DOI: 10.1111/sjp.12627

ORIGINAL ARTICLE

The Southern Journal of Philosophy

In opposition to alethic views of moral responsibility

Robert Pál-Wallin

Department of Philosophy, Lund University, Lund, Sweden

Correspondence
Robert Pál-Wallin, Department of Philosophy, Lund University, Helgonavägen 3, Lund 221 00, Sweden.
Email: robert.pal-wallin@fil.lu.se

Abstract

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* maintain 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. Accordingly, defenders of Alethic views argue that the fittingness of a potential reactive emotion is a matter of whether the representational content of that emotion accurately matches the targeted agent. The aim of this article 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.

1 | INTRODUCTION

What does it mean to be *morally responsible* for doing—or failing to do—something? According to one very influential view on moral responsibility, being responsible is just a matter of being an appropriate target of some negative or positive reactive emotion. To illustrate: Imagine driving to work on a Monday morning. You stop at an intersection and wait patiently for the traffic light to turn green. Unexpectedly, a car crashes into you from behind. Apart from a minor head injury from hitting the steering wheel, you are physically safe and sound. Confused and shook up, you step out of your car to assess the situation.

Now, suppose you discover that the car crashed into you because the driver was heavily intoxicated. In that case, it would be quite natural to become angry with the driver. But, more

importantly, directing certain negative emotional reactions toward the driver—emotions such as anger, resentment, or indignation—will, in light of the driver's intoxication, seem perfectly justified. After all, driving intoxicated is (for most of us) something objectionable, something that we can legitimately demand or expect people to refrain from. When people fail to live up to that demand, certain negative emotions are licensed, and it may even be appropriate to express them by overtly blaming the driver. On the other hand, suppose instead that you discover that the car crashed into you because the driver suffered a sudden heart attack. In that case, anger would not only be an unusual response, but it would also strike us as an inappropriate response. We cannot legitimately demand that people do not suffer sudden heart attacks while driving, and any negative emotional reactions directed toward the unfortunate driver will seem misplaced and unwarranted.

The key difference between the first version of the case and the second is precisely that whereas the first driver is an appropriate target of various negative emotional reactions, the second driver is not, and this is meant to explain why the former is morally responsible for crashing into you, whereas the latter is not.[1] The guiding thought here is that we must understand what it is to be morally responsible by attending to what it is that human beings normally do in interpersonal transactions with one another. And what we do is precisely illustrated by the case above: we engage emotionally with one another in ways deemed appropriate by the circumstances.

This central aspect of our ordinary human life has led to a widespread recognition among responsibility theorists that we can understand what it is to *be* morally responsible for some action or omission by examining the reactive emotions and the conditions under which they are appropriate (see e.g., McKenna, 2012; Shoemaker, 2015, 2017; Strawson, 1962; Wallace, 1994). 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.[2]

One issue that has attracted a lot of attention in the current responsibility debate is the question of how we should understand the notion of *appropriate* appealed to in a biconditional like ***MR***. Recently, defenders of what we might call *Alethic views of moral responsibility* (e.g., Graham, 2014; Rosen, 2015; Strabbing, 2019) have argued that the relevant sense of appropriateness is to be understood in terms of *representational accuracy*.[3] The appeal to representational accuracy is motivated by the widespread idea that emotions are at least partly constituted

[1]We might perhaps still say that both drivers are responsible, *in a causal sense*, for crashing into you and damaging your car. But attributing responsibility in this causal sense would just be a way of providing a causal explanation for why the event (i.e., the car crash) occurred, just as we do when we say things like: "The blizzard is responsible for widespread train delays" or "A virus is responsible for his respiratory infection." Attributing *moral responsibility* involves something else than merely establishing a causal connection.

[2]I wish to stay inclusive with respect to the kinds of things agents can be responsible for (or on account of). To that effect, "φ" in the biconditional should be read as a placeholder for not only different types of actions, but also for omissions, judgments, beliefs, attitudes, character traits, etc. None of the arguments presented in this article will hinge on whether or not one accepts this inclusive view.

[3]Rosen does not actually use the phrase "representational accuracy", but he argues that an emotion is appropriate (in the intended sense) with respect to its intentional object when the emotion's "ingredient thoughts are *true*" (2015, p.72, italics added). And Rosen takes these thoughts to be *representational* states. By contrast, Strabbing (2019) frames the discussion in terms of *fittingness*, which she explicitly identifies with representational accuracy. Strabbing's use of fittingness is slightly misleading since it seems to suggest that fittingness *just is* a matter of representational accuracy and nothing else—a tendency that has become something of an unfortunate orthodoxy in the philosophy of emotion. *Pace* Strabbing, the notion of fittingness is open to a variety of different interpretations (see, e.g., Howard, 2018; Howard & Rowland, 2022). Notwithstanding terminological differences, it is apparent that both Rosen and Strabbing take the relevant sense of appropriateness to be a matter of accurate correspondence between the representational content of the emotion and its intentional object. I will continue to use the term "representational accuracy" when referring to the intended Alethic understanding of appropriateness.

by representational mental states (e.g., thoughts, judgments, perceptions etc.) with a mind-to-world direction of fit. In light of this assumption, proponents of Alethic views contend that the appropriateness of a candidate reactive emotion is solely a function of whether its representational content matches the represented agent.

For example, suppose that resentment involves the three distinct thoughts *x*, *y*, and *z*.[4] The Alethic view of moral responsibility then yields the verdict that insofar as *x*, *y*, and *z* are all true of an agent A and A's φ-ing, A is morally responsible. Although I agree that a philosophical examination of the emotions and their fittingness can aid us in understanding what it is to be morally responsible, I will argue in this article that the Alethic way of doing it—by identifying the fittingness of the reactive emotions with representational accuracy—is wrongheaded.[5]

The article is structured as follows: In Section 2, I will present a few clarificatory remarks concerning the nature of emotions. I will then outline a sketch of the Alethic view of moral responsibility as presented by Gideon Rosen (2015) and Jada Twedt Strabbing (2019) and shed light on its inherent commitment to a particular view of the emotions, namely *representationalism* (Section 3). In Section 4, I briefly outline two versions of the Alethic view—which I shall call *Strong Alethicism* and *Modest Alethicism*—and underscore the main difference between them. I offer arguments against *Strong Alethicism* in Section 5 and against *Modest Alethicism* in Section 6. I conclude in Section 7.

2 | CLARIFICATORY REMARKS

To avoid confusion, we should start by highlighting two distinctions that are often made in the philosophical literature on emotions. The first is the distinction between being affected by an emotion in the *episodic* sense and being affected by an emotion in the *dispositional* sense (see, e.g., Deonna & Teroni, 2012). For example, to grieve the passing of a loved one does not necessarily imply that one is caught in a never-ending state of occurrent sorrow. At times, grief rests dormant and makes room for other emotions to take center stage. Nevertheless, in its latent slumber, grief may be lurking in the background and *disposing* the grief-stricken to experience bouts of intense sadness when—under certain circumstances—the beloved springs to mind. This is the sense in which one can be affected by an emotion in *the dispositional sense*. By contrast, when we are struck by an emotion in *the episodic sense*, the emotion in question takes center stage and manifests itself temporarily, as an episodic affective experience. The ensuing discussion will focus on emotions in the *episodic* sense.

Another common distinction in the literature is the one that theorists tend to draw between *emotions* and a closely related class of affective states, namely *moods*. While there might be disagreement over the precise nature of these mental phenomena, it is generally maintained

[4]Two quick remarks: (1) The appeal to *resentment* in this case is motivated by the pervasive idea that it is the fittingness of *guilt*, *resentment*, and *indignation* that fixes the conditions for moral responsibility. Although I do not personally believe that an inquiry into the nature of responsibility should be restricted to these three emotion types, I choose to follow Rosen (2015) in focusing on resentment for the purpose of the argument. (2) The fact that the reader is urged to suppose that resentment involves *thoughts* is not essential to the example. The point of the example does not hinge on whether we imagine resentment to involve *a thought*, *a judgment*, *a perception*, or *a belief* as long as we hold fixed that these mental states have representational content.

