy, however, is not and cannot be the same immediacy that we have with respect to the conscious mental states. Then, in what sense is our access to standing states immediate?

For Moran, "immediate" just means "non-observational". We have an immediate access to our belief that p in a sense that we do not make an inference from observational, empirical data. That is, we do not consider some causal, explanatory reasons for or against thinking that we do, in fact, have the belief in question. Rather, we are immediately aware of our belief by being aware of justifying reasons for or against its content, its truth.

Moran writes:

> ...it is because the deliberator declares the authority of reason over his thought and action that at the conclusion of his thinking there is no further thing he does to make that conclusion his actual belief or his intention (Moran, 2001, p. 131).

That is, what's so special about rational, justifying reasons as evidence, as opposed to empirical, behavioral evidence acquired through observation, is that in the case of the former, the conclusion of the judgment of what is the case, guarantees (and requires) the authoritative self-ascription of the belief without any extra, mediating steps in-between. On the other hand, a belief that I self-ascribe based on behavioral evidence requires me to do something additional to make the conclusion my own belief. I need to identify myself with the condition about which I know from the third-personal inference. Thus, by adopting a third-personal perspective I cannot have any more authority than anyone else who similarly infers my mental state from the observational evidence.

This is how Moran understands immediacy, and this is the immediacy that constitutes an inevitable aspect of the person's capacity to avow, assuming responsibility for his mental state or attitude.

1.3.2 Conclusion

Before moving on, to summarize main features of Moran's version of agential model that I will apply to self-knowledge of standing emotions, there are several points that are essential:

1. Some mental states/attitudes are active in a sense that they respond to justifying reasons.
2. The special first-personal authority of these mental states/attitudes cannot be sufficiently explained by purely epistemic models of self-knowledge.
3. The special authority of these mental states/attitudes can be accounted for by agential model: we authoritatively know our active mental states/attitudes by exercising our rational agency, by taking a deliberative stance with respect to them. By reflecting on justifying reasons for or against them, and by expressing avowals in conformity with the transparency condition, we commit ourselves to, we take responsibility for these states/attitudes.

The aim of my work is to find out whether the kind of agential model of self-knowledge that endorses the claims given above can account for whether and how we can have authoritative first-personal self-knowledge of our standing emotions.

In the following chapter, I will give a picture of the state of art in philosophy of emotions, describing several main trends and distinctions, relevant for my purposes. Then, I will account for the phenomenon of standing emotions, explaining its main features and characteristics.

Chapter Two:

Emotions as Standing States

2.1 Philosophical Discourse About Emotions

As I mentioned before, in this thesis I inquire whether Moran's agential model of self-knowledge can be applied to standing emotions. Since the latter are a kind of emotions, in order to account for them, I need to situate them in a broader picture of the philosophy of emotions. Thus, I will start with explaining some main features of philosophical understanding of emotions.

Anger, fear, joy, sadness, disappointment, frustration, shame, pride, disgust, amusement, boredom, envy, jealousy, admiration, fascination, embarrassment, regret, hope – these are some paradigmatic examples of the phenomenon we call "emotion". It is a widely shared belief that emotions are conscious episodes having components (Scarantino & de Sousa, 2021):

> Consider an episode of intense fear due to the sudden appearance of a grizzly bear on your path while hiking. At first blush, we can distinguish in the complex event that is fear an *evaluative* component (e.g., appraising the bear as dangerous), a *physiological* component (e.g., increased heart rate and blood pressure), a *phenomenological* component (e.g., an unpleasant feeling), an *expressive* component (e.g., upper eyelids raised, jaw dropped open, lips stretched horizontally), a *behavioral* component (e.g., a tendency to flee), and a *mental* component (e.g., focusing attention) (Scarantino & de Sousa, 2021, p. 3).

There are different ways to categorize the traditions in the philosophical study of emotions. To make things easier, two opposing poles can be identified in relation to which we can localize any tradition. Both of these poles emphasize one of the prima facie incompatible aspects of the nature of emotions – their active and passive, cognitive and conative, intentional and bodily aspects. Depending on the fact which of the above-mentioned poles a particular tradition is

closer to, traditions take some of those components of emotions to be essential. Now, let's have a look at those opposing poles.

To start with the view which accounts for the features occupying the right side in the pairs mentioned above (active and passive, cognitive and conative, intentional and bodily aspects), Martha Nussbaum's (Nussbaum, 2004) description would be illuminating. She describes the view as "adversary" to her own view stating that emotions are evaluative judgments:[5]

> […] the view that emotions are "nonreasoning movements," unthinking energies that simply push the person around and do not relate to conscious perceptions. Like gusts of wind or the currents of the sea, they move, and move the person, but obtusely, without vision of an object or beliefs about it. In this sense they are "pushes" rather than "pulls." This view is connected with the idea that emotions derive from the "animal" part of our nature, rather than from a specifically human part—usually by thinkers who do not have a high regard for animal intelligence. Sometimes, too, the adversary's view is connected with the idea that emotions are "bodily" rather than "mental," as if this were sufficient to make them unintelligent rather than intelligent (Nussbaum, 2004, p. 186).

The main representative of the view described above by Nussbaum is so-called James-Lange feelings theory with its various versions. The underlying idea of this theory is that emotions are the perceptions of felt bodily changes (Cannon, 1927). Thus, this type of theories emphasizes the *feeling, phenomenological, bodily aspects*, which do seem to be essential for the nature of emotions.

As for Nussbaum's account, it belongs to the opposite pole, which can be named as cognitivist theories. There are versions of cognitivist theories. In their simplest form, emotions are evaluative judgments (Nussbaum 2001, Solomon 1980). However, what any version of

[5] According to Nussbaum's theory, emotions are forms of evaluative judgement, where judgement is defined as an assent to an appearance (Nussbaum, 2004).

cognitivist theories should account for is that emotions are *evaluative* in some sense and that they are *intentional* – directed at or about some things in the world.

Some philosophers have tried to find a common ground between these two opposing poles and somehow integrate the above-described prima facie incompatible characteristics of emotions. Perhaps the perceptual theories have been most successful. It seems that perceptual theories, asserting that emotions are a kind of perception or that emotions share essential features with sensory perception (Brady, 2013), managed to capture most of those incompatible characteristics from both sides and offer the unifying explanation that would account for the dual nature of emotions. However, expectedly, this type of accounts is not without its deficiencies and challenges.[6] It is not my aim here to argue for or against the plausibility of any particular view. For now, I will point out that standing emotions, although sharing the above-mentioned duality of emotions to some degree, seem to be siding the cognitivist pole more.

As I said, I will not argue explicitly for or against either of these views here. However, I want to claim that a theory that I would favor should not neglect the following facts:

1. Emotions have phenomenology
2. Emotions are intentional
3. Emotions are or involve some kind of evaluation of their objects

The first two facts are widely accepted in the philosophical literature on emotions. By having phenomenology, it is meant that emotions have a what-it-is-like character. While experiencing an emotion it feels like something to be in that mental state (bodily sensations might contribute

[6] For an illuminating critique of perceptual theories see (Deonna & Teroni, 2012), chapter 6.

to this). By saying that emotions are intentional, what is usually meant is that they are directed at or about particular objects.

As for the claim that emotions are or involve some kind of evaluation of their objects, I mean that emotions involve evaluations like "dangerous" or "shameful". This, in turn, implies that similar to perception or judgement, emotions are subject to the correctness conditions[7] (they are "appropriate", "fit", "merited", etc.).

