



## BIROn - Birkbeck Institutional Research Online

Friend, Stacie (2022) Emotion in fiction: state of the art. *British Journal of Aesthetics* 62 (2), pp. 257-271. ISSN 0007-0904.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/45947/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>  
contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

or alternatively

# Emotion in Fiction: The State of the Art

Stacie Friend

[This paper has been accepted by the *British Journal of Aesthetics*, to be published in January 2022]

*In this paper I review developments in discussions of fiction and emotion over the last decade, concerning both the descriptive question of how to classify fiction-directed emotions, and the normative question of how to evaluate those emotions. Although many advances have been made on these topics, a mistaken assumption is still common: that we must hold either that fiction-directed emotions are (empirically or normatively) the same as other emotions, or that they are different. I argue that we should reject this dichotomy.*

## 1. A Familiar Story

Emotions toward fictional characters are not only common; they are among the most valuable aspects of our engagement with fiction.<sup>1</sup> Imagine how much our experience would be diminished if we did not despise Iago in Shakespeare's *Othello*, pity Sethe in Toni Morrison's *Beloved*, or admire Amma in Bernardine Evaristo's *Girl, Woman, Other*. Yet the responses are also puzzling: Why do we care about characters when we know they do not exist?

Answers to this question in analytic philosophy trace their roots to two seminal papers: Colin Radford's 'How Can We Be Moved by the Fate of Anna Karenina?' (1975) and Kendall Walton's 'Fearing Fictions' (1978). Radford claimed that emotions toward fictional characters were irrational, whilst Walton distinguished them from ordinary emotions. Since the 1990s, Radford's and Walton's arguments have typically been framed in terms of the so-called 'Paradox of Fiction' (PoF). The PoF is a set of individually plausible but jointly incompatible propositions. Here is one version:

- (1) We experience emotions toward fictional characters, situations and events.
- (2) We do not believe (we disbelieve) in the existence of fictional characters, situations and events.
- (3) To have an emotion towards something, we must believe that it exists.

Over time, a consensus grew that improvements in the theory of emotion had defused the paradox by demonstrating that (3) was false, thereby undercutting the motivation for Radford's and Walton's conclusions and leading Robert Stecker (2011) to ask, 'Should we still care about the paradox of fiction?'

Since Stecker raised this question a decade ago, a variety of challenges to the consensus have emerged, whilst philosophers have brought new methodologies to bear on the puzzles. My purpose in this paper is to trace these developments. Before doing so, though, it is useful to sketch the origins of the consensus position. Here is a standard narrative:

Radford was the first to formulate the Paradox of Fiction (PoF).<sup>2</sup> His conclusion that emotions toward fictional characters are 'incoherent' and 'inconsistent' follows from the commitment to a set of contradictory propositions. Walton offered a solution to

---

<sup>1</sup> There are related topics which I set aside here, e.g., emotional responses to authors (Holliday, 2018) or how literature offers insight into our emotional lives (Simecek, 2015).

<sup>2</sup> See Friend (2020, p. 404, n.3) for a selection of papers making this claim.

the PoF by denying that emotions toward fictions are genuine. This entails a rejection of (1), removing the contradiction and with it the threat of irrationality. However, both Radford's and Walton's conclusions assume that emotions require belief in existence. This assumption made sense in the 1970s, when *judgementalism* or *cognitivism*—according to which emotions are wholly or partly constituted by beliefs—was widely accepted. But developments in the philosophy and psychology of emotions since then have demonstrated that emotions do not require beliefs. This means that we can solve the PoF by rejecting (3). If we reject (3), we are not committed to a contradiction, so we are not irrational. And we have no need to reject (1), so we can maintain that emotions toward fiction are genuine.<sup>3</sup>

The various claims in this story are by no means universally accepted. But the overall picture is often taken for granted.<sup>4</sup>

Nonetheless, the familiar story is false. Drawing on recent discussions which identify its flaws, I will sketch alternative interpretations of both Walton and Radford. These reinterpretations clarify the significance of two debates, concerning the *descriptive* issue of how to classify emotions toward fictions, and the *normative* issue of how to evaluate those emotions. Recent discussions contribute to answering these questions. I argue, however, that many of them inherit from Walton and Radford the mistaken assumption that we must choose between two mutually exclusive options: Either fiction-directed emotions are (empirically or normatively) the same as other emotions, or they are different. Rejecting this assumption opens more promising avenues for understanding our emotional engagement with fiction.

## 2. Classifying Emotions

As noted above, debate over the descriptive question was prompted by Walton's (1978) argument, standardly interpreted as answering Radford by denying (1). It is helpful to clarify what is wrong with this interpretation before turning to recent approaches.

