Affective Engagement

Essays on Emotion, Value, and Normative Phenomenology

Pál-Wallin, Robert

2025

Document Version:
Publisher's PDF, also known as Version of record

Affective Engagement

Essays on Emotion, Value, and Normative Phenomenology

ROBERT PÁL-WALLIN
DEPARTMENT OF PHILOSOPHY | LUND UNIVERSITY

Affective Engagement

Emotions are simply fascinating phenomena. They constitute such a crucial dimension of life as we know it and permeate almost every aspect of our ongoing engagement with the world we inhabit. And yet they remain, in so many ways, elusive and bewildering. No wonder why the topic of emotions has been, and continues to be, such a vibrant field of philosophical inquiry. In this doctoral thesis, I investigate two key issues at the heart of the ongoing philosophical debate concerning the emotions – their *nature* and their *normativity*. Briefly put, the thesis might be said to be about the ways in which affectivity imbues life with meaning and substance, how it structures our relationships to other people and to ourselves, and the indispensable role it plays in constructing a world in which things matter.

Printed by Media-Tryck, Lund 2025
NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

The Joint Faculties of Humanities and Theology
Department of Philosophy

ISBN 978-91-90055-34-2

Affective Engagement
Essays on Emotion, Value, and Normative Phenomenology

Affective Engagement

Essays on Emotion, Value, and Normative Phenomenology

Robert Pál-Wallin

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Humanities and Theology at Lund University to be publicly defended on Saturday the 18th of October 2025 at 10:00 in LUX C126, Helgonavägen 3, Lund

Faculty opponent
Professor Antti Kauppinen (University of Helsinki)

Organization: LUND UNIVERSITY

Document name: Doctoral Dissertation **Date of issue:** 2025-10-18

Author(s): Robert Pál-Wallin

Title and subtitle: Affective Engagement: Essays on Emotion, Value, and Normative Phenomenology

Abstract: In this doctoral thesis I investigate two key issues at the heart of the ongoing philosophical debate concerning the emotions – their nature and their normativity. I argue against the deeply entrenched view according to which emotions are representational mental states and fitting in virtue of accurately representing their objects. Drawing on recent attitudinal accounts of the emotions, I develop my own view according to which emotions are *sui generis* affective modes of engagement through which we express our cares, concerns, and various commitments – i.e., elements that constitute our evaluative orientations and practical identities as agents. In light of this conception of the emotions, I go on to argue that the issue of emotional fittingness is essentially an agent-relative matter, the answer to which can only be settled by taking into consideration the full range of the particular agent's evaluative orientation.

The thesis consists of an introductory kappa and four self-contained papers, all of which touch on a variety of different issues pertaining to philosophy of emotion, value theory, normativity, moral psychology, moral responsibility, and phenomenology. The objective of the thesis can thus be seen as not only examining various isolated questions concerning the emotions, but also as making sense of the way in which such questions hang together in the broader scheme of things, things that are part and parcel of the lives we lead. In short, the thesis might be said to be about the ways in which affectivity imbues life with meaning and substance, how it structures our relationships to other people and to ourselves, and the indispensable role it plays in constructing a world in which things matter.

Key words: Affective Experience, Emotion, Empathy, Fear, Fittingness**,** Moral Responsibility, Normativity, Phenomenology, Reactive Attitudes, Reasons, Value

Language: English **Number of pages**:

ISBN (print): 978-91-90055-34-2

ISBN (digital): 978-91-90055-35-9

Affective Engagement

Essays on Emotion, Value, and Normative Phenomenology

Robert Pál-Wallin

LUND
UNIVERSITY

Cover artwork by Eva Pálová & Robert Pál-Wallin

The Joint Faculties of Humanities and Theology

Department of Philosophy

Lund University

Lund 2025

ISBN 978-91-90055-34-2 (print)

ISBN 978-91-90055-35-9 (electronic)

Printed in Sweden by Media-Tryck, Lund University,
Lund, 2025

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Till Ida & Vilja

Table of Contents

List of Papers

Paper I

Pál-Wallin, R. (2025). In Opposition to Alethic Views of Moral Responsibility. *The Southern Journal of Philosophy*. https://doi.org/10.1111/sjp.12627

Paper II

Pál-Wallin, R. (unpublished manuscript). Emotional Fittingness, Individual Nature, and Relational Values.

Paper III

Pál-Wallin, R. (manuscript under review). When Your Reasons Become My Reasons: On Internalism and Empathic Resonance.

Paper IV

Pál-Wallin, R. (2025). Fear as a Reactive Attitude. In A. Harbin (Ed.). *The Philosophy of Fear: Historical and Interdisciplinary Approaches* (pp. 107–122). New York: Bloomsbury Academic.

Acknowledgements

The end is nigh. I have to submit this thesis in just a few hours. I am severely sleep deprived and probably suffering from caffeine poisoning. But I still have one crucial task ahead of me – a task I have been postponing for as long as I can, for a very simple reason: It is painful to accept the fact that this journey is coming to an end. Oh, what an absolute joy it has been. But the time has come. This thesis is about emotions, well now it's time to show some. Here we go.

First of all, I need to express my heartfelt gratitude to my three supervisors Matthew Talbert, Paul Russell, and Toni Rønnow-Rasmussen. I could not have done this without your wonderful support and encouragement.

Matt, you have been a fantastic main supervisor and mentor. Your genuine interest in my work, your insightful thoughts, and your sharp feedback has been indispensable in the development of this project. With your guidance, support, and patience, I feel like I have flourished and matured as a philosopher. Thank you for the all the engaging discussions and the friendship. It has been a privilege.

Paul, thank you for welcoming me to the Lund Gothenburg Responsibility Project with open arms. You have shown such faith in both me and my project, and been a great source of encouragement. Also, thank you for showing me that philosophy is not merely an enterprise of technical navel-gazing, but that it is the task of the philosopher to see the bigger picture.

Toni, you may have been my second supervisor during my doctoral studies, but you were my main supervisor both when I wrote my bachelor thesis and when I wrote my master's thesis. It is in large part because of you that I came to see the beauty in philosophy, and why I chose to pursue a PhD. You have been there for me through thick and thin, both in the capacity of a philosopher and a friend.

Thank you for the well-needed therapy sessions. You are truly a fitting target of my deepest esteem and gratitude.

I have had the utmost pleasure to share an office with the best colleagues one can hope for. Anton Emilsson, Marta Johansson Werkmäster, Jiwon Kim, Marianna Leventi, Elsa Magnell, Gloria Mähringer, Martin Sjöberg, and Alexander Velichkov, I feel so blessed and honored to have had the opportunity to pursue this PhD alongside you. Thanks for the memories, the laughs, the deep talks, the support, and the whisky. Thank you all for being such amazing friends.

A special thanks is owed to Alexander Velichkov for taking the time to read and comment on large parts of this thesis at a very late stage in the process, and to Anton Emilsson for providing excellent feedback at a very late stage in the process, and for yelling at me over the phone and giving me that last encouraging push I so desperately needed.

I want to thank all the amazing people that have been part of, or in some way affiliated with, the Lund Gothenburg Responsibility Project. I am so grateful for having been part of such a vibrant research group with so many talented philosophers. In addition to those already mentioned, I want to thank Gunnar Björnsson, Olle Blomberg, Hadi Fazeli, Yuliya Kanygina, Shervin Mirzaeighazi, Björn Petersson, Wlodek Rabinowicz, András Szigeti, Daniel Telech, Patrick Todd, and Caroline Torpe Touborg.

My philosophical understanding has evolved in large part due to the many interactions I have had with the brilliant people at the department. Several of them I have already mentioned, but I also want to thank Niklas Dahl, Andrés Garcia, Mattias Gunnemyr, Frits Gåvertsson, Max Minden Ribeiro, and Jakob Werkmäster, for the many engaging discussions over the years from which I have learned a lot.

I would also like to extend a special thanks to Dan Egonsson, who was the very first teacher I had when I took the introductory course to practical philosophy in Lund. Your passion and openness got me hooked on philosophy. Thank you for your kindness and warm friendship.

Many more than those already mentioned have contributed with conversations concerning many ideas that have shaped this thesis in several respects; at seminars

and conferences, over lunch, fika and dinner, in bars and on walks, you have all contributed to a very stimulating intellectual environment. In this respect, I wish to thank David Alm, Dani Alnashi, Henrik Andersson, Trond Arild Tjøstheim, Balder Ask Zaar, Monika Betzler, Thibault Boehly, Mark Bowker, Fanny Chef Holmberg, Ellen Davidsson, Ylva von Gerber, Anders Herlitz, Hubert Hågemark, Ingvar Johansson, Pierre Klintefors, Sarah Koeglsperger, Patrizio Lo Presti, Jenny Magnusson, Ana Maria Mora Marquez, Maybí Morell Ruiz, August Olsen, Erik J. Olsson, Erik Persson, Signe Savén, David Shoemaker, Samantha Stedtler, Jakob Stenseke, Robin Stenwall, Andreas Stephens, Alexander Tagesson, Matthew Tompkins, Melina Tsapos, Gabriel Vogel, Annika Wallin, Anton Wrisberg, Anthony Öhrnström, and Fredrik Österblom.

This PhD would not have been possible without the administrative work and generous support of Kristina Arnrup Thorsbro, Anna Cagnan Enhörning, Tobias Hansson Wahlberg, Petter Johansson, Martin Jönsson, Tomas Persson, Eva Sjöstrand, Annah Smedberg-Eivers, and Anna Östberg. Thank you so much for all your help, your kindness, and your patience with me when I had no idea what I was doing.

In 2022, I had the great fortune to spend a month in Geneva with the Thumos Research Group. A special thanks to Julien Deonna and Fabrice Teroni for making my research visit possible and for providing me with invaluable feedback on my work. Agnés Baehni, thank you for the wonderful friendship, for all the lovely conversations, and for being such a reliable party companion. I also wish to extend my heartfelt gratitude to Roberto Keller for making my visit in Geneva so enjoyable, and for the affability with which you made me feel like I belonged. Thanks also to Angela Abatista, Radu Bumbacea, Mathilde Cappelli, Florian Cova, Steve Humbert-Droz, Michele Ombrato, Antoine Rebourg, Laura Silva, and Juliette Vazard.