[5]This seems to be a fitting place to mention that my overall aim in this article bears close resemblance to what Justin D'Arms has argued for in his recent article titled "Fitting Emotions" (2022). Let me therefore point out a significant dissimilarity between D'Arms's article and the present one as a way of motivating why my aims here are worth pursuing. In essence, while both D'Arms and I are in the business of arguing against a representational understanding of the fittingness of emotions, D'Arms's project is more general in scope and does not address the problematic implications of adopting such a view in an account of moral responsibility in particular. The arguments in this article are tailored precisely to that end. I thus believe that my article—and the arguments presented here—complements D'Arms's article and strengthens the case against any philosophical account that construes the fittingness of emotions in terms of representational accuracy.

that emotions have a certain "directedness" or "aboutness" that moods seem to lack. That is, emotions are seen as essentially *directed toward* or *about* various objects, events, or state of affairs in the world. Philosophers refer to this kind of directedness of mental states as *intentionality* (see, e.g., Crane, 1998; Searle, 1983). For example, one can be angry *with* one's partner, one can be afraid *of* the spider in the kitchen, or one can be amused *by* a joke etc. By contrast, being in a particular mood (anxious, melancholic, joyful, etc.) does not necessarily involve a directedness toward anything at all. In the rest of this article, I will assume that emotions do have this kind of intentionality, which makes them distinguishable from nonintentional mental states, and I will refer to the objects toward which emotions are directed as their *intentional objects*. In sum, the subsequent discussion will focus on emotions understood as *episodic* and *intentional* affective phenomena.

3 | ALETHIC VIEWS OF MORAL RESPONSIBILITY

In his article *The Alethic Conception of Moral Responsibility* (2015), Gideon Rosen sets out to present and develop a framework for thinking about moral responsibility in terms of appropriate emotional responses.[6] In particular, he aims to articulate a view of (i) what it is for someone to be morally responsible for some X, (ii) what conditions need to be satisfied in order for someone to be morally responsible for some X, and (iii) why those conditions are as they are (2015, p. 65). By way of stipulation, Rosen claims that moral responsibility is a matter of being either morally praiseworthy or morally blameworthy for some action or omission,[7] and that to be praiseworthy and blameworthy is a matter of being an appropriate target of praise and blame.[8] Thus, whether an agent is blameworthy for some X (e.g., action, omission, etc.) is simply a matter of it being appropriate to blame that agent on account of X.

What is it then, *to blame* someone? Although Rosen is open to the idea that there can be mild forms of blaming responses, he is primarily interested in blame that is harsh in nature. To that effect, Rosen states that he is following Strawson (1962) and Wallace (1994)—among others—in holding that "to blame X for A is to *resent* X for A, or to be *indignant* toward X for A, or to feel *guilty* about having done A (if you are X)" (2015, p. 67).[9] In sum, Rosen presents a view or moral responsibility according to which an agent A is blameworthy just in case resentment, indignation, guilt, or some other negative reactive emotion toward A is appropriate. What then, is it for a reactive emotion to be *appropriate*?

It is by now commonly accepted among philosophers that there are several 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, a reactive emotion 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, reactive emotions can also be deemed appropriate in the sense of being

[6]Similar ideas are sketched by Peter Graham in his "A Sketch of a Theory of Moral Blameworthiness" (2014). I shall be focusing on Rosen's developments of these ideas.

[7]This understanding of moral responsibility—as a matter of being *either* morally praiseworthy or morally blameworthy—excludes the possibility of being morally responsible for morally neutral acts. One might think that an account of moral responsibility must make room for such a possibility. This might very well be the case, but it should not distract us from our aims here. For the purposes of this article, we can simply follow Rosen and treat his conception of moral responsibility as a stipulative one (2015, p. 66).

[8]In the interest of space, I will mostly focus on blame and blameworthiness.

[9]Affective accounts of blame are rather common in the contemporary literature. See for example, Menges (2017), Tognazzini (2013) and Wolf (2011). Marta Johansson Werkmäster (2022) has argued that we should conceive of blame as a sentiment rather than as an affective episode. Still, on Johansson Werkmäster's account, blame will often manifest itself through token blaming emotions.

supported by *moral* considerations, for example, if having the emotion would lead to a morally good/desirable outcome.[10]

However, as various theorists have pointed out, any analysis that aims to analyze the evaluative or normative status of some object X in terms of appropriate attitudes is vulnerable to the familiar *conflation problem* (or, *the wrong kind of reason problem*).[11] In essence, the problem is that there seem to be occasions where adopting an attitude E is appropriate (or supported by reasons) vis-à-vis an object X despite the fact that X is not E-able (and vice versa). Hence, any analysis of the kind that appeals to the appropriateness of attitudes in explaining what value consists in must be able to differentiate between kinds of appropriateness (or between kinds of reasons) that do and do not bear on the value of the intentional object. So, while considerations of prudence and moral goodness can bear on the question whether or not an agent should—all things considered—have the emotion, philosophers tend to regard these types of reasons to be of the wrong kind for settling the conditions under which an agent *is* blameworthy or praiseworthy.

Enter the Alethic view. Rosen poses the following substantive question: "In what does the appropriateness of an appropriate response *consist*?" (2015, p. 70) and takes up the task of articulating *The Alethic View*. He writes:

> *The Alethic View*: The relation is definable in psychological and semantic terms. Like all emotions, the reactive emotions involve belief-like mental states or *thoughts*. Fear involves the thought that one is in danger, misery the thought that things are going badly, and so on. These emotions are appropriate in the intended sense only when the relevant thoughts are true, and this suggests an analysis: For an emotion to be appropriate *just is* for its ingredient thoughts to be true. In particular, for it to be appropriate to resent X for A *just is* for the thoughts implicit in resentment to be true of X and A.
>
> (ibid., pp. 70–71)

In a similar vein, Jada Twedt Strabbing (2019) defends her own version of the Alethic view of moral responsibility—*The True-Thoughts View*—according to which an agent "is accountable for an action if and only if a reactive attitude is appropriate toward her on account of the action" (2019, p. 3122). Like Rosen, Strabbing supports the view that emotions have representational content, which enables them to be assessed as either accurate or inaccurate with respect to their intentional objects. She writes:

> Reactive attitudes can be assessed in terms of accuracy because, as object-directed emotions, they have representational content, making them accurate if and only if their representational content is true.
>
> (2019, p. 3122)

Armed with an account of the reactive attitudes as essentially object-directed emotions with representational content, Strabbing claims that an instance of a reactive emotion "is fitting if and only if its constitutive thoughts are true" (ibid., p. 3124). Anew, we see the machinery of the Alethic account at work. In essence, this strand of moral responsibility theory can be said to endorse three related claims: (i) being morally responsible is a matter of being a fitting target

[10]The fact that having a reactive emotion would lead to a morally good/desirable outcome is not the only moral consideration that might count in favor of having the emotion. There could perhaps also be *deontic* considerations counting in favor of having the emotion, such as it being morally *right* to have it, or morally *required* to have it, etc.

[11]D'Arms and Jacobson (2000a, 2000b) call it *the conflation problem*, whereas Rabinowicz and Rønnow-Rasmussen (2004) call it *the wrong kind of reason problem*.

of a reactive emotion; (ii) emotions have representational content; and (iii) the fittingness of emotions should be understood as representational accuracy.

Now that the key features of the Alethic view have been mapped out, it is time to unveil its fundamental underpinnings. As we shall see in the next subsection, what lies at the heart of the Alethic view is a commitment to a particular conception of the emotions, namely *representationalism*.

3.1 | Alethic views' commitments to representationalism

As we have seen, the accounts of moral responsibility defended by Rosen and Strabbing are both varieties of ***MR*** (outlined above), according to which moral responsibility is analyzed in terms of *appropriate* negative or positive reactive emotions. The distinctive feature of both Rosen's and Strabbing's accounts is its fundamental commitment to the idea that reactive emotions are—at least partly—constituted by representational mental states with a mind-to-world direction of fit. Thus, the Alethic view builds upon a view of the emotions that has recently been labeled *representationalism* (see, e.g., Deonna & Teroni, 2022; Naar, 2021; Schroeter et al., 2015). On this view, emotions are (whatever else they are) in the business of representing their objects, which makes them suitable targets of assessment in terms of whether their representational content accurately matches the represented object.