One problem is that there is no evidence, either that Walton was replying to Radford, or that he was concerned with resolving the contradiction of the PoF. Still, Walton has made claims which appear to support (2) and (3) whilst denying (1). In describing moviegoer Charles, he writes, 'Charles knows perfectly well that the slime is not real and that he is in no danger' (1978, p. 6). Then: 'It seems a principle of common sense ... that fear must be accompanied by, or must involve, a belief that one is in danger' (pp. 6-7). He continues, 'Charles does not believe that he is in danger; so he is not afraid' (p. 7). Walton labels the physiological and psychological symptoms Charles experiences 'quasi-fear', and says that when Charles experiences quasi-fear in response to imagining the slime, he is 'make-believable afraid' (p. 13). According to the consensus, Walton thereby denies that affective responses to fiction are genuine emotions; instead, they are *quasi-emotions*, where quasi-emotions are make-believe facsimiles of the genuine article.

Commentators have highlighted inadequacies in this interpretation. First, there are reasons to doubt Walton's commitment to cognitivism (Matravers, 2014, p. 109). Walton recognizes counterexamples to the belief requirement (Dos Santos, 2017, p. 268). For instance, he says that a phobic about dogs might be afraid of Fido whilst believing that Fido is not dangerous (Walton, 1990, p. 208). More importantly, Walton never denies (1), that we experience genuine emotions toward fictions.<sup>5</sup> Walton agrees that Charles's response to the slime, or a reader's response to *Sethe*, is a genuine *emotion*; what he denies is that it is

---

<sup>3</sup> Stecker (2011) rightly points out that there may be other reasons to reject (1).

<sup>4</sup> A recent example is Adair (2019).

<sup>5</sup> Walton (2015) is explicit. See Friend (2016); Williams (2019); Gilmore (2020, p. 9, n.13).

genuine *fear* or *pity*—that is, an emotion of the same *kind* as pity or fear in other contexts. Walton takes for granted that emotions in response to fiction can have the same phenomenological and physiological profile (‘quasi-fear’ or ‘quasi-pity’) as emotions in other contexts; but this does not mean that they are of the same type.

Ultimately, Walton’s purpose is not to resolve a logical paradox or to delineate the ‘genuine’ emotions, but to describe a natural consequence of his account of engagement with fiction. For Walton, fictions and other representations invite us to participate in games of make-believe. Within these games we not only imagine people, situations and events; we also respond in a variety of other ways, including emotionally. Emotions within games of make-believe differ in kind from emotions in other contexts, not because they are less real, but because they occur within the scope of an imaginative project and therefore play a distinctive functional role in cognition (Stecker, 2011).

The issue, then, is not genuineness, but whether the emotions play such a role. Why does Walton think they do? In discussing Charles, he insists on what Dos Santos (2017) calls a ‘motivational requirement’: We take the phobic to be genuinely afraid because she is motivated to flee Fido, by contrast with Charles. Derek Matravers (2014) adapts the idea to emotions for others, interpreting Walton as saying that these require ‘the possibility of interaction between oneself and the object of emotion’ (p. 110). Matravers argues, though, that if this is right, the relevant contrast cannot be between fiction and non-fiction—histories may elicit emotions without any possibility of interaction—but between *confrontations* and *representations*.

The idea behind these proposals is that for Walton, ordinary emotions are characterized by a connection to action that is missing when the emotions occur within imaginative projects. But *contra* Matravers, the relevant connection is not merely practical; it is not just that temporal or spatial distance prevents action. Rather, emotions-within-imagination, including emotions toward fictional characters, cannot *even in principle* play the kind of action-guiding role that other emotions can play—typically because they involve imaginings rather than beliefs. I suggest that for Walton, these emotions are *quarantined* within the imaginative project, as Gendler (2003) proposes for imaginings themselves: their implications and effects are limited to a certain domain, which (among other things) prevents interaction with other mental states in producing action (Friend 2016; Cova and Friend forthcoming).

This is an empirical claim. And like their predecessors, philosophers who have recently criticized Walton draw on empirical evidence. For instance, Catherine Wilson (2013), Heather Adair (2019) and Jonathan Gilmore (2020) all marshal evidence from the psychology of emotion to argue (as Gilmore puts it) that ‘our affective responses to fictions are of the same kind, and are to be explained in the same way, as those we have for things in the real world’ (p. 59).<sup>6</sup>

There are, though, good reasons to doubt that this evidence is decisive. First, the experiments which underpin psychologists’ accounts of emotions typically deploy representations—pictures, stories, videos—taking for granted that the results apply to emotions across the board (Gilmore, 2020, pp. 77-78; Cova and Friend, forthcoming). Furthermore, studies in psychology or neuroscience are usually not concerned directly with fiction, so that conclusions are extrapolations from data about different topics (Stock, 2014; Pelletier, 2019). The few psychological studies explicitly comparing emotions toward fiction and non-fiction have been flawed in a variety of ways: for example, by failing to hold content constant across conditions (Humbert-Droz et al., 2020).