My paper "Fear as a reactive attitude" would perhaps never have seen the light of day if it hadn't been accepted by Ami Harbin to be included in *The Philosophy of Fear: Historical and Interdisciplinary Approaches*. For that, I am immensely grateful. Thanks also to Rodrigo Diaz and Christine Tappolet for providing such great feedback on the draft.

A special thanks is owed to my formidable mock opponent, Hichem Naar. Thanks for engaging so affectively with my work and for the excellent feedback you provided. Thanks for the lovely conversations. It was a pleasure to get to know you.

I am incredibly grateful to Gunilla Albertén at Media-Tryck for all her help with the production of this thesis. Thank you for your patience and for being so considerate.

Music is my oxygen, and I am so lucky to be able to play with such talented musicians and fantastic friends. In this respect, I want to thank Leo Mandin, Oliver Olsson, Viktor Rinneby, and Pontus Örnstrand.

Undoubtedly, I would not have been able to complete this thesis without the love, support, and encouragement from my wonderful family. For that, I wish to thank my parents Eva Pálová, Andreas Holmgren, and Per Borgström, my brother Oliver Holmgren, and my extended family Stene Wallin, Karin Wallin, Emma Wallin, Johan Wallin, Lisa Gard, Tobias Gard, Hugo Gard, and Daniel Gard. Thank you all for your unwavering belief in me.

Ragnar, you wonderful, crazy dog. Always willing and ready to cuddle. Thank you for comforting me when I need it, and for reminding me that sometimes, going for a walk is more important than philosophy.

Ida and Vilja, my beautiful sweethearts. You are everything. I would have never completed this thesis if it weren't for you Ida. Your unparalleled encouragement, assistance, comfort, and love has made it all possible. Thank you for your extraordinary patience with me during my moments of utter despair, and for providing me with the emotional and practical support I needed to persevere. You are my bedrock. Vilja, I am so immensely proud to be your father. You endow my life with sheer bliss, and you have taught me more about emotions than philosophy ever could. Oh, one more thing Vilja, please return my copy of *Ethics and the Limits of Philosophy* when you are done with it. I love you both so damn much.

Malmö, September 2025

Abstract

In this doctoral thesis I investigate two key issues at the heart of the ongoing philosophical debate concerning the emotions – their nature and their normativity. I argue against the deeply entrenched view according to which emotions are representational mental states and fitting in virtue of accurately representing their objects. Drawing on recent attitudinal accounts of the emotions, I develop my own view according to which emotions are *sui generis* affective modes of engagement through which we express our cares, concerns, and various commitments – i.e., elements that constitute our evaluative orientations and practical identities as agents.

In light of this conception of the emotions, I go on to argue that the issue of emotional fittingness is essentially an agent-relative matter, the answer to which can only be settled by taking into consideration the full range of the particular agent's evaluative orientation.

The thesis consists of an introductory kappa and four self-contained papers, all of which touch on a variety of different issues pertaining to philosophy of emotion, value theory, normativity, moral psychology, moral responsibility, and phenomenology. The objective of the thesis can thus be seen as not only examining various isolated questions concerning the emotions, but also as making sense of the way in which such questions hang together in the broader scheme of things, things that are part and parcel of the lives we lead.

In short, the thesis might be said to be about the ways in which affectivity imbues life with meaning and substance, how it structures our relationships to other people and to ourselves, and the indispensable role it plays in constructing a world in which things matter.

1. Introduction

Emotions are simply fascinating phenomena. They constitute such a crucial dimension of life as we know it and permeate almost every aspect of our ongoing engagement with the world we inhabit. And yet they remain, in so many ways, elusive and bewildering. No wonder why so many prominent philosophers throughout history – thinkers like Aristotle, René Descartes, Baruch Spinoza, David Hume, Adam Smith, Friedrich Nietzsche, Edith Stein, and Jean-Paul Sartre to name a few – were all so intrigued by the emotions, and keen on making sense of them. And still to this day, the topic of emotions continues to be a flourishing and vibrant field of philosophical inquiry.

The issues concerning the emotions that are typically raised and investigated in the contemporary philosophical literature are multifarious. Some issues are of a *metaphysical* or *ontological* flavor – they raise questions about the nature of emotions, what they consist in, and how they are structured. Other issues are *epistemological*, and raise questions about whether emotions can, or in what ways they can, provide us with epistemic goods like knowledge, or understanding. Philosophers are also concerned with various *normative* and *ethical* questions regarding the emotions. Are emotions liable to various kinds of normative, or moral, assessments? Can emotions be rational? Appropriate? Fitting? Can we have normative reasons to have emotions? Can we be responsible for our emotions?

This is but a minor sample of the numerous questions that are at the heart of the ongoing debate in the contemporary philosophical study of the emotions. My aim with this doctoral thesis – which consists of an introductory kappa and four self-contained papers – has been to contribute to that debate by investigating both the *nature* and the *normativity* of the emotions. This may sound like a fairly isolated

enterprise, but matters in philosophy tend to interlock and hang together in a spectacular, if not daunting, mosaic. A seemingly narrow philosophical inquiry can thus rather quickly prompt all sorts of unexpected questions. And that is why – as the reader may come to notice – the essays in this thesis touch on a variety of different issues pertaining to not only the philosophy of emotion, but also value theory, normativity, moral psychology, moral responsibility, and phenomenology.

In "Philosophy and the Scientific Image of Man" (1962), Wilfrid Sellars claims that it is the aim of philosophy to "understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term" (p.1). It would certainly be an exaggeration on my part to claim that this thesis successfully lives up to that aim, but I can in good conscience contend that my ambition has been to make sense of how issues pertaining to various areas of philosophy hang together. In that sense, my aim has been, as Sellars puts it, to have an "eye on the whole" (ibid., p.2). The objective of the thesis can thus be seen as not only examining certain questions concerning the emotions, but also as an attempt at making sense of the ways in which such questions hang together in the broader scheme of things, things that are part and parcel of the lives we lead.

But of course, one has to begin somewhere. For me, the key question that has been at the forefront of my mind during my years as a doctoral candidate, guiding my philosophical inquiries and the process of writing these essays, is the question of how we should think about the *fittingness* of emotions. It is a familiar fact that we sometimes treat our emotional reactions as *inappropriate, unwarranted,* or *out of place* with respect to their objects and the circumstances in which they are elicited. Other times, our emotional reactions are taken to be perfectly apt, or as philosophers may say, *fitting*.

The notion of fittingness here is commonly characterized as referring to a distinct type of *normative* relation that is said to hold between an attitudinal response and an object when that object in some sense *merits*, *deserves*, or is *worthy of* that response. But how precisely should we understand the idea that an object can

merit, or be worthy of, a certain kind of response? Under what conditions does the fittingness relation hold between an attitude and an object? Are the fittingness conditions the same for all kinds of attitudes (beliefs, emotions, desires, intentions, etc.)?

Presumably, since our attitudes differ from one another in various ways, we should expect the fittingness conditions of these various attitudes to differ as well. Indeed, it seems rather intuitive to think that the fittingness of an attitude should flow from the nature of that attitude-type. If that is correct, then the answer to the question under what conditions emotions are fitting with respect to their objects turns on the answer we give to the question of what emotions are.

My own approach to these questions has in large part been molded by a dissatisfaction with a deeply entrenched way of thinking about the emotions in the philosophical literature – a view which I shall here refer to as *Robust Representationalism.* This view embodies the conjunction of two distinct claims – one ontological and one normative – about the emotions:

Robust Representationalism

The ontological claim:	Emotions are, at least in part, representational mental states with a mind-to-world direction of fit.
The normative claim:	A token emotion is fitting exclusively in virtue of the fact that it represents things accurately.

It is worth noting that it is perfectly plausible to reject the normative claim while endorsing the ontological claim. One cannot, however, consistently endorse the normative claim without also endorsing the ontological claim. If we embrace the idea that emotions are fitting entirely in virtue of the fact that they represent things accurately, then we simply cannot deny that emotions are, at least in part, representational mental states.

What all of this means will become clearer as we proceed through the subsequent introductory chapters. To that end, in what follows I shall provide some theoretical background that will help the reader get a better grip on the subject matters of this thesis.

2. The Nature of Emotions

Emotions are mental phenomena. But what kind of mental phenomena are they? In particular, one might ask in what way emotions differ from other mental states such as beliefs, perceptions, desires, moods, etc. Although we might have an intuitive grasp of the differences between these things, philosophical reflection can often shatter our common-sense understanding and reveal the difficulties in pinpointing precisely how these mental phenomena come apart. My point of entry here shall be to draw attention to the concept of *intentionality*.

2.1. Intentionality

There is widespread agreement among philosophers that emotions are *intentional* mental phenomena (D'Arms 2024; Deonna and Teroni 2012; Echeverri 2019; Müller 2021; Scarantino 2014; Scarantino and De Sousa 2021; Teroni 2007).[1] This does *not* mean – as one might suppose – that token emotions are deliberately brought about, i.e., formed *with intention*, or *at will*. Our episodes of anger, fear, sadness, gratitude, and joy, are not chosen in the way that one might choose to have a second cup of coffee.[2] Rather, to claim that emotions are intentional mental states is simply to claim that emotions exhibit *intentionality*, i.e., a feature of

[1] There are of course philosophers who reject the idea that emotions are necessarily intentional (see e.g., Shargel 2015)

[2] Emotions are thus not under our *voluntary control* (see e.g., Smith 2005). This is not to deny that we can choose to place ourselves in circumstances where it is likely that we will succumb to a certain emotional experience. I might for example choose to listen to a song which reminds me of a sad event in my past and as a consequence feel sad again. Accordingly, we do enjoy *indirect control* over our emotions.

mental states by which they are *directed at* or *about* various things in the world (Searle 1983). To illustrate, consider the following examples that may occur in everyday life:

(i) Alex fears the dog in front of him
(ii) Jenny is angry with Marcus
(iii) Rami admires the Garden of Earthly Delights
(iv) Robert is afraid that he will fail his thesis defence

In all of these examples, the subjects' emotional states are directed at specific things in the world. Alex's fear is directed at a non-human animal, Jenny's anger is directed at an individual person, Rami's admiration is directed at an inanimate object (a triptych painting by Hieronymus Bosch), and Robert's fear is directed at a potential state of affairs. These particular things to which our token emotions are directed are often referred to as the emotion's *intentional object* (Echeverri 2019; Massin 2021), or *particular object* (Teroni 2007).