Because of this feature, many philosophers are led to think that emotions have mind-to-world direction of fit. In philosophy of mind, some mental states are said to have mind-to-world direction of fit. For example, it is very common to think that beliefs "ought to fit with, or get it right about, or match up to, the states of affairs they describe or represent (Schwitzgebel, 2024, p. 10). Therefore, we can say that beliefs have mind-to-world direction of fit.

Mind-to-world direction of fit is contrasted with world-to-mind direction of fit, which is exemplified by mental states/attitudes such as desires or intentions. Instead of having the correctness conditions, desires for instance, have the satisfaction conditions – they can be satisfied or frustrated, but not correct or incorrect (Deonna & Teroni, 2024, p. 46).

According to representationalists about emotions, these mental states are representations of evaluative properties (Deonna & Teroni, 2024). As one of the most popular versions of representationalism claims, the latter fact is in virtue of which emotions are a kind of evaluations. According to Roberts (Roberts, 2003), Milona (Milona, 2016) or Tappolet (Tappolet, 2015), "…emotions are evaluations because they are experiences of the relevant

[7] As Deonna and Teroni put Searle's view (Searle, 1983), beliefs and perceptions have the correctness conditions, that is, "they have a content in the light of which it is possible to assess whether they fit the facts or not" (Deonna & Teroni, 2012, p. 6).

evaluative properties" (Deonna & Teroni, 2024, p. 46). Therefore, the correctness conditions as understood by representationalists, imply that emotions would be either correct or incorrect depending on whether they represent the objects correctly or incorrectly. For example, my fear of a bear would be correct if the bear is, in fact, dangerous; or my feeling of being ashamed would be incorrect if the thing I have done is not shameful.

This characterization is crucially important for my purposes in this thesis. As I will show later, while deciding on what kind of reasons do standing emotions respond to (which, in turn, will allow us to decide whether they can be applied the Transparency Condition), we will have to consider this aspect of their nature – their having the correctness conditions in a sense of correctly representing the world.

One additional point about emotions concerns the most important distinction that can be made considering the aim of my thesis. This is the distinction between occurrent and standing emotions. I will account for this aspect in the following section.

2.2 What are Standing Emotions?

As I mentioned in Chapter One, there is a distinction in philosophy of mind between occurrent and standing mental states. In the previous section, one of the distinctions I mentioned but did not elaborate much was between occurrent and standing emotions. Now, I want to account for this distinction in detail to make clear our target phenomena - standing emotions.

Perhaps, the idea of standing emotion is best explained by considering an example. When I say "Maria is angry because her best friend forgot about her birthday", I can mean, at least, two different things. First, I may be implying that Maria is undergoing an emotion of anger right

now, she is currently in the grip of anger. Second, I may have in mind that Maria has an enduring, standing emotion of anger, involving, among other things, the disposition to feel short-term episodes of anger every time she is conscious of the thing that she's angry about - say, that her best friend forgot about her birthday. The first case would refer to a conscious episode of emotions, that is an event, it begins, unfolds and ends at a particular time. The second would be picking out what I call a "standing emotion", or what others call a "sentiment" (Naar, 2018; Sánchez & Salice, 2023), or an "affective disposition" (Deonna & Teroni, 2012) that Maria has even when she is not in the grip of anger, when her attention is taken away by something, say, by a breaking news she is listening to.

"Standing emotion" is not a term that would ring a bell for a non-philosopher. Even within the philosophy, it is extremally uncommon to use this term[8]. "Sentiment" has been used as a mere synonym of emotion during the history of philosophy, and today it has more than one meaning. As for "emotional/affective disposition", I think it might mislead one in taking standing emotions to be just the dispositions to enter short-term emotional episodes, and thus, to unnecessarily limit the scope of the phenomenon in question (I will explain this point in this chapter). Therefore, I decided not to use the latter two terms. Instead, following the distinction between conscious/occurrent and standing mental episodes/states accounted for in Chapter One, it seems very natural for me to call a non-conscious, standing counterpart of occurrent emotion – "standing emotion".

There are two main characteristic features of standing emotions: their lengthy duration (standing emotions can and usually do last for days, months or years), and their intermittent

[8] Chan (1999) is the only author to my knowledge that consistently uses the term.

consciousness (we are not conscious of our standing emotions all the time, but even when we are not, we still have them).

However, it's important to note that, although characteristic, duration does not seem to be an essential feature of standing emotions. It is quite rare, but occurrent emotions can last longer - we can be angry, fearful or sad for hours, or maybe during the whole day. But these states are not necessarily standing emotions. On the other hand, intermittent consciousness seems to be an essential feature. If the emotion ceases to exist when we are not conscious of it, then it is not a standing emotion. For example, if I am no longer angry at someone as soon as my mind becomes occupied by something else, then the emotion is occurrent, conscious or episodic, as opposed to standing or dispositional.

Thus, Peter Hacker (2017) distinguishes between *temporary* and *enduring/persistent* emotions. Within the former, he differentiates between *momentary* or *acute* and the ones that *last for a while* which he calls *episodic. Momentary/acute* emotions are: "a flash of anger, a pang of envy, a flush of pride, or a sudden thrill of fear that immediately dissipates" (Hacker, 2017, p. 29). As for *episodic temporary* emotions,

> "One may feel furious with another for some actual or imaginary slight and remain angry and ill-tempered for the rest of the afternoon. One may feel frightened or terrified by a present danger and cower in terror until it passes an hour or more later. One may feel proud of one's spouse throughout the whole celebratory dinner in his or her honour" (Hacker, 2017, p. 29).

Therefore, *Episodic temporary* emotions are the ones which are felt as long as we are conscious of them.

We need to draw a boundary between *episodic temporary* emotions described above and our target phenomena – standing emotions. The type of emotions that I am interested in persists through time even when we are not conscious of them (for example, when we think about

something else or when we are asleep). “One does not cease to love one’s beloved when one sleeps any more than one ceases to know or believe what one knows or believes when one falls asleep)” (Hacker, 2017, pp. 29-30). This similarity with belief is salient. Beliefs as well as *enduring/persistent* emotions are standing, non-conscious mental states/attitudes. I don’t cease believing that my name is Levan when I sleep or when I don’t think about this fact.

Hacker claims that, unlike *episodic temporary* emotions, *persistent* emotions lack ‘genuine duration’. That is, they “…do not lapse with loss of consciousness” (Hacker, 2017, p. 31). This is the reason for him to conclude that they are neither mental states nor states of mind (Hacker, 2017, p. 31). However, I think it would be much more helpful if we made a distinction between two concepts in metaphysics of mind. Namely, events and states.

For example, according to Davidson, “Events are taken to be unrepeatable, dated individuals such as the particular eruption of a volcano, the (first) birth or death of a person, the playing of the 1968 World Series, or the historic utterance of the words, “You may fire when ready, Gridley” (Davidson, 2003, p. 87). Events have temporal parts — earlier parts and later parts (e.g. the earlier part of the First World War). On the other hand, a state is a thing having a property at a time, or over a period of time: x has property F from t1 to tn. Unlike events, states do not have temporal parts. Examples of mental events are pains, perceiving, thinking etc., while examples of mental states are beliefs, intentions, desires, and so on. Therefore, taking this useful distinction in mind, I think Hacker’s above-mentioned conclusion that just because persistent emotions lack ‘genuine duration’, they are neither mental states nor states of mind, is misguided. In fact, persistent emotions, like beliefs, are standing mental states.