---

<sup>6</sup> Gilmore (2020) is discussed in the symposium in this special issue.

A promising development is therefore a new, interdisciplinary approach to fiction and emotion. Two separate teams of philosophers and psychologists have recently run studies contrasting emotions toward fiction and non-fiction (Sperduti et al., 2016; Sennwald et al., ms). Both teams presented participants with the same stimuli (video clips or texts) labelled either ‘real’ or ‘fictional’. In both studies, participants reported subjectively less intense emotion when viewing or reading ‘fictional’ sad stimuli. However, the results of implicit measures (skin conductance, heart rate, motor expression) conflicted: one study (Sperduti et al., 2016) found no difference between conditions, whereas the other (Sennwald et al., ms) found lower intensity in skin conductance for ‘fiction’.<sup>7</sup> The authors of the two studies drew conflicting conclusions about whether the emotions were different in kind (see Pelletier, 2019).

Despite the disagreement, the interdisciplinary approach has the potential to shed new light on emotional responses to fiction.<sup>8</sup> I doubt, however, that the illumination will be a definitive answer to the question of whether emotions toward fiction differ in kind from other emotions. This is not only because empirical results always require interpretation. It is because the question itself is misguided.

### 3. A More Nuanced Picture

I have argued that we should set aside the question of whether fiction-directed emotions are ‘genuine’; the issue is instead whether they are the same kind as emotions in other contexts. But even this question is problematic. It presupposes that there are only two choices: either the emotions are the same kind, or they are not. The evidence points to a more nuanced picture.

Consider two of the criteria for distinguishing emotions: experiential intensity and motivational profile. Neither lends itself to a dichotomy of emotion-kinds. Intensity comes in degrees, and differences of degree do not straightforwardly sustain a difference in kind (Cova and Friend, forthcoming). For instance, the experiments described above did not identify precise degrees of sadness associated with all ‘fictional’ (or ‘real’) stimuli. Even in our day-to-day lives we experience different degrees of sadness—ranging from mild disappointment to abject misery—in response to different events, at different times, in different circumstances. The same applies to other emotions.

Nor do emotions divide neatly into those which are motivational and those which are not. Those defending the genuineness of emotions toward fiction often point out that emotions can be motivating even though circumstances prevent action (e.g., Adair, 2019; Gilmore, 2020). But this does not show that all emotions are of a piece; admiration need not result in action even if circumstances permit. And there are a range of different relations that emotions bear to action. Some actions make sense only for confrontations; we run away from rabid dogs before us, not representations thereof. Other actions make sense only if we believe the situations are real (though represented), as when we visit graves, participate in commemorations, or donate money. Even emotions involving imagination vary in relation to action. Some fantasizing might be motivationally quarantined, but many other uses of imagination are not: for instance, imagining potential outcomes in planning or ‘affective forecasting’ to determine how one will feel in different situations.<sup>9</sup>

Emotions toward fiction and fictional characters are no different. Some fictions (like videogames) are interactive, designed to generate action directed toward the storyworld

---

<sup>7</sup> Note that the former used silent, five-second video clips, whereas the latter used five-minute audio-visual clips or one-page texts.

<sup>8</sup> See Humbert-Droz et al. (2020) for a proposal for future experiments.

<sup>9</sup> See Currie (2020, pp. 133-134) on affective forecasting.

itself.<sup>10</sup> Other influences are less direct. Though pity of *Sethe* does not motivate action directed *toward Sethe*, it could motivate action related to the real slaves who inspired Morrison's story, such as promoting reparations for their descendants. Therefore, rather than describing such emotions as quarantined—implying total isolation—we should describe them as *compartmentalized*, to a greater or lesser degree (Friend, 2020). Emotions are more compartmentalized to the extent that they remain contained within an imaginative project, and less so to the extent that they influence other mental states and behaviours.