Cognitive theories of emotions (e.g., Nussbaum, 2001, 2016; Solomon, 1973) and perceptual theories of emotions (e.g., Prinz, 2004; Roberts, 2003; Tappolet, 2016) are both varieties of representationalism, and both camps acknowledge the idea that emotions have representational content.[12] According to cognitive theories, emotions are essentially constituted by an evaluative judgment, thought, or a belief. The representational content of an emotion is understood in terms of the semantic content of its constitutive evaluative judgment, thought, or belief. In opposition to this cognitivist picture, several theorists have argued that the representational content can be spelled out without equating it with the semantic content of any alleged constitutive belief or judgment (see, e.g., Döring, 2003; Prinz, 2004; Rossi & Tappolet, 2019; Tappolet, 2016).[13] This is the enterprise of perceptualism, which tries to make sense of the emotions as representing their objects under an evaluative guise in virtue of being perceptual experiences of evaluative properties—analogous to the way in which visual perceptions are experiences of properties like color, shape, etc.[14]

At any rate, the Alethic view is inherently committed to some form of representationalism, according to which emotions represent their objects as having evaluative properties, either in virtue of being—at least in part—concept-laden evaluative judgments/beliefs/thoughts, or in virtue of being direct evaluative experiences of their intentional objects.[15] Indeed, this is *the* distinctive feature of the Alethic view, along with its additional

[12]Representationalism with respect to the emotions is thus orthogonal to the cognitive-noncognitive divide in the philosophy of emotion.

[13]In fact, one may even consider it well-advised to avoid a strong form of cognitivism with respect to the emotions considering the many problems that have been leveled against such theories (see, e.g., D'Arms & Jacobson, 2003; Deigh, 1994).

[14]Perceptual theorists come in various stripes. For example, in his earlier works, Jesse Prinz (2004) defended a perceptual theory of emotions according to which emotions are noncognitive embodied appraisals. In short, Prinz took emotions to be constituted by patterned bodily responses to features of the organism's environment. The patterned bodily changes that constitute an emotion, Prinz argued, have been *set up to be set off by* certain properties of an object or event, for example, offensiveness or dangerousness, etc. Just like a fire alarm has been set up to be set off by smoke, bodily changes of emotions have been set up to be set off by significant stimuli in our environment. Thus, bodily changes that are constitutive of anger can be said to represent offensiveness in a noncognitive, nonconceptual, embodied way. Prinz's teleosemantic account of emotional representation differs from the way in which other perceptualists (e.g., Tappolet, 2016) cash out emotional representation. In recent work, Prinz seems to have abandoned the view that emotions have representational content in favor of an enactivist approach (see Shargel & Prinz, 2017).

[15]I am not excluding alternative approaches to cashing out the ways in which emotions could be said to represent their intentional objects.

assumption that the relevant question of appropriateness ought to be construed as a matter of accurate matching between the representational content of the emotion and its intentional object.[16]

4 | TWO VERSIONS OF THE ALETHIC VIEW

There is at least one important difference between the views espoused by Rosen and Strabbing. Rosen seems to accept the stronger claim that moral responsibility is *explained by* or *grounded in* the fittingness of reactive emotions. On the other hand, Strabbing explicitly states that she wants to remain neutral with respect to which side of the biconditional (***MR***) explains the other. In other words, while Rosen endorses ***MR*** as an analysis of *what it is to be* morally responsible, Strabbing defends ***MR*** as merely a biconditional. Accordingly, we can distinguish between a strong form of alethicism and a modest one (I shall call these *Strong Alethicism* and *Modest Alethicism*, respectively) and articulate the key difference between them in the following way: *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. By contrast, *Modest Alethicism* is neutral with respect to this explanatory priority. As we shall see, different kinds of arguments can be advanced against proponents of Alethic views depending on whether they endorse *Strong Alethicism* or *Modest Alethicism*. It is thus important to bear in mind the different commitments of *Strong Alethicism* and *Modest Alethicism*, or else the arguments advanced may lose their bite. I shall now offer two arguments against *Strong Alethicism*.

5 | ARGUMENTS AGAINST *STRONG ALETHICISM*

5.1 | The redundancy objection

The first argument I shall present is called *The Redundancy Objection*, and the aim of it is to show that if the ultimate end of *Strong Alethicism* is to account for moral responsibility in terms of representational accuracy, then the detour via emotional content ends up explanatorily redundant. Let us therefore begin by considering the fundamental structure of *Strong Alethicism* outlined below:

First step: Moral Responsibility (MR) is grounded in Fitting Emotion (FE)

Second step: Fitting Emotion (FE) is grounded in Accurate Representation (AR)

It is apparent from the structure of *Strong Alethicism* that if MR is grounded in FE and FE is grounded in AR, then it follows that MR is ultimately grounded in AR.[17] The resulting explanatory structure of *Strong Alethicism* can therefore be presented as follows:

[16]In addition, it is worth noting that proponents of Alethic views also believe that an examination of the reactive emotions serves an important epistemological role, since we could—in principle at least—unearth the necessary and sufficient conditions for being morally responsible by way of examining the representational content of the reactive emotions. I will come back to the epistemic issue in Section 6.

[17]One might perhaps doubt whether it does follow from the conjunction, "MR is grounded in FE" *and* "FE is grounded in AR," that "MR is ultimately grounded in AR." Perhaps the grounding relation at play here is not transitive. While I do not wish to deny the possibility that this might be the case, I cannot adequately explore this issue here. At any rate, Rosen himself accepts transitivity of grounding (2015, p. 73).

Strong Alethic Explanation (SAE): AR → FE → MR

The arrows in **SAE** signify the explanatory direction such that it is the *fact* that a representation is accurate that *makes it the case* that an emotion in which the representation is embedded is fitting, which again, *makes it the case* that the represented object is morally responsible.[18] If the upshot of *Strong Alethicism* is that MR is ultimately explained by AR, it starts to look as if the appeal to emotions is explanatorily redundant. The explanatory work that the emotions are supposed to do in the argument—as grounding moral responsibility—could equally be done by any other mental state with the same representational content. Let me elaborate.

Deonna and Teroni (2014, 2015) have convincingly argued that to properly understand and distinguish various emotion types (as well as other mental attitudes) from one another, we should resist appealing to different representational (or propositional) contents. The difference between, say, the attitude of *believing* and the attitude of *resenting* does not seem to be well explained by invoking different contents, since plausibly, a token belief and a token resentment can have identical contents. For example, I can *believe* that a colleague of mine talked badly about my philosophical work behind my back, but I can also *resent* that my colleague did it. In both cases, my belief and my resentment seem to involve the very same semantic content, namely: [*My colleague talked badly about my philosophical work behind my back*]. And it should be obvious that I could adopt a wide array of different attitudes toward such content. I could *suspect* it, *desire* it, *feel sad* about it, be *angry* about it, be *confused* about it, etc. Instead, Deonna and Teroni urge us to locate the difference between different emotions (and other mental attitudes) in the nature of *the attitude itself*, rather than in its putative content.

Consequently, any assumption to the effect that reactive emotions have representational content *and* that it is the accuracy of the representational content that explains the moral responsibility of the targeted agent, will render the appeal to the reactive emotions explanatorily superfluous. To borrow an example from Rosen (2015): Suppose that resentment involves the three distinct thoughts *x*, *y*, and *z*. *Strong Alethicism* and its explanatory commitment (**SAE** outlined above) would then yield the verdict that insofar as *x*, *y*, and *z* are all accurate of an agent A and A's φ-ing, resentment toward A is fitting, which makes it the case that A is morally responsible for φ-ing. However, it is perfectly conceivable that we could have a *belief* with the content [*x, y, z*], or a *suspicion* with the content [*x, y, z*], or an *assumption* with the content [*x, y, z*] and so forth. If this is correct, it follows that according to *Strong Alethicism*, moral responsibility could be explained in terms of any type of mental state insofar as that mental state has a representational content, the accuracy of which makes it the case that the targeted agent is morally responsible. This simply won't do. If an account of moral responsibility makes an appeal to the fittingness of *emotions*, then the account should pick out some distinct feature of *the emotions* in order to explain moral responsibility. The redundancy objection shows why an appeal to representational accuracy is wrongheaded.