It makes perfect sense to say that while belief is a standing mental state, judging or assenting – a conscious episode of acknowledging or realizing that you believe something – is a mental

event. Similarly, while persistent emotions are standing mental states, conscious/occurrent (or as Hacker calls, *temporary*) emotions are mental events. Now, there is an obviously close but not simple relationship between these standing states and their counterparts. We need not go into details here.

I will address now two crucial aspects about ontology of standing emotions: the relation between standing emotions and dispositions; and the reducibility of standing emotions to occurrent emotions.

2.2.1 The Relation Between Standing Emotions and Dispositions

Standing emotions involve dispositions. Namely, having a standing emotion implies being disposed to experience occurrent, short-term emotional episodes. For example, when I love someone I get excited when I am around them, I get sad when they are gone, I get jealous when I see them with others, angry when someone tries to harm them, etc. However, I think, to claim that standing emotions are identical to dispositions to ender short-term emotional episodes is to unnecessarily limit their scope. We have sufficient reason to think that dispositional approach (identifying standing emotions with the above-mentioned dispositions) does not fully account for the nature of standing emotions. While a standing emotion implies disposition, the converse is not true. One can be disposed to enter into a short-term emotional state toward a certain object without having a standing emotion toward the object (Chan, 1999):

> My encounters with a waitress at a restaurant may be so unfortunate that every time I visit the restaurant, I get into a row with her, with both of us ending in boiling anger. However, it need not be the case that I have a standing anger toward her. Indeed I may have correctly dismissed what had happened as silly, temporary hysteria and now have completely forgotten about her (Chan, 1999, p. 497).

Moreover, according to Chan, dispositional accounts cannot give satisfactory explanations of emotional behavior, because they leave no room for the causal power of standing emotions (Chan, 1999). If we identify a psychological state in purely dispositional terms, Chan asserts, that will be merely locating it in a statistical or law-like generalization, which means to deny its having any causal power. Standing emotions, however, do have causal powers: they color our perspectives and influence our behavior (Chan, 1999).

Chan offers the "realist view", according to which standing emotions are "enduring complex mental states, which have as their constituents all or some of the following elements: judgments, desires, beliefs, and affects" (Chan, 1999, p. 502).

On the whole, I take Chan's argument to give us a reason not for rejecting dispositional approach as such, but for judging the view that standing emotions are just dispositions to ender short-term emotional episodes as unreasonably limiting.

2.2.2 The Reducibility of Standing Emotions to Occurrent Emotions

Are standing emotions sui generis affective states, or do they reduce to some other affective phenomena such as occurrent emotions?

Hichem Naar (2018) points out that the distinction in common language between affective states predicated to the subjects for a relatively short time (occurrent emotions) and affective states predicated to the subjects for a relatively long time (standing emotions, which he calls "sentiments") can be interpreted in two ways: either as implying a genuine ontological distinction or not. (Naar, 2018).

Accordingly, Naar distinguishes between the *distinctness view* (which takes the conceptual/linguistic distinction to imply the ontological one, claiming that standing emotions

and occurrent emotions are equally basic mental states), and the *pattern view*, which maintains that standing emotions are merely collections of occurrent emotions. He considers the *pattern view* to be a default view in the philosophical and psychological literature on emotions, stating that occurrent emotions are ontologically basic and standing emotions are derivative of them. That is, according to this view, one's account of standing emotions will be determined by one's account of occurrent emotions.

From these considerations, which, as mentioned above, are dubbed as the *pattern view*, Naar formulates the Priority Thesis:

> Priority Thesis: A subject possesses a given sentiment over a particular period of time if and only if, and because, the subject experiences a certain pattern of emotions over that period of time (Naar, 2018, p. 153).

Priority Thesis, as Naar explains, incorporates the sufficiency claim and the necessity claim. According to the sufficiency claim, if a subject experiences a certain pattern of occurrent emotions of love, anger, disappointment, etc. with respect to an object O, she thereby has the standing emotions of love, anger, disappointment, etc. with O. As for the necessity claim, in order to have the standing emotions of love, anger, disappointment, etc. with O, the subject must experience a certain pattern of occurrent emotions of love, anger, disappointment, etc.

After considering some counter examples against both the sufficiency and necessity claims, he concludes that the *pattern view* (along with any other reductive approach) should be rejected in favor of the *distinctness view* which holds that standing emotions are distinct mental states, as basic as occurrent emotions.

My goal is not to argue for this view. I simply take it as the correct understanding of the nature of standing emotions. What's important is that this distinction is going to play a crucial role for us. I aim to show that there are some significant challenges that occurrent emotions face,

especially with respect to reason-responsiveness and agency, that standing emotions avoid. This is made possible by the fact that occurrent and standing emotions are distinct from each other.

In the next chapter, I will answer the main question of my thesis – whether Moran's agential model can be applied to standing emotions. In order to do that, I will discuss whether we can exercise agency with respect to standing emotions, and also, what kind of agency is under consideration. After showing that we are talking about rational agency, I will address a challenge concerning the Transparency Condition. By discussing the applicability of the Transparency Condition, I will thereby inquire what kind of reasons standing emotions respond to.

Chapter Three:

Standing Emotions and Agency

3.1 Relevant Sort of Agency

The main concern of my thesis is to find out whether the agential model of self-knowledge can be applied to standing emotions. In order to decide that, the most crucial thing we need to understand is whether we exercise agency with respect to these mental states. Before I address that core question, however, I think it's important to make it clearer what kind of agency is under consideration. I already mentioned shortly in chapter One that the kind of agency Moran talks about is rational agency. However, this concept needs to be explained in more detail. I will do the latter after accounting for the fact that we hold each other responsible for our (standing) emotions - prima facie indicating that these states admit agency.

3.1.1 Responsibility

One of the peculiar facts about the nature of standing emotions that supports the idea that they are the right sort of mental states to admit agency, is a common practice of taking ourselves and others responsible for these mental states. As Nancy Sherman writes:

> There is no shortage of examples of our practice of holding people (including ourselves) responsible for their emotions. Within relationships, some emotional interactions are attuned; others are misattuned and lack empathetic sensitivity. Such interactions may be occasions warranting moral reproach, or perhaps gentle suggestions of how we would have preferred to be treated. Sometimes it is an emotional episode that calls for judgment; other times, a more long-term disposition. Thus, I can reproach myself for an unjustified outburst of hostility toward a friend, but can also reproach myself for a pattern of aggression that has a longer history and involves a series of repetitions. Similarly, I may reproach my mother in-law for a moment in which she expressed little compassion for her grandson's travails, but can also reproach her for a more enduring character trait whereby she can be

> emotionally removed from others' felt losses or disappointments (Sherman, 1999, p. 298).

In this passage, both occurrent and standing emotions are mentioned. Indeed, it seems to be the case that we feel responsibility for (some of) our short-term emotional episodes too. However, it should not be hard to see that standing emotions are a better candidate for responsibility than the occurrent ones. While it is very hard to imagine any standing emotion for which we would not hold the person responsible, on the other hand, we can have a handful of occurrent emotions that seem to hardly admit any kind of responsibility. These are the kind of occurrent emotions already mentioned earlier: automatic, reflex-like, bodily reactions. For example, we would abstain from holding someone responsible for being startled due to hearing a sound of a shotgun, or for being immediately disgusted by seeing a piece of meat infested with maggots, or for being surprised for watching an illusionist doing a trick, and so on. This kind of emotional reactions hardly admits any responsibility. On the other hand, it is very hard to find similar cases for standing emotions. It makes sense to hold someone responsible for the long-term standing anger at her friend, for her years-long hate for a political party, for her enduring disappointment in modern academic life, and so on. Even a case such as romantic love admits responsibility. Despite a long tradition in classical or pop culture to depict romantic love as something happening to you without you doing anything, when you are engaged in a continuous romantic relationship (even if it is unrequited), it does not sound surprising to feel responsible and for others to hold you responsible for that.