Now, motivation and phenomenological intensity are not the only dimensions relevant to classifying emotions. Adair (2019) points to the scientific consensus that 'emotion involves (1) physiological, (2) evaluative, (3) motivational, and (4) phenomenological components' (p. 1061). Each of these components encompasses variations and gradations, as we have already seen for intensity and motivation. The evaluative component similarly resists reduction to a simple dichotomy, despite the standard contrast between belief and imagining. Emotions are associated with a wide range of different *cognitive bases* (Deonna and Teroni, 2012). As Teroni (2019) writes, 'Cognitive bases can vary immensely. At one end of the spectrum, there are emotions based on perceptual experiences; at the other, there are emotions based on complex inferences about unperceivable entities, as when one is afraid that a theory is wrong' (p. 116).<sup>11</sup>

The contents of these evaluations are also more complex than usually recognized. Most discussions assume that emotion-types are distinguished by 'formal objects' or 'criterial qualities', the evaluative properties attributed by the emotion, such as *danger* for fear and *undeserved suffering* for pity. Kris Goffin (2018) argues that the empirical evidence supports a more complicated account, according to which emotional content is comprised of a variety of lower-level appraisals which are not unique to any emotion-type (cf. Smith, 2017, p. 165). Consequently, two experiences of the same emotion-type may differ dramatically in (1)-(4). Anger at a child for drawing on the wall is likely to be felt and manifested in different ways from, say, anger at a corrupt government official or anger at a close friend who betrayed one's trust, let alone anger at the treacherous Iago.

The upshot of these considerations is this: Emotions are multidimensional, and each dimension—physiological, phenomenological, motivational or evaluative—is complex, admitting of a variety of degrees and distinctions. There is no dimension along which a dichotomy between 'fictional' and 'ordinary' emotions can be sustained.<sup>12</sup>

One might reply that only certain components of an emotion episode are relevant to questions of sameness or difference in kind. How compelling this reply is turns on other commitments regarding the nature of emotion. Within the discussion of fiction, I suspect that the insistence that we must make a choice is an artifact of the focus on genuineness. *Genuine/non-genuine* is a dichotomy not subject to gradations, and it is typically one that matters. By contrast, whether sadness concerning victims of floods in another country *is* or *is not* 'of the same kind' as sadness about a parent's death or Anna Karenina's fate does not seem an urgent question; and I doubt it admits of a non-arbitrary answer.<sup>13</sup>

Saying that imagining or fictionality can make a difference to emotion need not be interpreted to mean that this is the only or the most important difference. Walton's contribution is to explore *how* imagination and fictionality matter to emotional experience

---

<sup>10</sup> Recent discussions of interactive fictions include Van De Mosselaer (2020); Antonsen (2021). Though they are focused on so-called *i-desires* rather than emotions, I think that many of the same considerations apply.

<sup>11</sup> Teroni does not construe cognitive bases as a component of the emotion proper.

<sup>12</sup> I purposely do not prioritise any dimension as most important; for example, I do not claim that differences in cognitive base explain differences in other dimensions. Thanks to a referee for pressing me on this point.

<sup>13</sup> Some authors explicitly dismiss the question as irrelevant to their inquiries (e.g., Currie, 2020, p. 72; Antonsen, 2021, p. 43). These dismissals would look unmotivated if there were a significant choice at stake.

(Wilson, 2013, p. 82). Such explorations benefit from rejecting the dichotomy in kinds. One reason is that different emotion-types may interact differently with fictionality. For example, Wilson (2013) argues that grief is a peculiarly ‘poetic emotion’, insofar as literature plays a role in our collective and individual experiences of grief over genuine loss; here the contrast between fiction and real life is not so significant. On the other hand, Filippo Contesi (2015; 2016) suggests that disgust experiences may differ depending on the source. Contesi argues against Carolyn Korsmeyer (2011) that disgust is *ideational* rather than (merely) sensory. That is, what matters to the disgust response is not just the sensory features, but what they are features *of*; the same odour may be experienced as pleasant when the source is identified as cheese but disgusting when it is identified as faeces (Contesi, 2016, p. 349). For the same reason, there can be differences in responses to fictional and non-fictional representations, and representations and direct experiences, of disgusting objects.

Or consider Michelle Saint’s (2014) discussion of the *paradox of onstage emotion*, the puzzle of explaining how an actor may experience, say, the anger of her character, despite having no (real-world) reason to be angry. Saint finds two solutions promising: a solution which takes the actor to experience Waltonian make-believe anger, and a situationist solution that takes the specific theatrical circumstances to afford ordinary anger. I deny that these options are mutually exclusive. Saint’s perceptive analysis suggests instead that the actor experiences one of multiple variations on anger, in this case one that involves an imaginative component.<sup>14</sup>

In short, once we drop the traditional dichotomy between ‘genuine’ and ‘quasi’ emotions, there is no reason to insist either that there is a fundamental difference between emotions toward fiction and emotions in other contexts, or that there are no differences at all (Stecker, 2011). Although this is an empirical claim, I will argue that the same holds for the normative dimension, where the complexity of emotions also undermines any dichotomy.

#### 4. Emotions and Rationality

The normative issue is the question of how to evaluate emotions toward fiction as appropriate or inappropriate. Prompted by Radford (1975), much of the literature has concerned whether the emotions are rational or irrational *as a class*. A second debate focuses on whether fiction-directed emotions are governed by distinctive norms, so that what makes *particular* responses appropriate differs between fiction and other circumstances. In my view, the first question—when properly understood—is a version of the second.