The idea that emotions are necessarily intentional – in the sense of having intentionality – is sometimes met with immediate suspicion. It is after all a familiar fact that some of the emotional states we endure just don't seem to be directed at, or about, anything at all. People may occasionally find themselves in a state of melancholy, or joy, or anxiety, without being melancholic, joyful, or anxious *about* anything in particular. Is this not a clear-cut counterexample to the idea that emotions are intentional mental phenomena? A standard reply here is to invoke the distinction between *emotions* and *moods*.

Emotions and moods are closely related affective mental phenomena, but whereas emotions are always directed at various things in the world, moods are not – or so the general thought goes. The putative counterexample to the idea that emotions are necessarily intentional can thus be answered by maintaining that insofar as the

aforementioned states of melancholy, joy, and anxiety are *de facto* not directed at anything, they are not emotions, but moods.[3]

There is another aspect to the intentionality of emotions worth highlighting here. Philosophers sometimes talk about the *formal objects* of the emotions (see e.g., Kenny 1963; Teroni 2007; De Sousa 2007). The formal object of an emotion is generally said to be the *value* or the *evaluative property* that is shared by all instances of the same emotion kind. What does this mean?

Let's illustrate with fear as an example. Instances of fear can be about, or directed at, all kinds of things in the world – snakes, clowns, job interviews, economic policies, and global warming among other things. Presumably, there should be some feature or property that all of these different things could exemplify and which all instances of fear can be said to be about. A common suggestion in the case of fear is the relevant kind of property is the property of being *dangerous* or *threatening*.

Correspondingly, for each type of emotion there should be some kind of value or evaluative property that all instances of that emotion kind can be said to be about. In the case of anger, the relevant kind of property is arguably the property of being offensive. In the case of sadness, it might be the property of being a loss. In other words, all token emotions can be said to be about two types of objects, namely a *particular object* or *intentional object* (the concrete thing in the world), and a formal object (the evaluative property that all instances of an emotion kind can be said to be about). We can illustrate this by reusing our previous examples:

[3] Another thing that might be said in response to doubts one might have with respect to the claim that emotions are intentional is that even though it may *seem* like a token emotion is not directed at anything, it does not necessarily mean that it is not. It does seem like an open possibility that our emotions could be directed at things in the world even though it may not seem to us that way.

(i) Alex's fear has an intentional object (*the dog in front of him*) and a formal object (the property of being *dangerous*)

(ii) Jenny's anger has an intentional object (*Marcus*), and a formal object (the property of being *offensive*)

(iii) Rami's admiration has an intentional object (*the Garden of Earthly Delights*), and a formal object (the property of being *spectacular*)

(iv) Robert's fear has an intentional object (*the state of affairs in which he fails his thesis defence*), and a formal object (*the property of being threatening*)

2.2. Direction of Fit

Appealing to intentionality may be a useful way of distinguishing between emotions and moods, but it is important to underscore that intentionality is not a unique feature of the emotions. Our mental repertoire comprises a number of different kinds of attitudes that all seem to exhibit the property of being directed at, or about, various things.

For instance, *beliefs* are commonly said to be directed at, or about, the *content* or *proposition* that is believed. Suppose that you believe that a water molecule consists of one oxygen atom and two hydrogen atoms, or that Bratislava is the capital of Slovakia. In the former case, your attitude of *believing* is directed at, or about, the proposition *a water molecule consists of one oxygen atom and two hydrogen atoms*. In the latter case, your attitude of believing is directed at the proposition *Bratislava is the capital of Slovakia*.[4]

[4] What propositions are, more precisely, is of course a matter of ongoing philosophical debate. We can note however that it is quite common to characterize propositions as, for example, "the primary bearers of truth and falsity" (Merricks 2009, p.207), or as "the referents of *that*-clauses" (McGrath and Frank 2024). Delving deeper into the question of the nature of propositions is a can of worms we need not open here.

Similarly, it is ordinarily thought that our *desires* are directed at various things in the world, namely those things that we happen to desire. I may for example desire some chocolate ice cream, or that I become a professor in philosophy. In both cases, my desire is directed at, or about, *the thing which I desire*, namely *chocolate ice cream*, or that *I become a professor in philosophy.*[5]

In light of the fact that these different mental states all seem to have the property of being intentional, it is reasonable to ask in what way we are supposed to distinguish between them. It is at this juncture that the notion of *direction of fit* may be of service, at least to some extent.

The guiding thought here is that we can differentiate between mental states in terms of whether they have a *mind-to-world direction of fit* or whether they have a *world-to-mind direction of fit* (see e.g., Anscombe 1957; Goldie 2000; Deonna & Teroni 2012). Mental states like *beliefs* are said to have a mind-to-world direction of fit because there is a sense in which beliefs should conform to the way the world actually is, i.e., beliefs are supposed to *fit the world*. In contrast, *desires* are said to have a world-to-mind direction of fit because there is a sense in which desires reveal to us how the world is 'supposed to be'. With respect to desires then, it is the world that is supposed to conform to the way that the desire 'suggests' it should, i.e., the world should *fit the (desiring) mind*.

Where then do emotions fit in the direction of fit schema? Are they on the side of belief? Or are they on the side of desire? Or are they perhaps on neither side? Truth be told, the question is very much a matter of ongoing debate. Some philosophers have argued that emotions are more like beliefs in this regard, and that our emotions should conform to the way that the world actually is. That is, they maintain that emotions have a mind-to-world direction of fit. Other philosophers have been reluctant to think about the emotions on the model of

[5] Philosophers may have different opinions about the chocolate ice cream case. Some would argue that, strictly speaking, I do not desire *chocolate ice cream*. What I desire is *that I have some chocolate ice cream.* In other words, they would maintain that, properly understood, the intentional object of my desire – what my desire is directed at – is not chocolate ice cream in itself, but the *state of affairs* in which I have chocolate ice cream.

belief and instead proposed that we should think of emotions as having perhaps neither direction of fit (Yang 2016), or maybe both directions of fit.

At any rate, it seems to me that the kind of answer one is prepared to give here will ultimately be a function of one's broader way of thinking about what emotions consist in and what sort of role they play in our relation to the world. Let us therefore proceed by taking a look at some prominent philosophical theories of what emotions are, i.e., what emotions *consist in*.

2.3. Theories of the Emotions

There are many different theories or accounts of the nature of emotions available in the philosophical literature. What these accounts aim to do is to provide explanations of what the emotions consist in and how they are structured, i.e., their ontology. Providing a taxonomy of these different theories is often easier said than done. There are several explanations for why that might be the case. First and foremost, there is often some overlap between accounts of different kinds. Account A may share some feature X with account B which is not shared by C. In turn, account B might share some feature Y with account C which is not shared by A, and account C might share some feature Z with account A which is not shared by B, and so on. Moreover, it is not uncommon for philosophers to develop their accounts by drawing inspiration from several different sources to the point where it is no longer clear in which category to place them. Nevertheless, in what follows I have tried to provide a rough overview of five different families of accounts that one might come across in the literature:

Cognitive theories:	Emotions are essentially constituted by *evaluative judgments*, *beliefs*, or *thoughts* that represent their intentional objects as instantiating a certain kind of value or evaluative property. Notable proponents: Greenspan (1988), Nussbaum (2001, 2016), Solomon (1973).
Perceptual theories:	Emotions are *perceptual experiences* by which the emotions' intentional objects are represented as instantiating a certain kind of value or evaluative property. Notable proponents: Döring (2003, 2007) Milona (2023), Prinz (2004), Tappolet (2016).
Feeling Theories:	Emotions are identified as *feelings* with a certain phenomenal character or hedonic quality. While some have argued that these feelings as evaluative, others have argued that they are feelings of bodily changes, e.g., physiological and somatosensory changes. Notable proponents: Goldstein (2002), Helm (2001), James (1884), Lange (1885).
Attitudinal /Reactive theories:	Emotions are specific kinds of *evaluative attitudes*, *modes,* or *reactions* that are (in)correct as a function of whether their intentional objects instantiate the relevant formal object. Notable proponents: Deonna & Teroni (2012, 2014, 2015, 2022), Massin (2021), Müller (2017, 2018), Zamuner (2015).
Enactivist theories:	Emotions are *embodied action-oriented processes* that are dynamically shaped in the relation between the subject and the external world. Notable proponents: Colombetti (2010), Hutto (2012), Shargel & Prinz (2018).

There are undoubtedly many interesting things to discuss concerning feeling theories and enactivist theories. However, in what follows I shall have to leave these theories behind in order to pay a bit more attention to cognitivist theories and perceptual theories, and to some extent attitudinal theories. The motivation for this is simple: Both cognitivist theories and perceptual theories maintain that emotions *represent* their intentional objects in a certain way. Both theories are thus committed to the ontological claim of the view I referred to in the introduction as *Robust Representationalism* – a view with which I explicitly declared my dissatisfaction. That is why I wish to focus on these theories.

We may recall that the ontological claim of *Robust Respresentastionalism* stated that emotions are, at least in part, representational mental states with a mind-to-world direction of fit. Cognitivists tend to cash out the way in which emotions are representational mental states in terms of some *doxastic* mental state, e.g., a *judgment* or a *belief*. In contrast, perceptualists maintain that emotions are representational mental states in virtue of being *perceptual experiences*. On this sort of picture, emotions are taken to represent their intentional objects as having certain evaluative properties by being perceptual experiences of those evaluative properties.