5.2 | The internal instability argument

The second argument I want to offer against *Strong Alethicism* is meant to exhibit a deep *internal instability* of the account.[19] The argument comes in the form of a dilemma that arises when the account is put to the test and faced with a series of questions. In answering those

[18]The locution "represented object" is here meant to denote the *intentional object* or *target* of one's emotion, which is typically taken to be an agent when the subject matter is that of moral responsibility.

[19]I am grateful to an anonymous referee for urging me to clarify this argument and for providing comments that helped me improve it. Thanks also to Alexander Velichkov for helpful comments on this section.

questions, *Strong Alethicism* can adopt one of two strategies, one of which exposes the account to an infinite regress, and the other one to a deep structural disunity. Hence, *Strong Alethicism* is inherently unstable and problematic.

In order to expose the internal instability of the view, we must first draw attention to a highly influential account in axiology, namely the *fitting attitude analysis of value* (henceforth the FA-analysis).[20] According to the FA-analysis, 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. 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 toward X (Todd, 2014, p. 90).[21]

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.[22] This is precisely the kind of analysis that is at work in *Strong Alethicism*. Recall that according to the view, an agent X is blameworthy *in virtue of* the fact that it is fitting to have a blaming emotion directed toward X. In other words, an agent's blameworthiness is *analyzed* in terms of it being *fitting* to have a negative reactive emotion vis-à-vis the agent. The FA-structure of *Strong Alethicism* can be displayed as follows:

> ***Strong Alethic Blameworthiness***: For an agent X to be blameworthy is for it to be fitting to have some blaming emotion E toward X.
>
> ***Strong Alethic Praiseworthiness***: For an agent X to be praiseworthy is for it to be fitting to have some praising emotion E toward X.

The structures here are analogous to the generic formulation of the FA-analysis we saw above, and *Strong Alethicism* is thus a kind of fitting attitude analysis of what it is to be morally responsible (i.e., blameworthy or praiseworthy). But as we saw earlier, *Strong Alethicism* does not only offer an answer to the question of what it is to be blameworthy, it also provides an answer to the question of what it is for a blaming/praising emotion *to*

[20]The FA-analysis of value is often claimed to have its origins in the works of Franz Brentano ([1889] 1969) and A. C. Ewing (1948). For a nice historical exposé of the FA-analysis, see Rabinowicz and Rønnow-Rasmussen (2004). On the theoretical merits of the FA-analysis, see, for example, Olson (2004), Rabinowicz and Rønnow-Rasmussen (2004), Rabinowicz (2013), and Rønnow-Rasmussen (2021).

[21]The term "fitting" refers here to a *normative* relation (or, as some theorists would say, a normative property). The precise way in which that normative relation is to be understood is, of course, a debated issue in the axiological literature, but we can note that it is not uncommon among theorists to gloss fittingness as either a distinct deontic relation or as a relation explicable in terms of deontic concepts. For example, one might think that the fittingness of an attitude A toward an object X is best construed in terms of it being *required* to have A toward X, or that one *ought* to have A toward X, or that it is *permissible* to have A toward X. In a recent article, Selim Berker (2022) has convincingly demonstrated why fittingness cannot be identified with, or understood in terms of, either deontic or evaluative concepts. Rather, Berker argues that we must think of *the fitting* as its own normative family—distinct from the category of the deontic and the evaluative—and construe reasons pertaining to the fittingness relation as "considerations that contribute toward the case in favor of" (Berker, 2022, p. 52) some relevant attitude. In that sense, "reasons of fit" are not "deontic in nature" (ibid., p. 52), but rather *contributory* or pro tanto. I think that Berker makes a compelling case in favor of treating fittingness as its own normative category, and I shall, in accordance with his conclusions, understand the reasons involved in the fittingness relation as pro tanto normative reasons that are distinct from deontic and evaluative considerations. Thanks to an anonymous referee for urging me to address this issue and for calling my attention to Berker's article.

[22]The FA-analysis can also be construed as (merely) a *conceptual analysis* according to which the focal point is to clarify and analyze *evaluative concepts* rather than to understand the nature of evaluative properties or values themselves (see, e.g., McHugh & Way, 2016; Rabinowicz, 2013). Personally, I am not as interested in analyzing emotion-related evaluative concepts as I am with understanding the ontological status of those values that are related to human reactive emotions.

be fitting. And this, according to the view, is a matter of the emotion representing its intentional object accurately. Hence, if it is fitting to resent X, then that must be because resentment represents X accurately. So far so good. But at this crucial point, the relentless philosopher should ask the following: Under what guise do emotions represent their objects? How should we understand their representational content? And what makes those representations accurate?

As I see it, there are two possibilities here. The first possibility is to say that emotions represent their objects under a wholly *descriptive* guise—they attribute purely descriptive properties to their objects. The second possibility is to say that emotions represent their objects under an *evaluative/normative* guise—they attribute (at least some) evaluative or normative properties to their objects. The standard representationalist answer is to claim that the representational content is essentially *evaluative/normative*.[23] And indeed, both of my main interlocutors—that is, Rosen (2015) and Strabbing (2019)—take the representational content of the reactive emotions to be at least partly normative (evaluative or deontic). For example, Rosen claims that resentment toward X for A is partly constituted by the thought that "it was *wrong* for X to do A" (2015, p. 77), and "In doing A, X showed an objectionable pattern of concern" (ibid., p. 77). The reference to "wrongness" in the first thought and to "objectionableness" in the second thought makes clear that Rosen takes the representational content of resentment to be at least partly normative.[24]

Let us, for the sake of simplicity, focus our attention on Rosen's second thought, and in particular the reference to "objectionableness." On Rosen's account, then, resentment toward X on account of doing A is, in part, to represent X as having shown an *objectionable* pattern of concern in doing A. And surely, for *that* representation to be accurate, it must be the case that X's pattern of concern de facto had the property of being *objectionable*. What kind of property is this? Rosen, and other proponents of *Strong Alethicism*, must either say that it is an evaluative/normative property that can be analyzed in terms of some fitting attitude, or that it is an evaluative/normative property that cannot be so analyzed. Which alternative should they opt for? Let us first consider the fitting attitude route.

The Fitting Attitude Route: Taking this route, Rosen can claim that what it is for something to be objectionable is for it to be fitting to object to it. This would give the account a nice kind of unity by way of sticking to the FA-approach all the way down. The account started off by analyzing what it is for X be blameworthy in terms of it being fitting to have a blaming emotion E toward X. We then learned that fittingness is a matter of representational accuracy and so for a blaming emotion E toward X to be fitting is for E to represent X accurately. Now, plug in resentment. Resentment toward X is, in part, to represent X as having shown an objectionable pattern of concern. For something to be objectionable is for it to be fitting to object to it. So, resentment toward X is, in part, to represent X as having shown a pattern of concern to which it is fitting to object.

At this juncture, proponents of *Strong Alethicism* must face the following question: What makes it the case that it is fitting to object to some X? With their commitment to understand fittingness in terms of representational accuracy, we can expect proponents of the view to say that if it is fitting to object to X, then that must be because objecting to X represents

[23]I take it that many theorists regard emotions as essentially carrying evaluative content precisely because it seems to offer a nice way of distinguishing them from other mental states in our mental ecology. This does not preclude proponents of *Strong Alethicism* to cash out the representational content of the reactive emotions in purely *descriptive* terms instead of evaluative terms, such that the content would be wholly explicable by reference to some *nonnormative* features of the intentional object. However, the contemporary debate between attitudinalists and representationalists in the philosophy of emotion centers on the issue of how to understand emotions as *evaluative phenomena*. Since representationalists locate the evaluative dimension of the emotions in the representational content and since *Strong Alethicism* is committed to some version of representationalism, it would be slightly odd for their advocates to cash out the representational content in purely descriptive terms.