To see why, we must first get clear about Radford’s argument. Contrary to the consensus view, Radford did not formulate the PoF, and his argument does not turn on a commitment to its three propositions (Friend, 2020). On the one hand, there is no doubt that Radford defends versions of both (1) and (2). With respect to (1), he maintains both that we are genuinely moved by fiction, and that there is no difference in kind between the emotions. With respect to (2), he denies that we ever suspend disbelief or ‘forget’ that we are engaged with the non-existent. On the other hand, Radford rejects (3) as a general statement about the nature of emotion. When Radford (1975) makes claims like ‘It would seem that I can only be moved by someone’s plight if I believe that something terrible has happened to him’ (p. 68), his concern is how we respond *in ordinary circumstances*, outside fiction (Cova and Teroni, 2016; Friend, 2016). Radford is right that in *those* circumstances, we do not normally feel fear when we do not believe there is danger, pity when we do not believe there is suffering, and so on.

In any case, Radford’s concern is not with propositions *about* our emotions, as expressed by the PoF, but with ‘our being moved’—that is, the emotions themselves. A

---

<sup>14</sup> Nils-Hennes Stear (2017) proposes something similar in his solution to the ‘paradox of sport’.

cognitivist might claim that because emotions involve beliefs, there is a contradiction between pitying Anna Karenina and disbelieving in her existence. However, despite the nearly universal assumption to the contrary, Radford has never espoused cognitivism (Friend, 2020, p. 408). Not only is a commitment to cognitivism incompatible with Radford's firm insistence on (1) and (2); in later replies to critics, he explicitly rejects any belief requirement on emotion. Like Walton, he suggests that phobias constitute a counterexample (Radford, 1995, p. 72).

Cova and Teroni (2016) propose that Radford's target is instead 'an apparent incoherence in our affective reactions': existence beliefs matter in ordinary life, but not for fiction (p. 937). To resolve the paradox, in their view, requires giving an account of emotions which explains why our responses are different in different contexts.

However, this interpretation does not get at the heart of Radford's puzzle. First, Radford already acknowledges differences between emotions in different contexts. He thinks these differences are irrelevant to the puzzle, which arises because we are moved *at all* by fictional characters. Second, the approach loses sight of the normative focus of Radford's argument. He does not claim that we are 'incoherent' in the sense of responding differently in different circumstances. Rather, his claim is that our emotional responses to fiction are *in themselves* incoherent: that is, irrational.<sup>15</sup> The question is why.

My proposal (Friend, 2020) is that the answer turns on Radford's defence of the claim that emotions toward fiction do not differ in kind from emotions in other contexts. If they do not differ in kind, Radford thinks, they should be subject to the same normative constraints.

It is widely agreed that emotions are governed by norms of *fittingness* or *correctness*. That is, the emotions are fitting only if they accurately represent their objects. Emotions are ordinarily incorrect if they have no object, for instance, if there is nothing to fear or pity. They are subject to an existence norm we can articulate as follows:

(E) It is correct to respond emotionally only if the object of emotion exists.<sup>16</sup>

An incorrect emotion need not be irrational; one might have good reason to think one's emotion is directed at an object even if it is not. What is irrational is *knowingly* to violate the norm. Adapting Radford's famous example: Suppose you tell me a 'harrowing story' about your sister, Sue, and then tell me she does not exist. Despite being convinced that there is no such person, I persist in feeling overwhelmed by grief and pity; I say, 'I know that Sue does not exist, but I still pity her'. This certainly appears incoherent.

Radford's contention is that the same applies to pity of Anna Karenina or Sethe. Again, if the emotions are of the same kind, they should be subject to the same norms. Addressing Radford's challenge thus requires demonstrating that there *is* a normative difference between emotions in different contexts: that is, a reason why (E) applies to pity of Sue but not pity of Sethe.

Though he does not pose the challenge this way, Teroni (2019) proposes a solution: that because emotions have distinct cognitive bases, their rationality turns on different considerations.<sup>17</sup> Teroni describes two ways of conceptualizing the cognitive bases of fiction-

---

<sup>15</sup> Teroni (2019, p. 123, n.32) claims that Radford 'does not speak of irrationality, but of incoherence'. Though Radford does not explicitly use the term 'irrational' in his 1975 paper, he does use it frequently in later replies to critics (e.g., Radford, 1990). It is clear in context that when Radford describes emotions as 'incoherent' or 'inconsistent', he means that they are irrational: they do not make sense.

<sup>16</sup> Many critics of Radford argue that the emotions are *practically* or *instrumentally* rational; but Radford's concern is with *cognitive* ('epistemic' or 'theoretical') rationality. See Adair (2019) for discussion of both.