Cognitive theories and perceptual theories are both varieties of a more general view of the emotions called *representationalism* (see Deonna & Teroni 2022; Naar 2022). It is important to note here that this is not tantamount to the view I call *Robust Representationalism*. Whereas representationalism is merely the ontological thesis that emotions are representational mental states, what I call *Robust Representationalism* is, as we have seen, a view that combines the ontological claim with a normative claim about the emotions. All proponents of cognitive theories and perceptual theories are proponents of representationalism, but they are not necessarily proponents of *Robust Representationalism* since they need not accept the claim that emotions are fitting *exclusively* in virtue of the fact that they represent their intentional objects accurately.

We will take up the issue of fittingness in the next chapter, but before we do so I think we should perhaps place the ontological claim under the microscope for a moment and raise the question what reasons there may be for accepting it. What is the motivation for thinking that emotions are representational mental states?

2.4. Are Emotions Representational Mental States?

At the beginning of this chapter, we noted the fact that emotions are intentional mental phenomena. A theory of the emotions should thus be able to account for that fact. Can representationalism – the view that emotions are representational mental states – do so? As Hichem Naar (2022) notes, the fact that emotions have intentionality is often taken as the "central pre-theoretical motivation for representationalism" (p.2). This is so for a very simple reason: All representational mental states represent their objects as being a certain way, and this necessarily entails that the mental state is *about* its object – it is simply "not possible for a mental state to be representational without being intentional" (Naar 2022, p.2). Accordingly, representationalism about the emotions can easily account for the fact that emotions are intentional. The crucial question is of course whether we *need* to accept representationalism about the emotions in order to account for emotional intentionality. Allow me to say a couple of things to suggest that we do not.

There is widespread agreement among philosophers of emotion that our emotional episodes depend – in some way – on other mental states, like perceptions, imaginations, beliefs, etc. For example, my fear of a grizzly bear seems to depend on me either *seeing* the bear, *imagining* the bear, *thinking* about the bear, or *remembering* the bear, etc. Those mental states on which our emotions depend are commonly referred to as *cognitive bases* (see e.g., Mulligan 1998; Deonna & Teroni 2012, 2014; Naar 2020; Milona & Naar 2020). As Deonna and Teroni point out; "emotions must latch on to information provided by other mental states" (2012, p.5), and the notion of a *cognitive base* is a technical term

referring to any type of mental state (belief, memory, perception, etc.) that may carry this sort of information. Similarly, Hichem Naar writes that:

> By contrast with perception, emotion is not an independent way of accessing its intentional object. In order to be afraid of something, I first need to see that thing, or to believe that it is there, or at least to think of it in some way. In other words, I need to have *a prior grasp* of it, a grasp which is not given by my fear (Naar 2020, my italics).

Now, given the plausible assumption that emotions depend on other mental states, or must latch onto information provided by other mental states, and given the quite plausible assumption that the kinds of mental states that *can* provide such information *are* representational mental states, one might wonder why we must posit that emotions too are representational.

A natural response for the representationalist is to maintain that we need to posit that emotions are representational because of the simple fact that emotions *themselves* are intentional. The fact that emotions depend on other representational mental states does not mean that emotions do not have their own intentionality that needs to be accounted for.

In response to this, anti-representationalists might perhaps push the claim about dependence a bit further and point out the possibility that the intentionality of an emotion perhaps is a feature that the emotion possesses *derivatively* by *inheriting* it from the mental representation on which the emotion depends (its cognitive base). One might argue that the antecedent mental representation 'points' or 'directs', as it were, the emotion towards objects in the world, but the emotion does not represent these objects. In that way, the intentionality of the emotion is explained without positing that emotions themselves are representational mental states. But these remarks are only tentative, I must admit.

Representationalists might also point out that even though it may be the case that our emotions depend on antecedent mental representations, we still need to posit that they are representational mental states because their representational content

is substantially different from the representational content of their cognitive bases. They might claim that although the representational content of cognitive bases provides us with information about the *descriptive* features of the intentional object, emotions provide *an additional layer of representation* which is essentially *evaluative*. Against this, I agree with Schroeter, Schroeter and Jones, who have argued that the burden of proof lies with the representationalist to show that "construing emotions as representing specific evaluative properties is part of the best, most economical explanation of the psychological processes underlying perception, behavior and cognition" (2015, p.363). To put it differently, we might say that if we are to accept the idea that emotions have evaluative representational content, then it must be shown that attributing such a content to the emotions is indispensable in a full explanatory picture of emotional phenomena (Schroeter, Schroeter & Jones 2015).

Is it indispensable to attribute evaluative representational content to emotions? Here is a reason to think that it is not: Although it is apparent that the types of mental states on which emotions depend – i.e., their cognitive bases – often represent objects in purely descriptive terms, it is, I contend, equally apparent that these mental states can represent objects in evaluative terms as well. For example, the mental states of *believing*, *judging*, *imagining*, and *perceiving,* etc., all seem to be types of mental entities capable of carrying evaluative content. That is, we can *believe*, *judge*, *imagine*, and *perceive* an object X to be dangerous, offensive, amusing, etc.[6] If this is correct, then there does not seem to be any decisive reason for attributing *evaluative content* to the emotions. We already have mental states in our 'mental arsenal' that are capable of representing things in the world under an evaluative guise, as it were. In conjunction with the claim that emotions are

[6] I am not assuming that the mental state of *believing, judging, imagining* or *perceiving* X as dangerous, offensive, or amusing, etc., necessarily involves mastery and deployment of the corresponding evaluative *concepts*. In other words, even if a token emotion does depend on some antecedent evaluative representation of its intentional object, the evaluative representation need not implicate the use of linguistically articulated concepts. Thus, one's capacity to feel afraid of, or amused by some X, need not depend on one's familiarity with the concepts of 'dangerous' or 'amusing'.

dependent on other mental states, it could be argued that when we experience an emotion, it is because of an antecedent evaluative apprehension of the intentional object which we then respond to emotionally. This would then seem to render the attribution of evaluative representational content to emotions explanatorily superfluous.

To be sure, these remarks do not carry the implication that emotions are not evaluative phenomena *in any sense*. The suggestion is merely that emotions seem to play a distinctive (evaluative) role in our mental economy; a role which cannot be assimilated to a mind-to-world directed representation. This would then set emotions apart from paradigmatic forms of representational mental states. Jean Moritz Müller (2017, 2018) endorses a view along these lines by way of arguing that emotions are *position-takings* towards apprehended values. Thus, on his view, emotions do not represent evaluative properties. Instead, they are modes of *acknowledging* the evaluative features of their objects. Likewise, Olivier Massin (2021) has recently defended a reactive theory of the emotions according to which emotions are attitudinal reactions to apprehended values.

Another reason one might think that emotions are representational mental states has to do with the following desideratum that a view of the emotions arguably should be able to account for: Emotions are liable to normative assessment, that is, we typically judge our emotions as (in)appropriate, or (un)warranted, or (un)fitting with respect to their objects and the circumstances in which they are elicited. Here, representationalism can be said to offer a neat explanation for this aspect of our emotional lives. Since emotions are representational mental states and represent their objects under an evaluative guise – i.e., they represent their objects as instantiating evaluative properties – emotions can be said to be fitting or unfitting depending on whether they represent their object accurately.

In order to deal with this issue properly, we need to turn to the issue of the normativity of emotions.

3. The Normativity of Emotions

Our emotional reactions are not simply taken as brute facts. We typically see these affective responses as somehow liable to normative assessment. Sometimes our reactions seem perfectly appropriate towards their objects and in the circumstances in which they are elicited, and sometimes they strike us as entirely out of place or inappropriate. Think of someone who is completely terrified by a cocker spaniel puppy. The felt terror will typically strike us – and perhaps even the terrified agent herself – as utterly misplaced, as there is no looming threat posed by the tiny creature. Or think about a case where your friend shows up twenty minutes late for the lunch meeting you had planned. You certainly have some cause to be a bit annoyed, but lashing out in full-blown rage seems way over the top. Here too, your emotional reaction is inappropriate. The philosophically interesting question is of course how we should make sense of this aspect of our emotional lives. What is it for an emotion to be appropriate or inappropriate with respect to its object?

3.1. Fittingness: A Distinct Normative Relation

It is common practice among philosophers who work on issues pertaining to the normativity of emotion to distinguish between the various senses in which emotions can be said to be *appropriate* with respect to their intentional objects (see e.g., D'Arms & Jacobson 2000a, 2000b; Deonna & Teroni 2012; Scarantino & de Sousa 2021). In one sense of the term, an emotional reaction can be deemed appropriate in virtue of it being supported by *prudential* considerations, as the case may be if the emoting agent's well-being or self-interest would be promoted

by having the relevant emotion. Conversely, emotions can also be deemed appropriate in the sense of being supported by *moral* considerations. For example, it might be the case that having a particular emotional reaction could result in a morally good/desirable outcome. Alternatively, there might be some *deontic* considerations that count in favor of having the emotion, such as it being morally *right* to have it, or morally *required* to have it. In both of these cases, there is a sense in which the emotion in question can be deemed appropriate, precisely in virtue of being supported by prudential or moral considerations respectively.