[24]Strabbing claims that resentment is partly constituted by the thought that "S could have done better" (2019, p. 3129), which she explicitly claims to be a normative thought.

X accurately. But now, a problem arises for the defender of *Strong Alethicism* since there is nothing preventing us from running the same sequence of questioning again. And to the extent that the proponent of *Strong Alethicism* sticks to its own FA-commitment, she will end up in an infinite regress. Let us illustrate this. Below is an illustration of how a proponent of *Strong Alethicism* would have to go about accounting for the property of being blameworthy and answering the subsequent questions of a relentless interlocutor. The arrows in the schema represent the grounding relations. Whatever is located on the left-hand side of the arrow is *grounded in* whatever is located on the right-hand side of the arrow (one can also simply replace the arrows with a "because"). So, in the first round, the proponent of *Strong Alethicism* will say the following:

Round 1: X is blameworthy → resentment of X is fitting → resentment of X represents X accurately → X has the property of being objectionable.

Now, we imagine the interlocutor posing the following question:

Question: Can the property of *being objectionable* be analyzed in terms of fitting attitudes?

Since our proponent of *Strong Alethicism* is dedicated to the FA-analysis of evaluative properties she must reply:

Answer: Yes.

We should now expect our proponent of *Strong Alethicism* to be able to give an account of the objectionable, and presumably her account will mirror her account of the blameworthy above:

Round 2: Y is objectionable → objecting to Y is fitting → objecting to Y represents Y accurately → Y has whatever property V that makes objecting to Y representationally accurate.

And the same questioning sequence can be run again and again.

Question: Can the property V be analyzed in terms of fitting attitudes?

Answer: Yes.

Round 3: Z is V → having some attitude A toward Z is fitting → A toward Z represents Z accurately → Z has whatever property W that makes A toward Z representationally accurate.

What this illustration shows is that insofar as the proponent of *Strong Alethicism* stays committed to its FA-aspirations all the way down the explanatory trail, the view risks ending up in an infinite regress. There is, however, a way to stop the regress, and that is to take the other route, namely, to deny that the property of *being objectionable* is an evaluative/normative property that can be analyzed in terms of some fitting attitude.[25] Let us consider this route.

[25]Just to be clear, the argument here does not hinge on the fact that we are examining the property of *being objectionable*. As long as proponents of *Strong Alethicism* accept that emotions have representational content (which they do) and that that content is, at least in part, evaluative/normative (which would be odd if they didn't), the argument can get going.

The Response-Independence Route: Taking this route, proponents of *Strong Alethicism* can stop the aforementioned regress by denying that the property that resentment represents its target as instantiating—that is, being objectionable—can be given a FA-treatment. They can argue that the property of being objectionable is a *response-independent* property, and not a property that can be analyzed in terms of fitting attitudes. Considering this route in a similar manner of illustration as above, we get the following:

Round 1: X is blameworthy → resentment of X is fitting → resentment of X represents X accurately → X has the property of being objectionable.

Question: Can the property of *being objectionable* be analyzed in terms of fitting attitudes?

Answer: No.

Denying that the property of being objectionable can be analyzed in terms of fitting attitudes acts as a stopper and sidesteps the aforementioned regress. But now, proponents of *Strong Alethicism* owe us a convincing story for why it is the case that *blameworthiness* can be analyzed in terms of fitting attitudes, but other evaluative/normative properties—like *objectionableness*—cannot. On the face of it, the property of being objectionable seems equally amiable to an FA-analysis as does the property of being blameworthy. If the property of being blameworthy can be analyzed in terms of it being fitting to blame, then why cannot the property of being objectionable be analyzed in terms of it being fitting to object?

Perhaps there is a story to be told here why blameworthiness ought to be analyzed in terms of fitting attitudes whereas other normative properties cannot, but the burden of proof lies with proponents of *Strong Alethicism* to provide it. Moreover, it must be stressed that even if they do provide such a story, they will still end up with a deep disunity in their account. The *Strong Alethic* account of moral responsibility will then result in a two-tier structure where the top tier offers a response-*dependent* analysis of something normative (of moral responsibility) that trickles down into a denial of response-dependence of the evaluative/normative at the bottom tier. The view starts off by putting a lot of faith in the merits of analyzing a normative property (moral responsibility) in terms of fitting attitudes but abandons its own faith once the underlying architecture of the view is laid bare. We can illustrate the resultant structure as follows:

Top Tier: The fact that attitude A is fitting vis-à-vis X *makes it the case that* X is A-worthy (or, A-able, A-some etc.).

Bottom Tier: The fact that X is A-worthy (or, A-able, A-some etc.) *makes it the case that* attitude A is fitting vis-à-vis X.

At the top tier, *Strong Alethicism* masquerades as a response-dependent account and assumes that fitting attitudes have explanatory priority over normative properties. But further down the edifice—at the bottom tier—the view metamorphoses into a response-independent account of normative properties and adopts a reversed explanatory priority between fitting attitudes and normative properties. This, I want to argue, is not some mere oddity; it is a substantive tension *within* the view. The tension consists in the fact that the view must accept *both* (i) that the fittingness of attitudes determines the instantiation of (at least some) normative properties and (ii) that (at least some) normative properties determine the fittingness of attitudes. Consequently, *Strong Alethicism* ends up with a substantive internal disunity. To be fair, this kind of disunity may perhaps be defended, but again, the burden of proof lies with

proponents of *Strong Alethicism* to provide a case—and the case better be a strong one—for why such an internal disunity should be accepted.

In conclusion, then, *Strong Alethicism* is vulnerable to a serious dilemma when the account is put to the test and faced with a series of questions. In answering those questions, *Strong Alethicism* must adopt one of two strategies. One of these strategies will expose the account to an infinite regress. In order to sidestep the infinite regress, the account must adopt the other strategy, but in so doing it will unavoidably open itself up to a substantive internal disunity, which at the very least needs to be defended. This, I submit, makes *Strong Alethicism* inherently unstable and deeply problematic.

6 | ARGUMENTS AGAINST *MODEST ALETHICISM*

6.1 | Unearthing responsibility conditions

The aim of the arguments presented against *Strong Alethicism* above was to underscore some of the problems associated with its explanatory commitment. As we saw earlier, there is a version of the Alethic view—namely *Modest Alethicism*—which does not rely on any explanatory priority between blameworthiness and the accurate representation of a blaming emotion. *Modest Alethicism* thus avoids some of the problems that *Strong Alethicism* faces. What, then, is the appeal of *Modest Alethicism*, and what sort of work are the emotions supposed to do if they do not provide an explanation of what it is to be blameworthy?

According to Strabbing (2019), paying attention to the reactive emotions is of great *epistemic value* since they are particularly apt to reveal the responsibility conditions. The methodology that Strabbing employs in uncovering the responsibility conditions is to "draw on intuitions about what agents are thinking in having reactive attitudes and about what thoughts extinguish reactive attitudes" (2019, p. 3126). Utilizing this methodology, Strabbing suggests that resentment is partly constituted by the thought that "in doing A, S expressed insufficient good will toward me" (ibid., p. 3127). Strabbing offers the following example in order to buttress the plausibility of this proposal:

> Suppose that Fred has promised to pick you up from the airport. He fails to show, and you resent him for it. "How thoughtless of him," you think. "If he cared more about me, he would have remembered." As you are brainstorming ways to tell him off, a mutual friend calls to say that Fred just lost his job and is completely distraught. Your resentment naturally disappears (if you are rational). Why? It is not because breaking his promise was permissible. Rather, the news negates your thought that Fred failed to pick you up from thoughtlessness or lack of caring—i.e., from insufficient good will—toward you.
>
> (ibid., p. 3127)

Suppose then, in line with Strabbing's proposal, that resentment involves the thought that the target of one's resentment expressed insufficient good will in doing A. Let us call this thought "C." C is thus the condition for being blameworthy that is revealed through the examination of resentment. On the assumption that this kind of methodology is a sound one, the question we should ask at this point is this: Should we stop with resentment, or should we examine the other negative reactive emotions as well? Resentment is surely not the only reactive emotion through which we engage with other people. We may experience disgust or fear in response to the abhorrent actions of other people. Sometimes we experience contempt. When people fail to live up to certain expectations, we may feel disappointment, or perhaps even hurt feelings if the expectation in question happens to have some great personal import. These are

all examples of significant reactive emotions, and if reactive emotions are supposed to reveal responsibility conditions, then presumably we should treat these as potential candidates as well.[26] Suppose, then, that an examination of these various emotional responses unveils different "constitutive thoughts"—we can call them D, E, and F—none of which entails C (the thought revealed to us by resentment). We should then conclude that there is a multiplicity of blameworthiness conditions—namely, C, D, E and F—that are all individually sufficient for blameworthiness.