Now, having responsibility intuitively implies having some kind of control. If I cannot change anything, it seems unfair to be held responsible. This intuition is closely related to the common understanding of agency, in general, as involving some kind of control. Thus, it is important to

discuss whether having control is necessary for having agency. I will address this question in the next section.

3.1.2 Voluntariness and Control

"[…] an agent is a being with the capacity to act, and 'agency' denotes the exercise or manifestation of this capacity" (Schlosser, 2019 , p. 1). "The capacity to act" naturally leads to the intuition that the concept of agency is tied to the concept of will or control. For example, R. Jay Wallace argues:

> …in our self-understanding as agents we take for granted that there is a kind of motivation that is directly subject to our own control. My contention is that we need to acknowledge such volitionalist motivations if we are to make sense of reasons for action as norms of reasoning, capable both of guiding our activity through deliberation and of retaining their normative force when such guidance breaks down (Wallace, 1999, p. 237).

As Hichem Naar claims, if we model our account of agency on intentional action as a paradigm case of a reason-guided response, which is plausibly voluntary, we will have a notion of agency which incorporates this aspect of voluntariness as well (Naar, 2022, p. 111). The question is, despite this intuitiveness, how essential is the aspect of control for agency?

In order to answer this question, I will start by discussing one of the arguments that Naar proposes (Naar, 2022). His formulation of the argument incorporates the claim that control is necessary for agency:

(1) If there are reasons for R, then R can in principle be guided by reasons. [Guidance-Reason Link]

(2) If R can be guided by reasons, then R is the sort of thing we can form voluntarily (alternatively: we have 'control' over R; we can form R at will).

(3) Emotions are not the sort of thing we can form voluntarily.

(4) Therefore, emotions cannot be guided by reasons, not even in principle. (From 2 and 3)

(5) Therefore, there are no reasons for emotions. (From 1 and 4) (Naar, 2022, p. 111).

Naar thinks that we can deny p2 since it is possible for there to be reasons for nonvoluntary responses. As he notes, belief is the paradigm example of a nonvoluntary response that is subject to reasons. We can control our beliefs only indirectly by actively searching for evidence that might lead to the conclusion – judgment. Naar's argument is that if the rejection of voluntarism about beliefs does not imply a rejection of their being subject to reasons, then emotions can also be nonvoluntary responses for which we have reasons.

In another work, in order to explain the idea of rational agency, Naar cites Joseph Raz (Naar, 2024, pp. 71-72):

> …in what sense are our normal feelings and emotions, desires and beliefs, etc., 'ours' or 'under our control'? My suggestion was that life is activity and we are active in so far as, as it seems to us, we function well, that is in so far as, as it seems to us, our moods, emotions, beliefs, desires, etc., are properly responsive to reason (Raz, 1999, p. 21).

Thus, so long as emotions are responsive to reasons, they can be characterized as active. In this respect, emotions are not different from beliefs – attitudes that can conform to reasons.

This is the kind of agency – rational agency - that Moran has in mind when arguing for his agential account. He tries to avoid using the word "control" with respect to rational agency. Thus, he differentiates between "assuming authority for one's attitudes (seeing them as "up to me") and "exerting control over them" (Moran R. , 2001, p. xx). When he says "control" he

means some kind of instrumental or indirect influence which is at odds with the kind of activity that is characteristic of rational agency:

> …In various cases a person may produce in himself various desires, beliefs, or emotional responses, either by training, mental discipline, drugs, the cooperation of friends, or simply by hurling himself into a situation that will force a certain response from him. But exercising this sort of control over one's attitudes is not the expression of "activity" relevant to autonomy or rational authority. In such cases of producing a desire in oneself, the resulting attitude is still one I am essentially passive with respect to. It is inflicted on me, even if I am the one inflicting it (Moran R. , 2001, p. 117).

Here Moran considers the case of desire, arguing that when already expressing the conformity with the person's reasons, exerting control over the desire is redundant:

> …the specifically first-personal responsibility that a person has for his own desire is essentially not instrumental. The person's responsibility here is to make his desire answerable to and adjustable in the light of his sense of some good to pursue. It is not a responsibility that reduces to the ability to exert influence over one's desires, and that is why the idiom of "control" is misleading in this context. When the desire is (already) the expression of the person's reasons, there is no need for exerting any control over it (Moran R. , 2001, p. 119).

Thus, the requirement for (at least, a traditional understanding of) control for agency is refuted by maintaining that the kind of agency at play which is relevant for autonomous person is rational agency.

To sum up, I compared emotions to beliefs and said that both are nonvoluntary, but reason-responsive. I argued that, on some understanding, due to their reason-responsiveness emotions and beliefs involve rational agency, and are active.

3.1.3 Difference Between Standing and Occurrent Emotions in Terms of Reason-Responsiveness

Now, I would like to point out one important difference between occurrent and standing emotions in terms of their being active (reason-responsive). Even if we take the above-described, more liberal approach regarding how to understand "being active" or agency and interpret it as reason-responsiveness, there might be some worries regarding occurrent emotions. Considering these worries will make it clear why standing emotions fare better than occurrent ones in terms of rational agency.

As I discussed in chapter Two, it is the fact that many occurrent emotions are automatic, purely somatic, or reflexive. As Naar notes, "emotions can certainly look like wild forces insensitive to reason – mere passions – and massively influenced by arational factors" (Naar, 2022, p. 113). He concedes that these emotions seem to be the wrong kind of mental states to be seen as responses to reasons. It is very hard to see a reflex-like response of being startled as the sort of thing that can be formed based on reasons. Rather, it looks like an automatic, bodily response. How should we account for the reason-responsiveness of this kind of occurrent emotions? Is being formed through reasoning necessary for a mental state to be reason-responsive?

As Kloosterboer argues, if reason-responsiveness implies that mental attitudes are

> formed through explicit deliberation, then many of our beliefs and other attitudes would not count as reason-responsive, since it is commonly accepted that many of our attitudes are not the result of deliberation but of unconscious processes (Kloosterboer, 2015, p. 249)

Alternatively, as she suggests, on a narrower interpretation of reason-responsiveness, awareness of one's reasons might not be necessary. Instead of being formed as a result of explicit deliberation, it is required from the mental states/attitudes not to contradict with the reasons one takes herself to have. That is, one is asked to make sure not seeing any defeaters for the attitude in question. "Being reason-responsive" according to Kloosterboer, "then means, in a minimal sense, that my attitudes are affected if I become aware of the presence of a defeater in the landscape of reasons" (Kloosterboer, 2015, p. 249).

However, the fact that not all beliefs are formed through rational deliberation does not mean that beliefs, in principle, cannot be formed through such process. This is what is essential for Moran. According to him, what is crucial for mental states/attitudes is that it must be, in principle, possible for them to be the conclusion of some kind of reasoning (practical or theoretical):

> Unlike a toothache, or a tooth for that matter, a belief is of the right grammatical kind to be the conclusion of one's reasoning. Similarly, what is essential for a desire to count as "motivated" in the relevant sense is for it to be the possible conclusion of some practical reasoning (Moran R. , 2001, p. 116).