However, as D'Arms and Jacobson (2000b) pointed out more than two decades ago, these assessments of an emotion's appropriateness can come apart from the emotion's *fittingness.* The notion of *fittingness* is widely understood as referring to a distinct normative relation that is said to hold between an attitude and an object when that object, in some way, *deserves*, *merits*, *is worthy of,* or *is an apt target of* that response (see e.g., Howard, 2018). Fittingness is thus a "paradigmatically normative notion" (Howard, 2018 p.2) which delineates a normative dimension along which our attitudes can be assessed with respect to how well they 'fit' their intentional objects.[7]

Let us briefly illustrate how the fittingness of an emotion can come apart from the various ways in which it can be deemed prudentially or morally (in)appropriate. Imagine that you are on a hike in the woods, and all of a sudden you come face to face with a brown bear. The bear starts to display signs of aggression, growling and baring teeth. Would it be fitting for you to feel fear? Well, if the bear really is getting ready to attack you and you still have a lust for life, then yes, fear would be a fitting response on your part. However, it could very well be the case that if you were to become frightened, it would actually impede your chances of escaping the situation in one piece. So, prudentially speaking, feeling fear may be inappropriate in these circumstances.

[7] It is worth noting that not all philosophers in the literature use the term 'fittingness' to refer to this specific normative relation. For example, Deonna & Teroni (2012, 2022) tend to speak of *correctness* rather than *fittingness*.

How can fittingness come apart from moral appropriateness? Suppose that you have been abroad on a business trip and that your partner has promised to pick you up from the airport on the day of your arrival. Come arrival day, your partner has forgotten all about it and is out drinking with friends. Being angry with your partner in this case may be entirely fitting. But suppose that your partner is exceedingly sensitive about letting other people down and not living up to promises. If you become angry with your partner, there is a good chance that they will be in an excruciating state of pain for a whole month. It may be, in these circumstances, morally inappropriate to be angry, even though it may be fitting.

A central issue for many philosophers who are interested in the emotions, value theory, and normativity, is the question of how to understand the fittingness relation more precisely. A widespread idea is that we can understand it in terms of *reasons*.

3.2. Fittingness in Terms of Reasons

The language of reasons is ubiquitous in our lives. When you ask yourself 'What should I do?' or 'What should I believe?', it is highly likely that you will try to settle the question by deliberating about what reasons there may be for doing this or that, or what reasons there may be for believing this or that. It is common practice among philosophers to distinguish between *normative reasons*, *motivating reasons*, and *explanatory reasons*. Briefly put, a *normative reason* to act in a particular way, or to believe a particular proposition, is standardly construed as "a consideration that counts in favor of it" (Scanlon 1998, p.17). A *motivating reason* can be seen as a consideration in light of which an agent is motivated to act in a certain way, or to believe a particular proposition (see Müller 2025). Finally, an *explanatory reason* is a fact that provides an explanation for why an agent acted in the way she did, or why she believes some proposition.

To illustrate the difference between these reasons, imagine that a friend of yours – not a very close friend – is having a birthday party. You have promised that you

will turn up, although you don't really feel like going. A mutual friend who plans on going to the party tells you that there will be an abundance of free alcohol. You have been feeling quite down all week and could really use a drink (or eight), so now you feel motivated to go. Alas, when you arrive at the party, there is no alcohol at all! The mutual friend was just bullshitting.

In this scenario, there are a couple of considerations that seem to count in favor of the action of *going to the party*. It is after all a friend of yours and you have been personally invited to her birthday party. In addition, you did promise that you would come. These facts are normative reasons for you to show up to the party. Your motivating reason for going however is the putative fact that *there is free alcohol at the party* – that is what actually motivates you to go. The thing is that the consideration in light of which you are motivated to go to the party – that there is free alcohol – is not a fact that can be cited as an explanation for why you actually showed up to the party, because there was, as a matter of fact, no free alcohol. So, the proper *explanation* – that is, the explanatory reason – for why you went to the party must be that you *believed* that there would be free alcohol. That is the fact that we can appeal to as a proper explanation for how you acted.[8]

Similarly, it is widely thought among philosophers that there are reasons for emotions. Not only explanatory reasons – in the sense that we can provide explanations for why a particular agent happens to be experiencing a particular emotion. But that there are *normative reasons* for emotions, that is, considerations that *count in favor* of having them.

Consequently, it has been suggested that the fittingness relation – the normative relation we appeal to when we judge that an object merits a certain response – can be analyzed in terms of normative reasons. For example, one suggestion would be to spell out the fittingness of an emotion in terms of *sufficient* reasons. Another suggestion would be to do so in terms of *decisive* reasons. Formulations of that kind might look something like this:

[8] For an overview of the philosophical debate over reasons for action, see Alvarez (2017).

Sufficient Reasons: A token emotion E *vis-à-vis* X is fitting in virtue of the fact that there are sufficient reasons to respond with E *vis-à-vis* X.

Decisive Reasons: A token emotion E *vis-à-vis* X is fitting in virtue of the fact that there are decisive reasons to respond with E *vis-à-vis* X.

This way of analyzing what it is for an emotion to be fitting is vulnerable to the well-known *Wrong Kind of Reason-problem* (see Rabinowicz & Rønnow-Rasmussen 2004). The problem was originally cast as a problem for a particular kind of analysis of value – the *fitting attitude analysis of value*, or the *buck-passing account of value*. I shall say more about this analysis in a while, but for now we can note that the gist of it is that an object has value in virtue of the fact that there are reasons to have either a negative or a positive attitude towards the object. What the wrong kind of reason problem highlights is that there are occasions where there are reasons to adopt an attitude towards a particular object even though the object is not valuable.

To borrow an example from Rabinowicz and Rønnow-Rasmussen (2004), suppose that an evil demon threatens to kill you if you do not admire him. In such a case, you have a very good reason to admire the demon. But the fact that you have good reason to admire the demon is not a fact that makes the demon admirable. The fitting attitude analysis however predicts just that: an object (the demon) is valuable (admirable) in virtue of the fact that there are reasons to have a positive attitude (admiration) towards it. As Howard (2018) notes, an account of fittingness according to which it is analyzed in terms of reasons faces the very same problem.

This might lead us to think that the kinds of considerations that are relevant for establishing the fittingness relation must belong to a distinct normative category. After all, the fittingness relation is commonly construed as a *distinct normative relation* that is said to hold between an attitude and an object when the object *merits* the attitude.

Indeed, Selim Berker (2022) has recently argued that fittingness cannot be understood as belonging either in the realm of the deontic or in the realm of the evaluative. That is, we cannot analyze or explain the fitting in terms of the deontic or the evaluative and we must consequently think of *the fitting* as its own normative family. According to Berker, reasons generated by fit cannot be conceived in terms of either sufficient or decisive reasons. Nor can they be understood as establishing requiredness or permittedness. Rather, reasons of fit are *contributory* reasons or *pro tanto* reasons, which Berker maintains can be understood as "considerations that contribute towards the case in favour of" (Berker 2022, p.52) some relevant attitude. Importantly, Berker finds it unclear why "reasons so construed should be viewed as deontic in nature" (ibid, p.52).

3.3. Fittingness as Accurate Representation

Another way of understanding fittingness – which is the most prevailing in the literature on emotions – is to think of it as a matter of accurate representation. The general idea here is basically that many of our attitudes – e.g., beliefs, emotions, judgments, and so on – seem to *represent* their intentional objects as *being a certain way*, or as *having a certain property/feature*. For example, believing that X is a blue whale is – on this conception – to represent X as having the property of *being a blue whale*, or to represent it as being the case that *X is a blue whale*.

Consequently, the suggestion is that if an attitude type is a type of representational mental state and represents their intentional objects as being a certain way, then it seems to make good sense to say that instances of that attitude type are fitting insofar as their representations are accurate. With respect to beliefs then, we might say that if a token belief represents its object accurately, the belief is fitting – it does after all *fit* or *correspond* to the way the world actually is.

Similarly, it has been widely assumed that emotions represent their intentional objects as being a certain way – this idea is, as we have seen, embodied in *Robust*

Representationalism. In particular, many have thought that emotions represent their objects under an evaluative guise, or as instantiating evaluative properties. The idea that emotions are fitting in virtue of accurately representing their intentional objects may seem quite intuitive and appealing if one is attracted to the idea that emotions are representational mental states with a mind-to-world direction of fit. But even here, one has cause for pause.

From the premise that emotions have representational content, it does not necessarily follow that their fittingness is wholly a function of their representational content being accurate. To be sure, those who tend to think of the fittingness of emotions *as a matter of representational accuracy* may simply insist that that is just what fittingness amounts to. Such a response would be either stipulative or question-begging and unacceptable. If the claim is not a stipulative one, then proponents of the *fittingness-as-representational-accuracy* view need to substantiate their claim. One way in which this could perhaps be done is by means of showing that one of the primary functions of the emotions *is* to represent objects in world. If representing the world is indeed one of the primary functions of emotions, it could be argued that insofar as an instance of an emotion succeeds in performing that function it can be positively assessed along that dimension, i.e., we can say that it is fitting.

So, is representing the world (or objects in the world) one of the primary functions of emotions? I must admit that I do not have a conclusive answer to that question. Nevertheless, I think we should be cautious about endorsing the view that the fittingness of emotions is exclusively a function of their representational content being accurate (assuming, of course, that they have representational content to begin with). It seems dubious at best to carve out a partial aspect of such complex mental states as the emotions and claim that their fittingness is a function of this partial aspect alone. Put differently, I believe that it is a mistake to simply disentangle the constitutive features of the emotions and treat them as neatly separated components that can be evaluated according to standards that apply to those component parts. For example, suppose a token of an emotion type E is constituted by the more "basic" mental states F and G, where F is a mental state

with a representational content.[9] Even if there is a limited sense in which it can be said that the token emotion E is *not* 'off the mark' when the content of F is accurate (or corresponds accurately to the represented object), it seems unclear why that fact alone should determine the fittingness of E *as such*.