<sup>17</sup> Teroni (2019) distinguishes both from 'blob-emotions': immediate, non-cognitive responses like Charles's fear of the slime.



directed emotions: either they are based on imaginings, or they are based on beliefs about fictional content. He advocates the latter, but either way, the claim is that emotions toward fictional characters are rational so long as they are sensitive to evidence about what is fictionally true (or as I prefer, *storified*), not what is true (p. 123; cf. Adair, 2019, p. 1068). So Sethe is rational insofar as she suffers *in the novel*; her non-existence in reality is irrelevant.

If this is right, and the fittingness of certain emotions turns on what is storified rather than what is true, then (E) is irrelevant to those emotions. Moonyoung Song (2020) makes the point clear: ‘in order to avoid making most fiction-directed emotions unfitting and accordingly inapt, the fittingness of fiction-directed emotions should be understood in terms of *fictional truth* rather than *truth*’ (p. 50). That is, when we evaluate emotions toward fiction, we must *replace* truth with story-truth, especially where the truth concerns the existence of the objects of emotion. Call this the *substitution approach*.<sup>18</sup>

The substitution approach identifies a normative difference between fiction-directed and reality-directed emotions: truth is relevant to the latter but not the former. It thereby provides a way to resist Radford’s conclusion. If emotions toward fiction are unconcerned with truth, then (E) is suspended, and pity of Sethe does not constitute a knowing violation of the norm. I agree with this conclusion. At the same time, however, I believe that the substitution approach is fundamentally flawed. This has implications not only for how we reply to Radford, but also for the recent debate over the norms governing fiction-directed emotions.

## 5. Continuities and Discontinuities

The debate pits the *continuity thesis*, according to which emotions toward fiction are subject to the same norms as emotions in other contexts, against the *discontinuity thesis*, which says the norms differ.<sup>19</sup> In recent discussions, the debate has concerned epistemic norms, and in particular the *aptness* of emotions: whether they are fitting and justified (Gilmore, 2011, 2020; Song 2020).

One would expect advocates of discontinuity to point to the suspension of (E) in support of their position, but they do not. One reason is that theorists on both sides take this as an example of ‘the uncontroversial discontinuities that no one would deny’ (Song, 2020, p. 49). A second reason is that the same considerations seem to motivate both the substitution approach and the continuity thesis. When we pity Sethe, detest Iago, or admire Amma, we seem to respond to the same evaluative features that prompt these emotions in other contexts: that Sethe suffers (in *Beloved*), that Iago is treacherous (in *Othello*), that Amma succeeds against the odds (in *Girl, Woman, Other*). In these examples, our responses parallel those we would have in ordinary circumstances. If we substitute story-truth for truth, they appear to be fitting for the same reasons—just as continuity theorists argue.

Rather than focusing on (E), discontinuity theorists argue that, *having substituted story-truth for truth*, what counts as apt may be different for fiction than in other contexts. Examples include cases of *discrepant affect* (Nichols, 2006), such as amusement at macabre events in a dark comedy or sympathy with unlikeable or immoral characters. Gilmore (2011, 2020) appeals to differences in how emotions are justified. For example, tense music might be enough to justify fear in response to a thriller solely because the filmmaker uses it to elicit that response, whilst some names in fiction justify conclusions about the character’s traits (think of Becky Sharpe). According to Gilmore, these cases of *discrepant justification* suggest that fiction-directed emotions are warranted for distinctive reasons.

---

<sup>18</sup> Recent versions of the substitution approach are articulated Teroni (2019, p. 124); Gilmore (2020, p. 86); Currie (2020, p. 62).

<sup>19</sup> The terminology originates in Gilmore (2011).

Song (2020) defends continuity against both arguments. Concerning discrepant affect, she argues that the appearance of discontinuity may be removed either by recognizing that *salience* plays a role in aptness (amusement may be apt if a work downplays the horror); or because it is storified that violence does not have its usual consequences (as in *Tom and Jerry* cartoons); or because we are responding to features of the work rather than the depicted situation. Concerning discrepant justification, Song poses a dilemma. Either the putative reason ‘*speaks to the object’s having the emotion’s criterial quality*’ (p. 56)—the tense music indicates that the object is dangerous in the story—or it does not. If it does, the emotion is justified in the same way as emotions in other contexts, by the object’s having the relevant evaluative features. If it does not, then it provides no *epistemic* justification.