What about standing emotions? Can they, in principle, be the conclusion of some practical or theoretical reasoning? If they can, then introducing the narrower interpretation of reason-responsiveness mentioned above will not be necessary.

First, we need to ask, what it means to be formed through rational deliberation or reasoning. As Hichem Naar points out, there are two, conservative and liberal understanding. According to the former, "concluding that p (…) *is* coming to believe that p" (Naar, 2022, p. 112). Therefore, as he claims, it seems unlikely that emotions (Naar discusses emotions in general) can be a conclusion of reasoning in this sense. As for the liberal interpretation, drawing upon Dancy (Dancy, 2004), McHugh (McHugh & Way, 2016), and others, Naar argues that

"reasoning is at bottom the formation of attitudes and performance of actions on the basis of considerations" (Naar, 2022, p. 112). This means that emotions would, in principle, be formed through reflection provided they are reason-responsive in a more general sense. Thus, it seems that we don't need to apply the narrower interpretation of reason-responsiveness to emotions (occurrent as well as standing), suggested by Kloosterboer.

What I would like to emphasize, however, is that even if we apply the narrower interpretation, there will still be a difference between the reason-responsiveness of standing and occurrent emotions. The nature of occurrent emotions is such that they don't always (perhaps even mostly) allow the subject to consider the presence of defeaters, and for the latter to have any effect on the state in question. On the other hand, standing emotions offer sufficient space for rational deliberation, reflection on reasons and evaluation of appropriateness of themselves.

The reason for this difference is the fact that unlike occurrent emotions, while having standing emotions, subjects are not constantly in the grip of intense affects. Moreover, it is crucial to keep in mind that standing emotions are not just the sum of occurrent emotional episodes felt from time to time. As mentioned earlier, mental episodes like judgment, for instance, are constitutive parts of standing emotions. Therefore, it should not be a surprise that standing emotions are more reflection-friendly than the occurrent ones.

Thus, while having standing emotions subjects are more likely to be self-conscious in a more robust sense of this word than while having occurrent emotions. Indeed, despite the fact that occurrent emotions are conscious by definition (in a sense that they are part of the stream of consciousness), subjects might not always be self-conscious while undergoing these experiences. What I mean here by "self-conscious" is that I might consciously feel anger, but I need not have a second-order self-conscious thought "I am angry right now". Furthermore,

even when subjects are self-conscious in this sense, usually this self-consciousness might only imply registering the fact that one is feeling a particular emotion. That is, while being in the grip of anger, for instance (especially intense, violent one), it is very hard to reflect on reasons for or against it. Usually, reflection becomes possible after the emotion dissipates.

Now, one might ask, even if this is so, what is the difference between reflecting on reasons after the intense occurrent emotion dissipates, on the one hand, and reflecting on reasons while having a standing emotion, on the other?

The difference seems crucial. Neglecting the fact that reflecting on the dissipated emotion means that we should retrieve the experience from our memory, which can be tricky sometimes, more importantly, reflecting on reasons after the intense occurrent emotion dissipates means that one would no longer be able to influence the emotion itself: either to maintain or to get rid of it. The only thing that a subject can do is to evaluate whether the emotion has been appropriate (in any sense) or not. On the other hand, engaging in rational deliberation and considering reasons for or against the state of mind while having a standing emotion means that the subject is able to alter the state itself. For instance, when I become aware of my standing anger at my friend, reflecting on reasons might result in my judgment that, in fact, she has not done anything to merit my anger. The fact that she forgot to pick up my delivery was not really her fault as I promised to remind her about that again and I forgot to do that myself. Furthermore, she has been stressed because of her work lately, and I did not explain it sufficiently how important the delivery was for me. Considering these reasons against my standing anger affects the mental state itself. After acknowledging these reasons, I might no longer be angry at her.

I think this is a very important aspect in Moran's account. As Kloosterboer points out, "being committed to a certain mental attitude means making up your mind and sticking with it" (Kloosterboer, 2015, p. 249). This "sticking with it" part of a commitment seems to be essential. It implies some kind of continuity which seems hardly taking place in case of occurrent emotions (despite the fact that they do have temporal length). Furthermore, in case of occurrent emotions, the reason-responsiveness often seems to be somewhat illusory. It is not an uncommon experience that when one gets very angry at something or about someone, and feels an extreme rage, one feels certain that his reaction is adequate. To refer to the example from the previous paragraph, I might have this kind of immediate emotional response of anger to my friend who forgot to pick up a very important delivery of mine. However, as soon as the emotion dissipates and I am in a better condition to reflect on the reasons to evaluate the appropriateness of my emotional reaction, I no longer see it as an adequate response to the corresponding facts. What is important here is that, despite this realization of having had an inadequate reaction, I can no longer influence the occurrent emotion in question, since it is already gone. This means that "sticking with" one's endorsement of his own mental state becomes very implausible in case of (at least some) occurrent emotions.

Considering all this, the distinction made in chapter One between conscious and standing mental states becomes crucial. If committing to a mental state implies some continuity in terms of sticking with it, then the agency and reason-responsiveness seem to be more robust with standing mental states (including standing emotions). Therefore, despite the fact, as mentioned in chapter One, that Moran refers to two basic categories of psychological state – occurrent and standing, the latter are in the focus. Indeed, what I have argued in the last few paragraphs shows that standing mental states appear to be a better candidate to be applied Moran's version of agential account.

3.2 Relevant Sort of Reasons

Having shown that emotions are active in a sense that they respond to reasons, what we need to understand now is what kind of reasons standing emotions respond to. As I mentioned earlier in this chapter, there are some worries that beliefs and emotions (generally) are reason-responsive in crucially different ways. Based on this claim, Naomi Kloosterboer argues that the Transparency Conditions cannot be applied to emotions. Thus, in this part of the chapter, I will answer Kloosterboer's challenge for emotions concerning the Transparency Condition and by doing that, I will thereby account for the nature of reason-responsiveness of emotions, in general, and standing emotions, in particular.

3.2.1 Reasons for Emotions and the Transparency Condition

In order to apply Moran's agential model to standing emotions, we needed to find out, first, whether standing emotions were active mental states - states with respect to which we could exercise our rational agency. I have already showed that standing emotions are reason-responsive, therefore, active in the relevant sense. Thus, the first requirement has been met. However, this is not sufficient. What we need to find out next is whether standing emotions can be applied the Transparency Condition. As I explained in Chapter One, the idea of the Transparency Condition is that we can "see through" our mental states/attitudes to their contents, that is, to the things they are directed at in the world. The fact whether standing emotions can be applied the Transparency Condition depends on the fact of what kind of reasons standing emotions respond to. I will address this issue now.

There are some worries that the Transparency Condition (at least, as understood by Moran) cannot be applied to emotions. Naomi Kloosterboer argues that emotions are conceptually related to concerns – things that we care about. Therefore, in order to acquire self-knowledge of one's emotions, one needs to possess the knowledge of one's other mental states/attitudes as well. However, the scope of the Transparency Condition does not cover that.

As I described in Chapter One, there are two kinds of questions one might ask while trying to find out his mental state. These two questions are taken to stand in the transparency relation to each other. First, what Moran calls the self-related question (Kloosterboer calls it an attitude-question), and second – the world-related question (Kloosterboer calls it a content-question[9]). In case of belief, an example of the former would be "Do I believe that P?" and of the latter – "Is P true?". Because of the alleged transparent relation between them, answering the latter means answering the former as well.