Of course, such a result is—in itself—not necessarily a problem. Those who have the intuition that blameworthiness can be instantiated in multiple different ways may indeed welcome such a result. However, a problem does arise for those who do not think that there is an abundance of blameworthiness conditions, and who are under the impression that only a limited class of emotions (say, guilt, resentment, and indignation) can play this epistemological role. The problem is this: anyone who insists that we should restrict our investigation to, say, guilt and resentment, must be able to provide some independent justification for why only these types of emotion qualify as relevant in unveiling responsibility conditions *without* presupposing the very conditions they are supposed to illuminate. Or, put differently: How does one determine which reactive emotions are *not* candidates for revealing the conditions for being blameworthy and praiseworthy without presupposing what those conditions are (or must be)?

Perhaps there is a way of providing such an independent justification, but I must admit that I cannot really see how this ought to be done. Accordingly, if no such justification can be provided, we seem to have no choice but to accept that any negative reactive emotion may reveal a condition for being blameworthy, and we might end up with as many individually sufficient conditions as there are negative reactive emotions. While I don't see this as a problem myself, I suspect that there are theorists who may regard this as an uncomfortable conclusion.[27]

6.2 | Criticizing emotional reactions and absences thereof

We can surely all recognize the fact that we sometimes criticize each other's emotional reactions. Not only does this dimension of our social lives seem psychologically inescapable, but there also seems to be some underlying normative justification for such a practice. One might thus wonder what this kind of criticism amounts to. With great persuasion, Sigrun Svavarsdottir (2014) has pointed out that although our unfitting emotions are indeed criticizable, it seems quite peculiar to think of the criticism as a charge of *inaccurate representation*. Indeed, think of everyday cases where we either criticize someone, or we ourselves are being criticized for having an (allegedly) unfitting emotion.[28] In such situations, it does not seem like what is being

[26]There has been a rising trend in recent years among scholars concerned with moral responsibility to examine other reactive emotions than the standard trio of resentment, indignation, and guilt. Listing only a few examples we have: *contempt* (Mason, 2003), *hurt feelings* (Shoemaker, 2019), *disappointment* (Telech & Katz, 2022), and *fear* (Pál-Wallin, forthcoming).

[27]To be clear, this is not a criticism against Strabbing in particular. The argument is only meant to expose a problem for those theorists who believe that only a particular set of emotions can unveil responsibility conditions, namely, they must provide a justification for why only their selected set of emotions can unveil these conditions without presupposing what these conditions are. Theorists who strive for ecumenism or pluralism with respect to the responsibility conditions will not be bothered by this argument, they may simply say: "Yes, we *should* indeed investigate all the emotions and establish all their different fittingness conditions!" Strabbing might be one of these theorists, but then again, the criticism is not intended for her. I cite Strabbing here only as a way of introducing the general idea that we can uncover the responsibility conditions by investigating the emotions and their putative constitutive thoughts.

[28]Admittedly, we usually don't use the expression *unfitting* in ordinary parlance when criticizing each other's emotions. However, we surely use locutions such as "inappropriate," "out of place," "misplaced," or "not making sense," etc. I take it that what people refer to when using expressions like these is the more technical notion of unfitting.

pointed out is that there is an inaccurate representation that needs correcting. Rather, the complaint seems to be one that has to do with how the agent *reacted to* or *was engaged* vis-à-vis the object. To be sure, such criticism is likely to be voiced through utterances like "Why are you angry? He didn't do anything offensive!" as an attempt to make apparent to the interlocutor that he is mistaken about the evaluative status of the targeted agent. But such remarks are hardly made as a way of conveying a representational misfire. Underscoring the mistaken evaluative status is surely just a way of communicating the fact that there was something fundamentally flawed with the way in which the angry agent was affectively engaged vis-à-vis the target.

Not only is it common for us human beings to criticize the unfitting emotional reactions of our fellow people, but we also seem to criticize each other for *failing* to have emotional responses deemed fitting by the circumstances. Here is an example that illustrates this familiar aspect of our interpersonal interactions:

> *Unemotional*: Kim and her spouse Aisha are attending a Christmas dinner party organized by Kim's company. During the predinner mingle, Kim and Aisha engage in a conversation with some of Kim's colleagues at which point one of the male colleagues, Saul, blurts out a sexist remark aimed at Aisha. The remark did not escape Kim's attention. On their way home from the Christmas party, Aisha confronts Kim and starts to grill her as to why she didn't get angry with Saul for the highly inappropriate remark.[29]

It is evident that the source of Aisha's discontent with her spouse Kim has to do with the fact that Kim *failed* to respond with anger even though the situation called for it (i.e., anger was the fitting response). The main issue here is to make sense of what Aisha's criticism amounts to. On the assumption that *Modest Alethicism* is true, and that the fittingness of an emotion indeed *just is* a matter of representing its object accurately, it would seem that the absence of a fitting emotion would amount to nothing more than an absence of an accurate representation. It would then seem that Aisha's criticism consists in a charge of *not representing things accurately*. This cannot be the correct verdict. As a case in point, let us now imagine that Kim—in response to Aisha's confrontation—acknowledges her failure to respond *with anger* vis-à-vis her colleague Saul, but in the same breath points out that it should not really matter since she did *form the belief* (or *judged*) that he was being chauvinistic and offensive.[30] Thus, Kim justifies the absence of anger by appealing to the fact that she did harbor a mental state (a *belief* or a *judgment*) the representational content of which involved an accurate evaluation of Saul—explicable in semantic terms—as chauvinistic and offensive.

Should Aisha be satisfied with such a reply? Not in the least. Not only does Kim's appeal to her true belief—and hence representational accuracy—strike me as an odd reply to Aisha's confrontation but also as being in bad taste. The issue at hand has little to do with a failure to represent things accurately. Aisha is discontent with Kim's failure to *engage affectively* in a way deemed fitting by the circumstances. Kim should have been *angry* with Saul. The fact that she *believed* or *judged* him to be offensive offers no solace. However, according to *Modest Alethicism*, which takes the fittingness of emotions to be nothing over and above representational accuracy, there is no basis for Aisha to be discontent. The absence of anger when anger is fitting would—according to *Modest Alethicism*—amount to nothing more than the absence of a mental representation that it would be accurate to have. If this is true, then it simply

[29]We can also imagine similar cases involving a failure to feel gratitude, guilt, shame, or any other reactive emotion.

[30]We can bracket any epistemic worries as to whether Kim is being insincere or confused when responding to Aisha and simply stipulate that she de facto did judge Saul to be chauvinistic and offensive.

should not matter that Kim did not respond with anger since, *ex hypothesi*, she did represent Saul accurately via her belief.

In response to this, it might be argued that considerations pertaining to fittingness only provide pro tanto reasons to have the relevant emotion. And since they only provide pro tanto reasons, they can be outweighed by other types of reasons (e.g., prudential or moral), such that the all-things-considered balance of reasons would count against having the emotion. With respect to Kim, then, it might be the case that she did recognize the fit-making reason to get angry, but it was outweighed by the moral/social reason not to make a scene.[31] Indeed, many theorists do seem to regard considerations of fittingness as only providing pro tanto reasons, and for argument's sake, we can grant this. However, the point of the case still stands, and we can illustrate this with two variations of the case.