Accordingly, the accuracy of the putative representational content of emotions should – at best – only specify the necessary conditions for their fittingness and not the necessary *and* sufficient conditions. Rather, the necessary and sufficient conditions for the fittingness of emotions must be unearthed by examining the emotions *holistically,* as attitudes in their own right. This is certainly not an easy endeavor, and it may feel tempting to understand emotional fittingness on the model of other mental states – like that of belief – as nothing more than a function of representational correspondence. However, I believe that the guiding methodology for those concerned with the ethics of emotion should be one which is likely to yield a richer understanding of the role emotions play in our lives and what sort of work they are supposed to do. A step in that direction is to start recognizing emotions as attitudes in their own right – over and above their (alleged) constituent parts. By extension, I believe that we need to re-orient our theorizing and start thinking about fittingness as an assessment of emotions *as such*, rather than as merely an assessment having to do with one of their partial components.[10]

3.4. In Virtue of What are Emotions Fitting?

As we have seen, representationalists maintain that emotions are representational mental states. Some of these representationalists go beyond the ontological claim

[9] The mental state "G" in this case could be a purely conative or motivational mental state.

[10] These claims are congenial to the account of emotional fittingness favored by Hichem Naar (2021). Naar writes that "[t]he relation of fittingness between emotions and their objects is a normative relation over and above the relation of representation" (p. 14). I am in strong agreement with Naar that we should think of the fittingness of emotions as a normative matter which cannot simply be reduced to representational accuracy.

and espouse the normative claim that emotional fittingness is a matter representational accuracy. Given the common assumption among representationalists that emotions represent their objects as instantiating *evaluative properties* or *values*, the resultant view will yield the verdict that a token emotion is fitting in virtue of the fact that it accurately represents its object as instantiating a certain value. Such a view entails a commitment to the idea that values are the *fit-makers* for emotions – values are the properties that make emotions fitting. This does not sit well with the well-known *fitting attitude analysis of value*.

The fitting attitude analysis of value (henceforth the FA-analysis) is the view according to which the value of an object is explained in terms of it being such that an attitude of a certain kind is fitting with respect to it. This type of analysis – which is often said to have its origin in the works of Franz Brentano (1969/1889) and A.C. Ewing (1948) – utilizes the idea of the fittingness of attitudes (or reasons for attitudes) as a way of analyzing what value consists in.[11] A generic formulation of the FA-analysis of value states that:

> (**FA**) For something (object, situation, and so on) X to possess value Φ is for it to be *fitting* to have some particular attitude A towards X (Todd 2014, p.90, emphasis in original).

In other words, the FA-analysis offers an explanation of value according to which values are understood in terms of fitting responses. It is, for example, the fact that anger is fitting *vis-à-vis* X that makes it the case that X is offensive. Similarly, for X to be fearsome is for it to be fitting to fear X, for X to be admirable is for it to be fitting to admire X, and so on. Consequently, according to the FA-analysis, the fittingness of attitudes is understood as *explanatorily prior* to values. This fact is precisely the reason why *Robust Representationalism* is incompatible with the FA-analysis. Recall that according to *Robust Representationalism*, emotions are fitting

[11] For a nice historical exposé of the FA-analysis, see Rabinowicz & Rønnow-Rasmussen (2004).

in virtue of representing their intentional objects accurately. This hypothesis, in conjunction with the thought that emotions represent their objects as instantiating evaluative properties/values – which arguably most, if not all representationalists endorse – adds up to a view according to which it is the *value* or *the evaluative properties* of objects that *makes* emotions fitting (for a nice discussion see e.g., Naar, 2021). As we noted above, it holds that values are the fit-making facts, or the *fit-makers* of the emotions.

Robust Representationalism thus implies that emotions represent their objects as instantiating evaluative properties and that emotions are fitting just in case the representational content accurately corresponds to the features of the intentional object, i.e., just in case the represented object has the evaluative property in question to begin with. It is committed to the idea that it is the value of the intentional object which explains why emotions of different kinds are fitting with respect to those objects. According to the FA-analysis, the explanatory direction goes the other way. It is the fact that an emotion is fitting that explains the value of the object. The two views are irreconcilable since they treat the explanatory priority between values and fitting emotional attitudes as running in opposite directions. Therefore, we cannot endorse both *Robust Representationalism* and the FA-analysis. Which of the two views should we sacrifice?

My own sympathies lie with the FA-analysis. As several theorists have noted over the years, one of the important virtues of the FA-analysis has to do with it being meta-ethically neutral in the sense that it does not necessarily commit one to either a cognitivist or a non-cognitivist metaethical position. An additional reason in favor of adopting the FA-analysis of value has to do with its recourses for demystifying values.[12] Whether the FA-analysis can, at the end of the day, live up to its purported virtues is a question that will have to be treated elsewhere.

[12] For more on the theoretical merits of the FA-analysis, see e.g., Olson (2004), Rabinowicz & Rønnow- Rasmussen (2004), Rabinowicz (2013), Rønnow-Rasmussen (2021).

4. Emotions and Responsibility

Issues pertaining to the nature and the normativity of emotions are not only philosophically intriguing in themselves, but they are also directly relevant to other areas of philosophical inquiry. In particular, the subject matter of emotions has enjoyed much attention in the contemporary philosophical literature on *moral responsibility*.

On the face of it, this may seem a bit surprising. What does moral responsibility have to do with emotions? Is the issue of moral responsibility not an issue concerning free will and determinism? Before we answer these questions, it may be of service to briefly distinguish between the various ways in which the term *responsibility* might be used.

4.1. Different Senses of the Term 'Responsibility'

The notion of *responsibility* (and its cognates) can be used in a variety of different ways. Sometimes we use the term to indicate that we occupy a certain social role that comes with certain duties and obligations. My role as a father entails that I have certain duties and obligations to my daughter – I am responsible for her. My role as a teacher at the university entails that I have certain obligations *vis-à-vis* my department and my students – it is, for example, my responsibility to prepare lectures, grade papers, and so on.

There is also a purely *causal* sense of the term *responsible* which might be used in order to convey that something was the cause of something else. For example, we might say that "The recent heat waves are responsible for the nation-wide

drought", or that "A new type of virus seems to be responsible for the sudden increase in hospitalizations". Human agents are of course also responsible, in this purely causal sense of the term, for all kinds of things. I may for instance be causally responsible for the fact that there is milk in the fridge.

Then there is the notion of *moral responsibility*. When we say that someone is morally responsible for some action or omission, we do not merely mean that they are causally responsible for that action or omission. To attribute moral responsibility to an agent A for some action φ is to do something else than to merely point out that A was the cause of whatever consequence that was brought about by the act of φ-ing. It is to say that, on the basis of having performed the action of φ-ing, A is, in some sense, liable to a certain kind of moral assessment, or a candidate for a certain kind of treatment. This raises the question what the conditions are under which an agent can be said to be liable for such an assessment or treatment. What are the conditions that must hold in order for an agent be morally responsible?

4.2. Strawson and the Reactive Attitudes

A widespread idea among philosophers has been that in order for an agent to be morally responsible for her actions, the agent needs to possess a certain kind of freedom of the will. Exactly what the relevant kind of freedom of the will is supposed to consist in is a contested issue, but it has often been thought that it must consist in the ability to do otherwise.

So construed, it has been argued by many that we cannot be morally responsible if the requisite free will is not compatible with the truth of determinism – the thesis that "given the past and the laws of nature, the future is determined in every detail" (van Inwagen 2017, p.82). If the future is determined, then so are our actions, and that means that we cannot act otherwise than how we actually happen to act. Consequently, these theorists have argued that the truth of determinism would render our responsibility practices – our practices of holding each other

responsible – ultimately unjustified. Others deny such a conclusion, arguing that the relevant kind of freedom is compatible with determinism.

The former party is standardly referred to as *incompatibilists*, and the latter party as *compatibilists*. Incompatibilists are typically divided into two camps: *hard determinists* and *libertarians.* Hard determinists claim that:

(i) The kind of free will required for moral responsibility is incompatible with the thesis of determinism
(ii) The thesis of determinism is true
(iii) Therefore, we lack the free will required for moral responsibility

Libertarians agree with hard determinists with respect to the first claim – the claim that free will is incompatible with determinism – but maintain that we do have the kind of free will required for moral responsibility since the world is in fact indeterministic.[13]

With his landmark essay "Freedom and Resentment" (1962), Peter Frederick Strawson invigorated the conventional debate over moral responsibility by drawing attention to a central commonplace of our human lives, namely:

> the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about these attitudes and intentions (1962, p.3).

In other words, while the traditional debate over moral responsibility was concerned with abstract questions regarding the freedom of the will and determinism, Strawson urged us to direct our attention to how we actually engage

[13] For more on the various positions in the debate see e.g., McKenna & Pereboom (2016), and Talbert (2016).

with one another in ordinary interpersonal relationships and appreciate the extent of our natural human commitment to participate in such relationships (ibid., p.7).

Once our attention is fixed on what is actually going on within our social practices and interpersonal transactions with one another, we can come to see and appreciate the fact that we – within the framework of such relationships – typically demand or expect "some degree of goodwill or regard on the part of those who stand in these relationships to us" (ibid., p.3-4). The critical observation that Strawson then goes on to make is that the relevant kind of demand or expectation for goodwill and regard is reflected in or expressed through our *reactive attitudes*. Strawson writes:

> The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings towards ourselves; or at least on the expectation of, and demand for, an absence of the manifestation of active ill will or indifferent disregard (ibid., p.8).

Roughly then, Strawson's central point is that by attending to our actual interpersonal relationships and our practices of engaging with each other, we can see that to be involved in these practices *is* to be prone to the reactive attitudes (1962, p.7). The reactive attitudes express, or reflect, our interpersonal demands and expectations, and these are part and parcel of our interpersonal transactions. And this is arguably all we need to attend to in order to answer questions about responsibility. As Strawson puts it:

> Only by attending to this range of attitudes can we recover from the facts as we know them a sense of what we mean, i.e. of all we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice (ibid., p.13).

In the wake of Strawson's groundbreaking essay, many philosophers have come to take seriously the idea that a proper understanding of our responsibility practices and their rationale must take into account the reactive attitudes.

Strawson's own list of the reactive attitudes, which comprises attitudes like gratitude, resentment, indignation, love, and hurt feelings, are clearly paradigm examples of emotions[14], and it is thus no big surprise that there has evolved a tradition of responsibility theorists who place the emotions at center stage in their theorizing (e.g., Graham 2014; Menges 2017; Rosen 2015; Shoemaker 2015; Tognazzini 2013; Wallace 1994).