In order to explain why two logically different questions can be answered by appealing to the same reasons, Moran introduces the concept of rational agency:

> It would not ... make sense to answer a question about my state of mind (e.g., my belief about the weather) by attending to a logically independent matter (the weather itself) unless it were legitimate for me to see myself as playing a role in the determination of what I believe generally, ... in the sense that ... the responsiveness to reasons that belongs to beliefs is an expression of the person's rational agency (Moran R. , 2012, p. 213).

Thus, in order for the Transparency Condition to obtain, our deliberative stance towards our mental states – to determine them by making up our mind (reflecting on reasons) – must be justified. That is, the mental states in question must, in fact, be reason-responsive. However,

[9] I will follow Kloosterboer in naming these two questions

the reasons that they respond to must be exclusively content-related. This point will become clearer as we proceed.

As Kloosterboer points out, when we rationally deliberate about our beliefs, the kind of reasons that is relevant for them concerns the truth-value of their content (Kloosterboer, 2015). That is, since beliefs are a kind of things that aim at truth, I don't need to appeal to any other reason in order to endorse a particular belief, other than the reason telling me that the belief is true. Therefore, the relevant reasons for making up my mind about my beliefs concern exclusively the content of these mental states/attitudes (Kloosterboer, 2015, p. 251). For example, if I want to know whether I believe that it is raining outside, the reasons that I will consider for or against this belief will be concerning whether it is true that it is raining outside. Thus, the case of belief is the most straightforward and the least controversial within the agential model.

On the other hand, the relation of emotions to reasons, as Kloosterboer argues, is different. The first thing she claims is that, although being necessary for the justification of an emotion (e.g. anger), the truth of a pure fact (grasped by the underlying belief, constituting or merely accompanying our emotion of anger) that our friend forgot the date is not sufficient for my anger to be justified:

> I am not angry with a friend for forgetting our date only because it is the case that she forgot our date. Of course, my anger is only justified if she has *in fact* forgotten the appointment[10] (…) if I want to deliberate about whether I am angry with my friend or not, saying this is true (*that is, saying that she really forgot our date – my note*) does nothing to explain why being *angry* is the right way to feel (Kloosterboer, 2015, p. 251).

[10] In fact, following Deonna and Teroni's distinction between correctness and justification of emotions (Deonna & Teroni, 2012), my anger would be justified but incorrect if I acquired my belief reliably, but my friend did not, in fact, forget the appointment.

What Kloosterboer claims is that verifying the fact that my friend forgot the appointment does not give us any reason ("does nothing to explain") why my emotion is appropriate ("the right way to feel"). As a result, a content-question cannot be "Is P true?" since, answering that question would only tell us that our friend forgot the date. For Kloosterboer, the latter seems to be necessary but not sufficient to answer an attitude-question of whether we have a particular emotion (anger).[11] Thus, the content-question of an emotion cannot be the same as the content-question of its underlying belief. This seems uncontroversial. Therefore, additionally, for my anger to be justified, I should have reasons for its being appropriate ("the right way to feel").

Kloosterboer assumes that emotions entail evaluation (Kloosterboer, 2015, pp. 250, 252). Thus, my anger would be an apt evaluation of the situation if "forgetting a date is the right sort of thing to be evaluated as offensive" (Kloosterboer, 2015, p. 252). Reasons to justify the latter are the reasons to judge that our emotion is the right way to feel, that is, appropriate.

Thus, an alternative adequate content-question is formulated by Kloosterboer: "Does the fact that my friend forgot our appointment have features that make anger an appropriate response?" Or more generally "Does P (or O) have features that make X an appropriate response?" (Kloosterboer, 2015, p. 252). Introducing "appropriateness" is better than merely talking about the truth of an underlying belief. However, the evaluation's being apt and the emotion's being appropriate, as she argues, is not sufficient for a fully justified emotion.

Before moving on to explain Kloosterboer's argument for the latter claim, to avoid confusion, I think it is necessary to clarify how she uses the concept of "justified" with respect to emotions, since it might be different from other uses in the literature. On one understanding, "a justifed

[11] It is another question whether the truth of the underlying belief is even necessary for the justification of the emotion. I will address this issue later.

emotion is one that the evidence accessible from the subject's perspective makes seemingly correct" (Deonna & Teroni, 2024, p. 46). On this understanding, justification and correctness of emotions might come apart: an emotion can be justified (when the underlying belief is acquired reliably) but incorrect (when the underlying belief is false) or correct (when the underlying belief is true) but unjustified (when the underlying belief is acquired unreliably) (Deonna & Teroni, 2024, p. 46).

Kloosterboer uses the word in a broader sense. According to her discussion, for our emotion to be justified, three conditions must be met:

1. The underlying belief must be true – e.g. it must be true that my friend forgot the date;
2. The evaluation inherent in the emotion must be apt. That is, the thing must be offensive, therefore the emotion – appropriate.
3. We must care about it personally.

I already discussed the first two requirements. I will account for the last one, which is the most important since, according to Kloosterboer, it creates a problem for the application of the Transparency Condition.

As Kloosterboer asserts, insufficiency comes from the fact that we only feel emotions when we care about the thing in question, when it matters to us personally (Kloosterboer, 2015, p. 252). There is a conceptual connection between our emotions and our concerns. Since we need to account for this aspect, we cannot answer the question "Do I feel anger?" without answering an additional question "Do I care about it?" (Kloosterboer, 2015, p. 253). The latter is an attitude-question and not a content-question due to its allegedly inevitable reference to the other mental states of mine: "concerns, plans, ambitions, fears, vulnerabilities, relations to other persons and so on" (Kloosterboer, 2015, p. 253). Thus, she concludes:

> The upshot of the conceptual relation between emotions and concerns is then that deliberation about what to feel cannot be limited to reflection on facts relevant to the specific evaluative content of the emotion, but includes considerations about what is important to someone (Kloosterboer, 2015, p. 253).

Now, I agree with Kloosterboer that an aspect of personal concern is essential for most of our emotions.[12] However, I think she is not right in claiming that in order to find out whether we have a particular emotion, we are inevitably stuck with an attitude-question. In order to show this, I will argue the following:

First, our emotion is appropriate if and only if it correctly represent the world. Thus, my anger is appropriate if the thing in question is insulting.

Second, judging that an emotion is appropriate is possible by merely considering the content-related reasons for it, without asking an additional attitude-question about our cares.

Third, judging that an emotion is appropriate is, normally, sufficient for us to know whether we care about the thing in question, and, therefore, to acknowledge that we do, in fact, have the alleged emotion.

Thus, when we judge that our anger is appropriate, we believe that it is correct (as mentioned earlier, it might not be, but this does not concern our purposes). We believe this because we have relevant (content-related) reasons for it. Our judgment about appropriateness (if appropriateness is understood as described above) implies that we care about it.

Let's consider an example. I've heard my coworker telling her friend that I am stupid and that she is surprised why the boss even let me work here. "Do I feel insulted by hearing this?" In

[12] However, not for all emotions. For instance, it is hard to see for what should I care personally in order for me to get amused by a joke.

order to answer this question, I am asking, “Is this insulting?” While answering this question I search for reasons that would show me that there is something in the world that makes this fact insulting. If I find sufficient reasons to judge positively, that would imply judging that my emotion of being insulted is appropriate.