In the first variation, we can simply *stipulate* that *there are no reasons counting against* Kim becoming angry with her colleague Saul. There are no moral, prudential, or other kinds of considerations weighing against responding with anger. The case would then, by stipulation, only involve reasons of fit. Now, if fittingness just is a matter of representational accuracy, then presumably "reasons of fit" are nothing over and above "reasons to represent things accurately," and in that case it should not matter if Kim represents accurately via the attitude of believing (or judging) or via the attitude of being angry. Again, there would then be no basis for Aisha's discontent. Kim is responding appropriately to the only existing reason in the situation, namely, a reason to represent things accurately.

In the second variation of the case, we can imagine that Kim indeed has some prudential or moral reason(s) not to make a scene. However, reasons "not to make a scene" will surely only concern *the outward expression* of the attitude, and not *the occurrence* of the attitude itself—after all, most human adults can experience emotions without expressing them in way that creates a scene.

On the assumption that Kim is the kind of person who has sufficient control over the outward expression of her emotions, she could very well be angry with Saul without creating a scene. In effect, the only reason that affects whether or not Kim should *be angry*—as opposed to *expressing her anger* in a way that would create a scene—is the pro tanto reason of fit.[32] And again, if that reason is nothing over and above a reason to represent things accurately, then it should not matter whether Kim represents things accurately via a belief, a judgment, or anger. An accurate representation is an accurate representation, and there would be no basis for Aisha's discontent since Kim is, *ex hypothesi*, responding in the fitting way. But the whole point of the case at hand is precisely that *it does make sense* for Aisha to be discontent with Kim. In light of the fact that Kim is Aisha's spouse—someone who she is supposed to share a life with, marked by shared values and common concerns—Aisha can reasonably expect of Kim not only to believe that Saul is offensive and chauvinistic, but also to be angry about it. To iterate a point made in Section 5, *believing* that something is the case and *being angry* about it are two vastly different kinds of attitudes, and it strikes me as quite clear that that difference should be captured in an account of the fittingness of these attitudes.

The guiding intuitions behind the case of Kim and Aisha is precisely that it seems to make perfect sense for Aisha to be discontent with Kim for not being angry in light of the facts, and that there is something very weird about Kim's attempt to defend herself against Aisha's

[31]I am grateful to an anonymous referee for urging me to address this issue.

[32]One might perhaps argue that Kim has prudential or moral reasons *not* to be angry on this very occasion due to some potential negative effects that that one instance of anger might have on her in the long term. We can grant this, but then it seems equally plausible to argue that Kim could have prudential or moral reasons *to be* angry on this very occasion due to some potential negative effects that the absence of anger (on this very occasion) might have on her in the long term—especially if we take into consideration the negative effects it might have on her relationship with Aisha. Hence, I think we can simply set these putative reasons aside.

confrontation by appealing to her accurate belief (and hence her accurate representation). Aisha's primary concern is not whether Kim managed to represent the facts accurately (in fact, Aisha may even assume that she did)—her concern is with the way Kim was left unaffected by those facts. But again, proponents of *Modest Alethicism*, which take the fittingness of an emotion to be a matter of representational accuracy, must contend that Aisha's discontent is unwarranted and that Kim's appeal to her accurate belief is perfectly valid. This strikes me as plainly wrong, and we should therefore conclude that emotional fittingness is not merely a matter of representational accuracy. To engage emotionally is not a matter of representing things in the world, it is a matter of expressing the way in which things in the world matter to one. Our emotions are expressions of our cares and commitments. They are ways of affirming that some things have a special kind of significance in one's life. An account of the fittingness of emotions should not only reflect all of this; it should place it at center stage.[33]

7 | CONCLUDING REMARKS

If the arguments presented here are along the right track, where does that leave us? Should we simply conclude that any attempt to understand moral responsibility in terms of fitting reactive emotions is doomed? I think not. Analyzing what it is to be blameworthy or praiseworthy in terms of the fittingness of reactive emotions strikes me as a fruitful method for shedding light on the nature of these evaluative properties. The arguments presented in this article have only been intended to underscore the many problems associated with a particular strand of responsibility theories, namely *Alethic views of moral responsibility*.

As we have seen, the problematic implications of Alethic views stem from their two core assumptions, namely: (i) that emotions have representational content and (ii) that the fittingness of emotions should be understood as representational accuracy. However, none of this should lead us to think that we now must abandon any project of analyzing moral responsibility in terms of fitting reactive emotions. The path ahead lies open for philosophers to do their due work and explore alternative ways of conceptualizing emotional fittingness. A promising step in that direction is—I submit—to start recognizing emotions as attitudes in their own right and to theorize about their fittingness in a way that does full justice to their multifaceted nature.

To be sure, much progress has been made in this area in recent years with the revival of attitudinalist approaches to the emotions (see e.g., Deonna & Teroni, 2012, 2014, 2015; Müller, 2017, 2018, 2019; Mulligan, 2007; Naar, 2020; Zamuner, 2015). Notwithstanding, there are lingering issues to tackle. For one thing, more theorizing needs to be done concerning the

[33]It may seem quite natural to think that beliefs about X are fitting if and only if they are correct/accurate. And if that is the case, then one might wonder why that definitional thought could not extend to other attitudes with a putative representational content, such as the emotions. Let me say two things: (1) I am inclined to object to the idea that beliefs about X are fitting if and only if they are correct/accurate. Fittingness (of attitudes) should be understood as a *normative* relation, and fit-making considerations must therefore be understood as considerations that provide *normative reasons* to have the relevant attitude. And I am inclined to say that the fact that a belief about X is accurate does *not*, by itself, provide a normative reason to have the belief. (2) Even if we grant the definitional thought that beliefs about X are fitting if and only if they are correct/accurate, we may still question why that definitional thought should extend to other attitudes just because they supposedly have representational content. From the alleged fact that a certain type of attitude has representational content, it does not follow that its fittingness is wholly a function of that representational content being accurate. A claim to that effect should be substantiated with good arguments. To me it seems dubious *at best* to carve out a partial aspect of such complex mental phenomena as the emotions and claim that their fittingness is a function of this partial aspect alone. Accordingly, even if we assume that emotions have representational content, the accuracy of that content should—*at best*—only specify the necessary conditions for their fittingness and not the necessary *and* sufficient conditions. The fittingness of emotions is a normative matter that must be understood by examining the emotions *holistically*, as attitudes in their own right. In that regard, I strongly agree with Hichem Naar, who writes that "the relation of fittingness between emotions and their objects is a normative relation over and above the relation of representation" (2021, p. 14). Thanks to an anonymous referee for a helpful discussion on this issue.

interplay between accounts of emotional fittingness, metaethical theories, and axiology. This will have to be a topic for another day. In the meantime, we should do our best to resist being enticed by the haunting specter of representation.[34]

ORCID

Robert Pál-Wallin https://orcid.org/0000-0003-4951-227X

REFERENCES

Berker, Selim. 2022. In *Fittingness: Essays in the Philosophy of Normativity*, edited by Chris Howard and R. A. Rowland. New York: Oxford University Press.

Brentano, Franz (1889). 1969. *The Origin of our Knowledge of Right and Wrong*. Edited by Oskar Kraus and Roderick Chisholm, translated by Roderick Chisholm and Elizabeth Schneewind. London: Routledge & Kegan Paul.

Crane, Tim. 1998. "Intentionality as the Mark of the Mental." In *Contemporary Issues in the Philosophy of Mind (Royal Institute of Philosophy Supplements)*, edited by Anthony O'Hear, 229–52. Cambridge: Cambridge University Press.

D'Arms, Justin. 2022. "Fitting Emotions." In *Fittingness: Essays in the Philosophy of Normativity*, edited by Chris Howard and R. A. Rowland. New York: Oxford University Press.

D'Arms, Justin, and Daniel Jacobson. 2000a. "Sentiment and Value." *Ethics* 110: 722–48.

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

D'Arms, Justin, and Daniel Jacobson. 2003. "The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)." *Royal Institute of Philosophy Supplement* 52: 127–45.