4.3. Alethic Views of Moral Responsibility

Building on the Strawsonian paradigm, several of the leading figures in the responsibility debate now espouse the idea that we can analyze what it is to *be* morally responsible in terms of the reactive emotions (see e.g., Wallace 1994; McKenna 2012; Shoemaker 2015, 2017). By examining the emotions through which we engage with one another and pinpointing the conditions under which these emotions are appropriate, we may unearth the conditions for being morally responsible. Consequently, more than a few theorists have come to embrace some version of the following biconditional:

> **Moral Responsibility (MR):** An agent A is morally responsible for φ if and only if it is appropriate to have – depending on the nature of φ – either a negative or positive reactive emotion *vis-à-vis* A on account of φ. (Pál-Wallin 2025, p.2).

A crucial question with respect to the above biconditional is of course how we are supposed to understand the notion of *appropriate*. As we saw in chapter 3, there

[14] Although it may be inaccurate to classify love or hurt feelings as distinct *emotion types*, it is undoubtedly the case that these phenomena are emotionally toned. Additionally, there may be *bona fide* reactive attitudes which are not emotions. *Trust* might be a case in point. Still, it is quite clear that a significant set of the reactive attitudes which Strawson considered are genuine types of emotion.

are several ways in which an emotion can be deemed appropriate, and so it is a vital task to figure out what the relevant sense of appropriate might be.

One alternative is to say, as R. Jay Wallace argues in *Responsibility and the Moral Sentiments* (1994), that the relevant sense of appropriate is to be understood in terms of *moral fairness*. On this sort of picture, for A to be morally responsible for φ would be a matter of it being *morally fair* to have a reactive emotion *vis-à-vis* A on account A's φ-ing. (Of course, this raises the question of what it is, or under what conditions it is, morally fair to have a reactive emotion towards another agent).

Another alternative that has recently been proposed by some theorists in the literature (e.g., Graham 2014; Rosen 2015; Strabbing 2019) is that the relevant sense is supposed to be cashed out in terms of *representational accuracy***.** Here, the idea that emotions are representational mental states and that their fittingness is a matter of accurate representation resurfaces. These co-called *Alethic Views of Moral Responsibility* have been subject to a devastating critique by yours truly in a recent paper (Pál-Wallin 2025). The paper is included in this thesis and follows directly after this chapter and a short summary of all the papers. I will therefore not spend too much time discussing these views here. I do, however, think that it might be a good idea to present to the reader the core arguments that undergird the Alethic Views in a more formalized manner than I do in the actual paper.

To that end, we should first note – as I do in the paper – that we can distinguish between a *strong version* and a *modest version* of the Alethic View – which I refer to as *Strong Alethicism* and *Modest Alethicism* respectively. For ease of exposition, I present the underlying argument of each view by showing how they would account for what it is to be blameworthy.

Strong Alethicism:

P1: X is blameworthy *in virtue of* the fact that it is fitting to have a blaming emotion directed towards X.

P2: All episodic instances of an emotion type have a representational content with a mind-to-world direction of fit.

P3: An instance of an emotion type E towards a particular object O is fitting *in virtue of* the fact that E's representation of O is accurate.

C: X is blameworthy *in virtue of* the fact that the representational content of a blaming emotion E directed towards X is accurate.

Modest Alethicism:

P1: X is blameworthy *if and only if* it is fitting to have a blaming emotion directed towards X.

P2: All episodic instances of an emotion type have a representational content with a mind-to-world direction of fit.

P3: An instance of an emotion type E towards a particular object O is fitting *if and only if* E's representation of O is accurate.

C: X is blameworthy *if and only if* the representational content of a blaming emotion E directed towards X is accurate.

The essential difference between *Strong Alethicism* and *Modest Alethicism* is that *Strong Alethicism* is "committed to a grounding relation between X's blameworthiness and the representational accuracy of a blaming emotion E such that it is the representational accuracy of the blaming emotion that *explains* (or *grounds*) X's blameworthiness" (Pál-Wallin 2025, p.7). *Strong Alethicism* thus offers a genuine analysis of *what it is to be* blameworthy (and praiseworthy).

Contrastingly, *Modest Alethicism* remains neutral with respect to this explanatory priority between X's blameworthiness and the representational accuracy of a blaming emotion E. Hence, *Modest Alethicism* offers only a biconditional.

Both of these versions of the Alethic View are targets of my critique in my paper "In Opposition to Alethic Views of Moral Responsibility" (2025), which we shall turn to right after a short summary of all the papers in this thesis. Enjoy.

Summary of Papers

Paper I: *In Opposition to Alethic Views of Moral Responsibility*

A standard analysis of moral responsibility states that an agent A is morally responsible for φ-ing if and only if it is fitting to have – depending on the nature of φ – a negative or positive reactive emotion *vis-à-vis* A on account of A's φ-ing. Proponents of *Alethic views of moral responsibility* argue that the relevant notion of fittingness in the analysis should be understood in terms of accurate representation. The allure of understanding emotional fittingness as representational accuracy arguably stems from the widespread idea that emotions are representational mental states with a mind-to-world direction of fit. Consequently, proponents of Alethic views argue that the fittingness of emotions should be cashed out in terms of whether the representational content of reactive emotions accurately matches the targeted agent. The aim of this paper is to argue against Alethic views of moral responsibility by means of exposing various problems that these accounts face in virtue of their inherent commitment to understand emotional fittingness in terms of representational accuracy.

Paper II: *Emotional Fittingness, Individual Nature, and Relational Values*

It is often maintained that emotions represent their intentional objects as instantiating evaluative properties, and that emotional fittingness is a matter of accurate representation. An instance of admiration is – on this view – fitting in virtue of the fact that it accurately represents the admired object as instantiating the property of being admirable. In this paper, I first outline reasons to resist this kind of picture of the emotions and their fittingness. Drawing on recent attitudinal accounts of the emotions, I sketch my own view of the emotions

according to which they are *sui generis* affective modes of engagement through which we express our cares, concerns, and various commitments – i.e., elements that constitute our evaluative orientations and practical identities as agents. In that sense, emotions inherently display a kind of *import-for-the-subject* structure which an account of their fittingness must accommodate. The resultant view construes fittingness as a relation of normative support relativized to an agent's *individual nature,* according to which an instance of an emotion E is fitting with respect to its intentional object *partly* in virtue of non-normative facts about the emoting agent. In light of the account of emotional fittingness defended here, a suggestion is made to understand sentimental values as essentially relational values, which supervene on the subject-object relation.

Paper III: *When Your Reasons Become My Reasons: On Internalism and Empathic Resonance*

Bernard Williams (1981) famously defended an internalist account of practical reasons according to which statements of the form 'A has a reason to ϕ' are true only if there is a *sound deliberative route* from A's existing motivations by which A could become motivated to ϕ. In other words, if A could not become motivated to ϕ via a process of sound deliberation that begins from her existing motivations, there is no reason for A to ϕ. This renders the internalist thesis unnecessarily and mistakenly restrictive: It neglects the important ways in which a reason to ϕ may apply to an agent precisely in virtue of there being *non-deliberative* routes from her existing motivations by which she could become motivated to ϕ. The aim of this article is to explore a potential way to expand the internalist picture by examining one such non-deliberative route – a process I refer to as *empathic resonance*. Empathic resonance designates a particular kind of relational entanglement and affective-motivational coupling between empathizer and empathizee that occurs in direct face-to-face encounters. The characteristic feature of this entanglement is that it exerts a gripping force on the empathizer that manifests itself as an experience of an *overwhelming directive to act on behalf of the*

empathizee. Consequently, empathic resonance constitutes a distinct kind of non-deliberative process of interpersonal attunement in virtue of which an internal reason for the empathizee to ϕ can be said to apply to the empathizer vicariously.

Paper IV: Fear as a Reactive Attitude

In the wake of Peter Frederick Strawson's landmark essay "Freedom and Resentment" (1962), much of the theorizing about moral responsibility has centered around the reactive attitudes – with a particular emphasis on guilt, resentment, and indignation. Although philosophical interest in previously unexamined reactive attitudes has grown rapidly in recent years, remarkably little has hitherto been said about fear as a candidate reactive attitude. The aim of this chapter is to explore the phenomenon of fearing other human agents *qua* agents. Drawing on P.F. Strawson's seminal work on the reactive attitudes, I aim to develop an account of agent-directed fear as a distinct reactive attitude (which I refer to as *reactive fear*). In order to pave the way for such an account, I shall in the first part of this chapter consider a potential objection to the central claim that people can fear other agents. The objection has its roots in a propositionalist view of mental states according to which fear is a propositional attitude and thus essentially about potential events, or state of affairs (something that is, was, or will be the case). I will provide reasons for why we should reject such a restrictive view of fear, and more specifically, why we need to embrace the idea that fear can be directed towards other human agents. In the second part of the chapter, I contrast my proposed account of reactive fear to some existing views of fear, according to which fear is seen as reactive in a very general sense, as a reaction to any number of objects. On the account I propose, reactive fear is a distinct fearing attitude *vis-à-vis* human agents *qua* agents. As I aim to show, this kind of attitude is a particular mode of engagement which reflects an underlying concern of the fearing agent not to have her practical identity harmed by the feared agent. I will conclude by discussing the upshot and importance of recognizing fear as a *bona fide* reactive attitude.

References

Alvarez, M. (2017). Reasons for action: justification, motivation, explanation. In Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition).

Anscombe, G. E. M. (1957). *Intention*. Oxford: Basil Blackwell.

Berker, S. (2022). The deontic, the evaluative, and the fitting. In Chris Howard, and R. A. Rowland (eds), *Fittingness: Essays in the Philosophy of Normativity*. Oxford University Press.

Brentano, F. (1969) [1889]. *The Origin of our Knowledge of Right and Wrong*, ed. Oskar Kraus and Roderick Chisholm, trans. Roderick Chisholm and Elizabeth Schneewind. London: Routledge & Kegan Paul.