Since it is not very clear what exactly Kloosterboer means by “the right way to feel” and “appropriate”, that is, whether she takes appropriateness to be merely a matter of representational correctness, I would like to address one potential challenge. One might argue that we do consider other kinds of reasons, too, for our emotions, and it is not the case that the only kind of reason that emotions correspond to concerns the fact whether things, in fact, have the alleged evaluative property. Therefore, understanding “appropriateness” as representational correctness is unjustified.

Following D’Arms and Jacobson (D'Arms & Jacobson, 2000a, 2000b), I believe there is an essential difference between the two senses of “appropriate”. On the one hand, “to call an emotion appropriate is to say that the emotion is fitting: it accurately presents its object as having certain evaluative feature” (D'Arms & Jacobson, 2000a, p. 65). On the other hand, “appropriate” might mean morally, prudentially right, or “all things considered” justified.

They give an example of “a wickedly clever joke”, which is told at the expense of a socially marginalized person. Someone might think that it is inappropriate to be amused because the joke is offensive. “But does this mean the joke isn’t funny?” – D'Arms and Jacobson claim that thinking that it does would amount to a dubious thesis of “comic moralism” (D'Arms & Jacobson, 2000b, p. 731). Thus, they claim, there are certain good reasons for or against appropriateness of emotions, exclusively bearing on the associated evaluative judgement

(D'Arms & Jacobson, 2000b, p. 732). This is known as the Wrong Kind of Reason Problem (WKR problem).

On the pain of committing a "moralistic fallacy" (D'Arms & Jacobson, 2000a), I claim that the question "Does this fact have features that make my feeling of being insulted an appropriate response?" means nothing more than the question "Is this insulting?" What I am asking by the latter is whether the fact has the alleged evaluative property of being insulting. Do I need to inquire about my other mental states or attitudes in order to know whether my emotion is justified? It seems to me that I don't. The only thing that I need to know in order to find out whether my emotion is justified is that the thing in question, in fact, has the alleged evaluative property. If it is the case that x is insulting and also that x happened to me, then my judgment that my emotion of feeling insulted is appropriate would be correct. Therefore, my emotion of feeling insulted would be justified, and consequently, I would have a reason to think that I do feel insulted.

Indeed, it would be very weird if I judged "x is insulting", therefore, "the emotion of feeling insulted is appropriate" and still, "I do not care about x" or "I don't feel insulted". Finding something to be insulting, shameful, disgusting, admirable implies for those things to matter to you, unless there is any kind of significant impairment in one's psychology or unless the conditions are extremally unusual.

Saying "that is insulting, but I don't care about it" or "this is admirable, but I don't care about it" means that one is not aware of or one is not in the condition to endorse the reasons that determine that the thing is, in fact, insulting or admirable. Grasping reasons for your emotions implies that you are motivated in a relevant way. Instead, what the above-mentioned judgment might imply is something like this: "I guess this should be considered as insulting/admirable,

since most people (or a particular, relevant group of people) would count it as insulting/admirable". This is a dispositional claim – guessing what people would feel/do, perhaps in normal or idealized circumstances, probably considering the cultural and other kinds of constraints and inferring from this consideration what would be an "appropriate" (in a sense of expected or normal) response. However, it should be clear that this has nothing to do with understanding and endorsing reasons, having internalized them. In fact, according to Moran's model, this case would not be considered as taking the deliberative stance towards our mental states. Saying that something is insulting, but we don't care about that implies that we judge something's being insulting based on some evidence (what other people would do, etc.) and not based on the reasons that concern the very thing in question, and that we become aware of as the result of rational deliberation.

Thus, the relation between these two kinds of questions/considerations, one about appropriateness and another about caring, is transparent in the sense described earlier. That is, by answering the content-question concerning the appropriateness of an emotion, we thereby answer the attitude-question concerning the personal significance or care as well.

It seems to me that the order of reasoning from content to attitude is more natural than the other way around as well. It is because we judge that things are certain way and, therefore, merit particular responses, that we judge that we care or don't care about them. If I judge that something is shameful, I will judge that my emotion of being ashamed is appropriate, and I care about the thing in question. If I judge that something merits anger, fear, disappointment, admiration, contempt, joy or sadness, I will judge that my corresponding emotions are appropriate and I care about things that my emotions are directed at.

So far, I have argued that, usually, acknowledging that something has the alleged evaluative property and merits a particular emotion (a particular emotion is appropriate) implies that we care about the thing in question. However, perhaps it happens sometimes that things are obscure to us and we are not certain that we care about something. It is important to realize that in this case, we would be uncertain not only about our cares, but also about the very fact whether the thing under consideration has, in fact, the alleged evaluative property (whether it merits a particular emotional response). These things come together. Therefore, elucidating the one implies elucidating the other.

Conclusion

Can Moran's agential account be applied to standing emotions? What was required from standing emotions as a kind of mental states to be eligible for being applied Moran's agential model?

First, it was required that we saw it as up to us to avow, to commit ourselves to those states, to take responsibility for their endorsement. As I have shown, standing emotions can be seen as meeting this condition. Second, it was required that standing emotions were reason-responsive mental states towards which we could take deliberative stance and apply the Transparency Condition. I have argued that standing emotions also satisfy this requirement. It is part of our common practice that we evaluate the appropriateness of our standing emotions based on reasons we have for or against them. Furthermore, the resulting judgment about their appropriateness is sufficient for us to acknowledge having the alleged standing emotion.

To summarize the argument of the thesis, I will consider an example. I notice that recently I feel a particular way about my friend. When I ask myself, "am I disappointed?", the answer might not be very obvious to me. I am not sure how to identify the mental state I am in during this recent period of time. Then, I might start thinking "is Luke's behavior disappointing?" After asking this question to myself, I start remembering some things about him which function as reasons for judging that his actions are indeed disappointing, and therefore, for acknowledging that I really do feel standing disappointment towards him. For example, it immediately comes to my mind that it has been several times in the past three weeks that he declined my suggestion to play boardgames together, saying that he was busy as an excuse. I also remember that he forgot my birthday last week and congratulated to me only two days later when another friend of ours mentioned it in the conversation. Moreover, I remember an

impression of mine which was kind of muted so far, but now it seems clear to me: recently, when we see each other and he asks me about myself, he does not really listen to what I say, as if he is absent, his mind occupied by something else, not interested in what I have to tell him. All of these are reasons for me to judge that a disappointment is an appropriate emotion to have. That is, to judge that his behavior is, indeed, disappointing. On the other hand, some things might suggest themselves as reasons against this judgment. For instance, I also remember that the other day during a party he was very warm and caring. Although acknowledging the fact that caring for and being warm towards someone does not merit disappointment, therefore, taking this as a reason against my standing emotion of disappointment, still, this reason does not seem as compelling to me as the reasons for thinking that the facts considered merit my disappointment. Therefore, I feel certain that, in fact, I do have this emotion towards my friend.

Additionally, in every above-mentioned case, I recall feeling an occurrent, short-term emotion that my mind registered as disappointment. Now I come to realization that what I feel is enduring, long-term emotion, which I am not conscious of all the time, but, still, it is there. On every occasion that I become conscious of it, I know that it has been there all this time. As a result of this acknowledgment, I feel responsible for my standing emotion of disappointment since the latter is my committed attitude, my endorsed stance and avowed state of mind, backed by reflection on justifying reasons and rational deliberation.

Attending to this example we can see that my disappointment towards my friend, Luke is a standing emotion. As I argued in Chapter Two, there are two characteristic features of standing emotions: their temporal length and their intermittent consciousness. Both of these are present in the case described above. First, my anger is long-standing, enduring state of mind, that

continues for days or weeks. Second, it persists through this time even though I am not always aware of it.