I am sympathetic to Song’s replies to the discontinuity theorist, and I return to this debate below. First, though, I want to focus on what the two sides have in common: the substitution approach. Despite their disagreement, discontinuity theorists rely on the substitution approach just as much as continuity theorists. According to the discontinuity theorist, emotions toward fiction are governed by distinctive norms. Arguments for this position rely on pointing to differences in our judgements, such as treating amusement at macabre events to be apt in response to fiction but not in other contexts. However, such differences constitute evidence of *normative* discontinuity only if the contexts are sufficiently similar in other respects, including the objects of emotion. If the emotions differed in different contexts because they were directed at different types of objects—for instance, real or unreal—then this would pose no challenge to continuity; we expect emotions toward different things to be different. The arguments support normative discontinuity only if the differing responses are to sufficiently similar objects. And the objects are similar only if we substitute story-truth for truth, discounting (for instance) the fact that in one case the macabre events are real and in the other not.

In my view the substitution approach faces two serious challenges, which undermine arguments on both sides of the (dis)continuity debate.

First, the approach requires a sharp distinction between those contexts in which truth determines aptness and those contexts in which it is replaced by story-truth. In which contexts do we make the substitution? In the literature we find three answers: when we are responding to (i) *works of fiction* or to (ii) *fictional characters*, or in the context of (iii) *imagining*. I reject all three (Friend, 2020). In brief: The difficulty with (i) is that works of fiction can prompt ordinary beliefs and corresponding emotions. If *Beloved* prompts me to pity the real slaves who inspired the story, my emotion is subject to ordinary norms of fittingness like (E). As for (ii), suppose that in reading Anne Rice I am terrified of the vampire Lestat. Despite knowing that he does not exist, my terror continues, so that I collect crosses and string garlic around the house. Though the fear is of a fictional character, it is irrational in virtue of violating (E). A similar problem besets (iii). If I imagine that my partner is having an affair, this does not justify anger towards him *in reality*; if I imagine burglars in the house, this does not justify (actually) calling the police.

The problems with (i)-(iii) arise because emotions prompted by fiction or imagining can be directed at the real world. When they are, we cannot substitute story-truth for truth; the emotions are subject to ordinary norms. This suggests an alternative: that we substitute when our emotions are restricted to the storyworld or our imaginative projects. It makes sense to fear Lestat *in the storyworld* or to be angry at my partner *within imagination*, as long as these emotions do not ‘escape’ to the real world. As we have seen, however, emotions toward fiction are rarely if ever entirely quarantined. Instead, they are compartmentalized to greater or lesser degrees, along a variety of dimensions. Matters of degree cannot sustain the sharp distinction between contexts assumed by the substitution approach, which requires either that

we substitute story-truth for truth, or that we do not. Emotions are too complex to support the substitution approach.

The second challenge is to the substitution approach's central claim: that there are some fictional contexts, however determined, in which truth is irrelevant to the aptness of emotions. This is just mistaken. Facts about the real world remain relevant even for emotions toward fictional characters and events. Currie (2020) argues that 'considerations of truth sometimes do and should constrain the responses of the spectator' (p. 64).<sup>20</sup> I agree with Currie that fictions which trade in false stereotypes or other inaccuracies may thereby invite inapt emotions (Friend, ms). For example, *Gone with the Wind* invites amused contempt at Prissy's antics, presupposing inaccurate racist tropes. Roland Emmerich's film *The Patriot* invites us to condemn terrible British atrocities during the American War for Independence, though the atrocities portrayed never actually transpired. If we reject the aptness of the invited responses in these cases, the concern is with what is true rather than (or in addition to) what is storified.<sup>21</sup>

Here is a further reason to recognize the relevance of truth (Friend, ms). Emotional intensity and duration are subject to considerations of correctness (e.g., Gilmore, 2020, 92). It is incorrect to respond with boiling anger to a situation that calls for mild annoyance or experience high levels of anxiety about something with which one can easily cope. Thus, other things being equal, the tragic death of a close friend warrants more intense and persistent grief than that of a stranger. Fittingness turns not only on whether the situation is tragic, but also on factors like personal relevance and coping ability.

There are good reasons to think that fictionality and non-existence play a similar role. We have already seen evidence that classification as fiction may reduce the intensity of emotions. More importantly here, reduced intensity is *appropriate* for emotions for fictional events. We criticize the reader who cannot recover from grief at Sethe's killing Beloved, or the spectator whose rage at Iago interferes with her personal life. These responses strike us as disproportionate. The claim here is not that we should *always* experience less intense emotions toward fictions than we do in other contexts; Chimamanda Ngozie Adichie's *Half of a Yellow Sun* justifiably elicits stronger emotions than a dry newspaper report on the Nigerian Civil War. The claim is rather that *other things being equal*—such as the quality and vividness of the narrative—knowing that a character does not exist or that an event never occurred should mitigate the intensity of emotion.