Colombetti, G. (2010). Enaction, sense-making, and emotion. In Stewart, J., Gapenne, O. & Di Paolo, E. (2010)(eds). *Enaction: Towards a New Paradigm for Cognitive Science* (pp.145-164) Cambridge MA: MIT Press.

D'Arms, J. (2024). What makes an emotion appropriate or inappropriate to its object? In *Emotion Theory: The Routledge Comprehensive Guide* (pp.605-627). Routledge.

D'Arms, J. & Jacobson, D. (2000a). Sentiment and value. *Ethics*, Vol. 110 (July 2000), pp. 722-748.

D'Arms, J.& Jacobson, D. (2000b). The moralistic fallacy: On the 'appropriateness' of emotions. *Philosophy and Phenomenological Research, 61*(1), 65–90.

De Sousa, R. (2007), Truth, authenticity, and rationality. *Dialectica*, 61: 323-345. https://doi.org/10.1111/j.1746-8361.2007.01104.x

Deonna, J. & Teroni, F. (2012). *The Emotions: A Philosophical Introduction*. London: Routledge

Deonna J. & Teroni F. (2014). In what sense are emotions evaluations? In Cain Todd & Sabine Roeser (eds.), *Emotion and Value.* Oxford University Press, pp. 15-31

Deonna, J. & Teroni, F. (2015). Emotions as attitudes. *Dialectica*; Vol. 69, N° 3 (2015), pp. 293–311

Deonna, J. & Teroni, F. (2022). Emotions and their correctness conditions: A defense of attitudinalism. *Erkenntnis*. https://doi.org/10.1007/s10670-022-00522-0

Döring, S. A. (2003). Explaining action by emotion. *The Philosophical Quarterly (1950-)*, *53*(211), 214–230.

Döring, S. A. (2007). Seeing what to do: affective perception and rational motivation. *Dialectica*, *61*(3), 363–394.

Echeverri, S. (2019). Emotional justification. *Philosophy and Phenomenological Research*, *98*(3), 541–566.

Ewing, A. C. (1948). *The Definition of Good*. London: Routledge & Kegan Paul.

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

Goldstein, I. (2002). Are emotions feelings? A further look at hedonic theories of emotions. *Consciousness & Emotion* 3:1 (2002), 21–33.

Graham, P. A. (2014). A sketch of a theory of moral blameworthiness. *Philosophy and Phenomenological Research, 88*(2), 388–409.

Greenspan, P.S. (1988). *Emotions and Reasons: An Inquiry into Emotional Justification* (1st ed.). Routledge.

Helm, B. W. (2001). *Emotional reason: Deliberation, Motivation, and the Nature of value*. Cambridge: Cambridge University Press.

Howard, C. (2018). Fittingness. *Philosophy Compass, 13*, 11.

Hutto, D. D. (2012). Truly enactive emotion. *Emotion Review*, *4*(2), 176-181.

James, W. (1884). What is an emotion? *Mind*, 9: 188–205

Kenny, A.J.P. (1963). *Action, Emotion and Will* (London: Routledge and K. Paul).

Lange, C. G. (1885). *Om Sindsbevægelser*. Kjøbenhaven: Jacob Lunds Forlag

Massin, O. (2021). The reactive theory of emotions. *European Journal of Philosophy*, 31(3), 785–802. https://doi.org/10.1111/ejop.12736

McGrath, M. & Frank, D. (2024). Propositions. *The Stanford Encyclopedia of Philosophy* (Fall 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = https://plato.stanford.edu/archives/fall2024/entries/propositions/

McKenna, M. (2012). *Conversation and Responsibility*. New York: Oxford University Press.

McKenna, M., & Pereboom, D. (2016). *Free Will: A Contemporary Introduction*. Routledge, Taylor & Francis Group.

Menges, L. (2017). The emotion account of blame. *Philos Stud*, 174:257-273.

Merricks, T. (2009). Propositional Attitudes? *Proceedings of the Aristotelian Society* 109: 207–232.

Milona, M. & Naar, H. (2020). Sentimental perceptualism and the challenge from cognitive bases. *Philos Stud*, 177:3071–3096. https://doi.org/10.1007/s11098-019-01360-7

Milona M. (2023). What perceptualists can say about reasons for emotion. *Canadian Journal of Philosophy*. 53(6): 502-518. doi:10.1017/can.2024.43

Mulligan, K. (1998). From appropriate emotions to values. *The Monist*, Vol. 81, No. 1, pp. 161-188

Müller, J. M. (2017). How (not) to think of emotions as evaluative attitudes. *Dialectica*, 71:281– 308. doi: 10.1111/1746-8361.12192.

Müller, J. M. (2018). Emotion as position-taking. *Philosophia, 46*(3), 525–540.

Müller, J. M. (2021). The spontaneity of emotion. *European Journal of Philosophy*. 1–19. https://doi.org/10. 1111/ejop.12625

Müller, J.M. (2025). Motivating reasons: The state of the question. *The Southern Journal of Philosophy*. https://doi.org/10.1111/sjp.12623

Naar, H. (2020). Emotion: more like action than perception. *Erkenntnis,* (2020) https://doi.org/10.1007/s10670-020-00324-2

Naar, H. (2021). The fittingness of emotions. *Synthese* **199**, 13601–13619. https://doi.org/10.1007/s11229-021-03391-2

Naar, H. (2022). Do emotions represent? In J. Deonna, C. Tappolet and F.Teroni (eds.), *A Tribute to Ronald de Sousa*. URL https://sites.cisa-unige.ch/ronald-de-sousa/

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

Nussbaum, M. C. (2016). *Anger and Forgiveness: Resentment, Generosity, Justice.* New York: Oxford University Press

Olson, J. (2004). Buck-passing and the wrong kind of reasons. *The Philosophical Quarterly*, Vol. 54, No. 215, pp. 295-300.

Pál-Wallin, R. (2025). In opposition to alethic views of moral responsibility. *The Southern Journal of Philosophy*. https://doi.org/10.1111/sjp.12627

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

Rabinowicz, W. (2013). Value: fitting-attitude account of. In H. LaFolette (Ed.), *International Encyclopedia of Ethics* (pp. 1-12). Wiley-Blackwell. https://portal.research.lu.se/en/publications/value-fitting-attitude-account-of

Rabinowicz, W. & Rønnow-Rasmussen, T. (2004). The strike of the demon: On fitting pro-attitudes and value. *Ethics,* 114(3), 391-423.

Rosen, G. (2015). The alethic conception of moral responsibility. In *The Nature of Moral Responsibility*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith. New York: Oxford University Press, 65–87.

Rønnow-Rasmussen, T. (2021). *The Value Gap*. Oxford University Press.

Scanlon, T.M. (1998). *What we owe to each other*. Cambridge, Mass.: Belknap

Scarantino, A. (2014). The motivational theory of emotions. In Justin D'Arms, and Daniel Jacobson (eds), *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics*. New York, NY: Oxford University Press

Scarantino, A. & De Sousa, R. (2021). Emotion. *The Stanford Encyclopedia of Philosophy.* Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2021/entries/emotion/

Schroeter, L., Schroeter, F. & Jones, K. (2015). Do emotions represent values? *dialectica* Vol. 69, N° 3 (2015), pp. 357–380

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

Sellars, W. (1962). Philosophy and the scientific image of man. Published in *Frontiers of Science and Philosophy*, edited by Robert Colodny (Pittsburgh: University of Pittsburgh Press, 1962): 35-78. Reprinted in *Science, Perception and Reality* (1963).

Shargel, D. (2015). Emotions without objects. *Biol Philos* **30**, 831–844. https://doi.org/10.1007/s10539-014-9473-8

Shargel, D., & Prinz, J. (2017). An enactivist theory of emotional content. In H. Naar & F. Teroni (Eds.), *The Ontology of Emotions* (pp. 110-129). Cambridge: Cambridge University Press. doi:10.1017/9781316275221.007

Shoemaker, D. (2015). *Responsibility From the Margins.* Oxford: Oxford University Press.

Shoemaker, D. (2017). Response-dependent responsibility; or, a funny thing happened on the way to blame. *Philosophical Review*, 126(4), pp. 481–527.

Smith, A. M. (2005). Responsibility for attitudes: activity and passivity in mental life. *Ethics*, *115*(2), 236–271.

Solomon, R. C. (1973). Emotions and choice. In *Not Passion's Slave: Emotions and Choice* (New York, 2003)

Srinivasan, A. (2018). The aptness of anger. *The Journal of Political Philosophy*: Volume 26, Number 2, 2018, pp. 123–144

Strabbing, J. T. (2019). Accountability and the thoughts in reactive attitudes. *Philosophical Studies* 176(12): 3121– 40.

Strawson. P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48: 1-25

Talbert, M. (2016). *Moral Responsibility: An Introduction*. (Key concepts in philosophy). Polity Press.

Tappolet, C. (2016). *Emotions, Value, and Agency*. Oxford University Press.

Teroni, F. (2007). Emotions and formal objects. *dialectica*, 61, 3, pp. 395–415.

Todd, C. (2014). Relatively fitting emotions and apparently objective values. In Cain Todd & Sabine Roeser (eds.), *Emotion and Value.* Oxford University Press

Tognazzini, N. (2013). Blameworthiness and the affective account of blame. *Philosophia*, 41:1299-1312.

van Inwagen, P. (2017). Moral responsibility, determinism, and the ability to do otherwise. In: *Thinking about Free Will*. Cambridge University Press; 2017:81-89.

Wallace, J. R. (1994). *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press.

Williams, B. (1981). Internal and external reasons. In *Moral Luck*, 101–113. Cambridge: Cambridge University Press.

Yang, S. (2016). Do emotions have directions of fit? *Organon F,* 23 (1) 2016: 32-49

Zamuner, E. (2015). Emotions as psychological reactions. *Mind & Language, 30*(1), 22–43.