I feel responsible and I would hold someone else with the same mental state responsible for having this standing emotion. I see it as up to me to make up my mind about what kind of mental state I am in. I do this as the result of rational deliberation, avowing to the state in conformity to the Transparency Condition. That is, by answering an attitude-question, "Do I have a standing disappointment towards my friend Luke?" I am answering the corresponding content-question, "Is Luke's behavior disappointing?" or "Does Luke merit my disappointment? / is my disappointment appropriate?" I start reflecting on content-related reasons for and against having this mental state, and by answering the content-question, in conformity to the Transparency Condition, I thereby answer the corresponding attitude-question.

Thus, the requirements are met. It should be clear by now that Moran's agential model can be applied to standing emotions. Perhaps some aspects with standing emotions are a bit more controversial than with beliefs. But, still, the agential account of self-knowledge of standing emotions is plausible – and, definitely, more plausible than the agential account of self-knowledge of occurrent emotions.

References

Bilgrami, A. (2006). Self-knowledge and resentment. Cambridge, Mass.: Harvard University Press.

Boyle, M. (2008). Two Kinds of Self-Knowledge. Philosophy and Phenomenological Research, 133-164.

Brady, M. S. (2013). *Emotional Insight: The Epistemic Role of Emotional Experience.* New York: Oxford University Press.

Burge, T. (1996). Our entitlement to self-knowledge. Proceedings of the Aristotelian Society, 91-116.

Cannon, W. B. (1927). The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *American Journal of Psychology*, 106–24. https://psycnet.apa.org/doi/10.2307/1415404

Chan, S. Y. (1999). Standing emotions. *Southern Journal of Philosophy*, 495-513. https://doi.org/10.1111/j.2041-6962.1999.tb00879.x

Dancy, J. (2004). Enticing Reasons. In R. Wallace, P. Pettit, S. Scheffler, & M. Smith, *Reasons and Value: Themes from the Moral Philosophy of Joseph* (pp. 91–118). Clarendon: Oxford University Press.

D'Arms , J., & Jacobson, D. (2000a). The Moralistic Fallacy: On the 'Appropriateness' of Emotions. *Philosophical and Phenomenological Research*, 65-90. https://doi.org/10.2307/2653403

D'Arms, J., & Jacobson, D. (2000b). Sentiment and Value. *Ethics*, 722-748. https://doi.org/10.1086/233371

Davidson, D. (2003). Mental Events. In N. C. (ed.), *Mental Causation and the Metaphysics of Mind* (pp. 85-101). Peterborough, CA: Broadview Press.

Deonna, A. J., & Teroni, F. (2012). *The emotions: a philosophical introduction.* New York: Routledge.

Deonna, J., & Teroni, F. (2024). Emotions and Their Correctness Conditions: A Defense of Attitudinalism. *Erkenntnis*, 45-64. https://doi.org/10.1007/s10670-022-00522-0

Evans, G. (1982). *The Varieties of Reference.* Oxford: Oxford University Press.

Farkas, K., & Crane, T. (2022). Mental fact and mental fiction. In T. P. Tamas Demeter, *Mental fictionalism* (pp. 303-319). London.

Gertler, B. (2018). Self-Knowledge and Rational Agency: A Defense of Empiricism. *Philosophy and Phenomenological Research*, 91–109. https://doi.org/10.1111/phpr.12288

Gertler, B. (2021, December 21). *Self-Knowledge*. From The Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/archives/win2021/entries/self-knowledge/

Gertler, B. (2022). Agency and Self-Knowledge. In L. Ferrero, *The Routledge Handbook of Philosophy of Agency.* New York: Routledge.

Hacker, P. M. (2017). *The passions: a study of human nature.* NJ: Wiley.

Howard, C. (2023 , March 21). *Fitting Attitude Theories of Value.* From The Stanford Encyclopedia of Philosophy : https://plato.stanford.edu/entries/fitting-attitude-theories/

James, W. (1884). "What Is an Emotion? *Mind*, 188–205. http://www.jstor.org/stable/2246769

James, W. (1890). *The principles of psycholog, Vol. 1*. Henry Holt and Co. https://psycnet.apa.org/doi/10.1037/10538-000

Kloosterboer, N. (2015). Transparent emotions? A critical analysis of Moran's transparency claim. *Philosophical Explorations* , 246-258. https://doi.org/10.1080/13869795.2015.1032327

McGeer, V. (1996). Is "Self-Knowledge" an Empirical Problem? Renegotiating the Space of Philosophical Explanation. *Journal of Philosophy*, 483-515. https://doi.org/10.2307/2940837

McGeer, V. (2008). The Moral Development of First-Person Authority. *European Journal of Philosophy*, 81-108. https://doi.org/10.1111/j.1468-0378.2007.00266.x

McHugh, C., & Way, J. (2016). Fittingness-first. *Ethics*, 575–606. https://www.jstor.org/stable/26540864

Milona, M. (2016). Taking the perceptual analogy seriously. *Ethical Theory and Moral Practice*, 897–915. http://www.jstor.org/stable/24762664

Moran, R. (2001). *Authority and Estrangement.* Princeton: Princeton University Press.

Moran, R. (2012). Self-Knowledge, 'Transparency', and the Forms of Activity. In D. Smithies, & D. Stoljar, *Introspection and Consciousness* (pp. 211–236). Oxford University Press.

Moya, C. J. (2006, December). Moran on Self-Knowledge, Agency and Responsibility. *CRÍTICA, Revista Hispanoamericana de Filosofía*, pp. 3–20.

Naar, H. (2018). Sentiments. In H. N. Teroni, *The Ontology of Emotions* (pp. 149-168). New York: Cambridge University Press.

Naar, H. (2022). Skepticism about reasons for emotions . *Philosophical Explorations*, 108-123.

Naar, H. (2024). Emotions and the Action Analogy: Prospects for an Agential Theory of Emotions. *Journal of the American Philosophical Association*, 64-78. doi:10.1017/apa.2023.11

Nussbaum, M. (2004). Emotions as judgments of value and importance. In R. C. Solomon, *Thinking About Feeling: Contemporary Philosophers on Emotions* (pp. 183-199). New York: Oxford University Press .

Raz, J. (1999). *Engaging Reason: On the Theory of Value and Action.* Oxford, UK: Oxford University Press UK.

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

Sánchez, A. M., & Salice, A. (2023). Introduction. In A. M. Sánchez, & A. Salice, *Emotional Self-Knowledge* (pp. 1-13). Routledge.

Scarantino, A., & de Sousa, R. (2021, June 21). *Emotion*. From The Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/emotion/

Schlosser, M. (2019 , December 21). *Agency*. From The Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/agency/

Schwitzgebel, E. (2024, March 21). *Belief*. From The Stanford Encyclopedia of Philosophy : https://plato.stanford.edu/entries/belief/

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind.* New York: Cambridge University Press.

Sherman, N. (1999). Taking Responsibility for our Emotions. *Social Philosophy and Policy* , 294-323. doi:10.1017/S026505250000248X

Tappolet, C. (2015). Values and Emotions. In I. H. Olson, *The Oxford Handbook of Value Theory* (pp. 80-95). Oxford University Press.

Wallace, R. J. (1999). Three conceptions of rational agency. *Ethical Theory and Moral Practice*, 217-242. https://doi.org/10.1023/A:1009946911117

Wright, C. (1991). Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention. In K. Puhl, *Meaning Scepticism.* Berlin: de Gruyter.