Deigh, John. 1994. "Cognitivism in the Theory of Emotions." *Ethics* 104(4): 824–54.

Deonna, Julie, and Fabrice Teroni. 2012. *The Emotions: A Philosophical Introduction*. London: Routledge.

Deonna, Julie, and Fabrice Teroni. 2014. "In What Sense Are Emotions Evaluations?" In *Emotion and Value*, edited by Cain Todd and Sabine Roeser, 15–31. New York: Oxford University Press.

Deonna, Julie, and Fabrice Teroni. 2015. "Emotions As Attitudes." *Dialectica* 69(3): 293–311. https://doi.org/10.1111/1746-8361.12116.

Deonna, Julie, and Fabrice Teroni. 2022. "Emotions and Their Correctness Conditions: A Defense of Attitudinalism." *Erkenntnis* 89(1): 45–64. https://doi.org/10.1007/s10670-022-00522-0.

Döring, Sabine A. 2003. "Explaining Action by Emotion." *The Philosophical Quarterly* 53(211): 214–30.

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

Graham, Peter A. 2014. "A Sketch of a Theory of Moral Blameworthiness." *Philosophy and Phenomenological Research* 88(2): 388–409.

Howard, Chris. 2018. "Fittingness." *Philosophy Compass* 13(11): e12542. https://doi.org/10.1111/phc3.12542.

Howard, Chris, and R. A. Rowland. 2022. "Fittingness: A User's Guide." In *Fittingness: Essays in the Philosophy of Normativity*, edited by Chris Howard and R. A. Rowland. New York: Oxford University Press.

Johansson Werkmäster, Marta. 2022. "Blame as a Sentiment." *International Journal of Philosophical Studies* 30(3): 239–53. https://doi.org/10.1080/09672559.2022.2121893.

Mason, Michelle. 2003. "Contempt as a Moral Attitude." *Ethics* 113(2): 234–72.

McHugh, Connor, and Jonathan Way. 2016. "Fittingness First." *Ethics* 126(3): 575–606.

McKenna, Michael. 2012. *Conversation and Responsibility*. New York: Oxford University Press.

Menges, Leonhard. 2017. "The Emotion Account of Blame." *Philosphical Studies* 174(1): 257–73.

Müller, Jean Moritz. 2017. "How (Not) to Think of Emotions as Evaluative Attitudes." *Dialectica* 71: 281–308. https://doi.org/10.1111/1746-8361.12192.

Müller, Jean Moritz. 2018. "Emotion as Position-Taking." *Philosophia* 46(3): 525–40.

Müller, Jean Moritz. 2019. *The World-Directedness of Emotional Feeling: On Affect and Intentionality*. New York: Palgrave-McMillan.

Mulligan, Kevin. 2007. "Intentionality, Knowledge and Formal Objects." *Disputatio* 2(23): 205–28.

Naar, Hichem. 2020. "Emotion: More Like Action Than Perception." *Erkenntnis* 87(6): 2715–44. https://doi.org/10.1007/s10670-020-00324-2.

Naar, Hichem. 2021. "The Fittingness of Emotions." *Synthese* 199: 13601–3619. https://doi.org/10.1007/s11229-021-03391-2.

[34]I am grateful to Toni Rønnow-Rasmussen, Matthew Talbert, Julien Deonna, Fabrice Teroni, Hichem Naar, David Shoemaker, Gunnar Björnsson, Wlodek Rabinowicz, Paul Russell, Björn Peterson, Ingvar Johansson, András Szigeti, Anton Emilsson, Alexander Velichkov, Marta Johansson Werkmäster, Roberto Keller, Agnés Baehni, Martin Sjöberg, Jiwon Kim, Mattias Gunnemyr, Daniel Telech, Laura Silva, and two anonymous referees for very helpful comments on this article.

Nussbaum, Martha C. 2001. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.
Nussbaum, Martha C. 2016. *Anger and Forgiveness: Resentment, Generosity, Justice*. New York: Oxford University Press.
Olson, Jonas. 2004. "Buck-Passing and the Wrong Kind of Reasons." *The Philosophical Quarterly* 54(215): 295–300.
Pál-Wallin, Robert. Forthcoming. "Fear as a Reactive Attitude." In *The Philosophy of Fear: Historical and Interdisciplinary Approaches*, edited by Ami Harbin. London: Bloomsbury Academic.
Prinz, Jesse. 2004. *Gut Reactions. A Perceptual Theory of Emotion*. New York: Oxford University Press.
Rabinowicz, Wlodek. 2013. "Value: Fitting-Attitude Account of." In *International Encyclopedia of Ethics*, edited by H. LaFolette, 1–12. Hoboken, NJ: Wiley-Blackwell. https://portal.research.lu.se/en/publications/value-fitting-attitude-account-of.
Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. 2004. "The Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114(3): 391–423.
Roberts, Robert C. 2003. *Emotions: An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.
Rønnow-Rasmussen, Toni. 2021. *The Value Gap*. New York: Oxford University Press.
Rosen, Gideon. 2015. "The Alethic Conception of Moral Responsibility." In *The Nature of Moral Responsibility*, edited by Randolph Clarke, Michael McKenna, and Angela M. Smith, 65–87. New York: Oxford University Press.
Rossi, Mauro, and Christine Tappolet. 2019. "What Kind of Evaluative States Are Emotions? The Attitudinal Theory vs. the Perceptual Theory of Emotions." *Canadian Journal of Philosophy* 49(4): 544–63.
Scarantino, Andrea, and Ronald de Sousa. 2021. "Emotion." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. The Metaphysics Research Lab, Department of Philosophy, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/emotion/.
Schroeter, Laura, François Schroeter, and Karen Jones. 2015. "Do Emotions Represent Values?" *Dialectica* 69(3): 357–80.
Searle, John. 1983. *Intentionality: An Essay in the Philosophy of Mind*. New York: Cambridge University Press.
Shargel, Daniel, and Jesse Prinz. 2017. "An Enactivist Theory of Emotional Content." In *The Ontology of Emotions*, edited by Hichem Naar and Fabrice Teroni, 110–129. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316275221.007.
Shoemaker, David. 2015. *Responsibility from the Margins*. Oxford: Oxford University Press.
Shoemaker, David. 2017. "Response-Dependent Responsibility; Or, a Funny Thing Happened on the Way to Blame." *Philosophical Review* 126(4): 481–527.
Shoemaker, David. 2019. "Hurt Feelings." *Journal of Philosophy* 116(3): 125–48.
Solomon, Robert C. 1973. "Emotions and Choice." In *Not Passion's Slave: Emotions and Choice*, edited by Robert Solomon. New York: Oxford University Press.
Strabbing, Jada Twedt. 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.
Svavarsdottir, Sigrún. 2014. "Having Value and Being Worth Valuing." *Journal of Philosophy* 111(4): 84–109.
Tappolet, Christine. 2016. *Emotions, Value, and Agency*. New York: Oxford University Press.
Telech, Daniel, and Leora Dahan Katz. 2022. "Condemnatory Disappointment." *Ethics* 132(4): 851–80. https://doi.org/10.1086/719512.
Todd, Cain. 2014. "Relatively Fitting Emotions and Apparently Objective Values." In *Emotion and Value*, edited by Cain Todd and Sabine Roeser. New York: Oxford University Press.
Tognazzini, Neal A. 2013. "Blameworthiness and the Affective Account of Blame." *Philosophia* 41(4): 1299–1312.
Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.
Wolf, Susan. 2011. "Blame, Italian Style." In *Reasons and Recognition: Essay on the Philosophy of T. M. Scanlon*, edited by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, 332–47. New York: Oxford University Press.
Zamuner, Edoardo. 2015. "Emotions as Psychological Reactions." *Mind & Language* 30(1): 22–43.

AUTHOR BIOGRAPHY

Robert Pál-Wallin is a doctoral candidate in practical philosophy at Lund University. His main areas of research include the philosophy of emotion, value theory, moral responsibility, and moral psychology, with a particular focus on issues concerning the normativity of emotions.