If this is right, it poses a serious problem for the substitution approach. What are the grounds for criticizing someone excessive grief over Beloved's death or debilitating rage at Iago? That the characters do not exist, and the events never occurred. These are truths about the real world. If they are relevant to proportionate intensity, then truth cannot be wholly replaced by story-truth in determining the aptness of *any* emotions toward fiction. We should reject the substitution approach.

The substitution approach offers a reason why (E) is inapplicable to emotions like pity of Sethe: that truth is irrelevant to the aptness of fiction-directed emotions. If that is mistaken, we need an alternative reply to Radford. We have seen that (E) is inapplicable to emotions insofar as they are compartmentalized. Though this observation does not support the substitution approach, understanding *when* emotions are compartmentalized offers a reply to Radford (Friend, 2020).

Why is it appropriate to carry over our feelings about Sethe to real slaves, but not our feelings about Lestat to the real world? Because *Beloved* is realistic about the horrors of slavery, whilst Rice's horrors are mere fantasy. In other words, we compartmentalize our

---

<sup>20</sup> Currie (2020) is discussed in the symposium in this special issue.

<sup>21</sup> Moral considerations may also be relevant; but I do not believe that the moral and epistemic can be sharply distinguished in such cases (Friend, ms).

emotions to the extent that we take the content of a fiction or imaginative project to be inaccurate in respects relevant to the emotion. Compartmentalization is a matter of degree, turning on different appraisals. Because Sethe does not exist, pity of her should be compartmentalized with respect to its target. But we recognize that Sethe's suffering represents the experience of actual slaves; to this extent the pity is un compartmentalized and can be carried over to the real victims. Emotions carried over to real-world objects are subject to ordinary norms.

This observation suggests a different approach to the continuity/discontinuity debate. The debate concerns whether the norms governing emotions are continuous between one context where truth is relevant, and another where it is replaced by story-truth. I have argued that there are no such contexts. Instead, we should expect normative continuity to the extent that emotions are un compartmentalized, and discontinuity to the extent that they are compartmentalized. This reflects the patterns exhibited by the examples above. Where we take a story to be realistic in certain respects, we respond as we normally would in those respects; *Beloved* is realistic about the horrors of slavery, so we respond with pity. Where we take the story to be unrealistic, as with Sethe's existence or the violence in cartoons, we respond differently; in *Beloved*, we ignore the fact that there is no one to pity and in *Tom and Jerry* we do not worry about consequences.

Consider in this the 'sympathy for the devil' phenomenon, another example of discrepant affect. Katherine Tullmann (2016) argues that positive feelings about unlikeable or immoral characters are due to *fascinated attention*: the story directs our attention toward the fascinating traits and away from the less admirable (exemplifying Song's observations about salience). Tullmann rightly points out that this combination may be exhibited as much by non-fiction as by fiction (p. 126). However, it would be a mistake to think, either that we have *exactly the same* responses to accurate portrayals as to made-up stories, or that the same responses would be equally apt. If the murders in Oyinkan Braithwaite's *My Sister, the Serial Killer* had happened—if Korede was a real person who had systematically covered up the murders committed by her sister—the story would merit less sympathy and amusement, and more condemnation, than it does. It matters to our responses that these characters (and murders) do not exist.

In short, rather than answering Radford by holding that truth is irrelevant to aptness in response to fiction, we should say that considerations of truth are irrelevant *only insofar as* an emotion is compartmentalized. We know that fictional characters do not exist; emotions toward them are therefore meant to be compartmentalized with respect to their targets. When we are talking about the real world, though, the opposite is true. That is why pity of Sue violates (E), but not pity of Sethe. This does not mean that non-existence is entirely irrelevant to aptness, however, as with discrepant affect. As we have seen, emotions are complex, involving different dimensions thus different degrees of compartmentalization.

## 6. Conclusion

The last decade has witnessed significant developments in our understanding of fiction-directed emotions. Better interpretations of Walton's and Radford's arguments have helped to clarify what is at issue in classifying and evaluating these emotions. Interdisciplinary approaches have begun to uncover similarities and differences in our affective experience across contexts, whilst renewed attention to the normative dimension has highlighted the range of different factors relevant to judgements of rationality and aptness.

I have argued, however, that many of these discussions mistakenly assume that we must say either that fiction-directed emotions are (empirically or normatively) the same as emotions in other contexts, or that they are different. I propose that we reject this dichotomy.

Emotional experience is multifaceted, and the normative landscape is just as complex as the empirical. Exploring these landscapes is a promising direction for the next decade.<sup>22</sup>

## References

- Adair, H. (2019) 'Updating Thought Theory: Emotion and the Non-Paradox of Fiction'. *Pacific Philosophical Quarterly*, 100, pp. 1055–73.
- Antonsen, P. F. (2021) 'Self-Location in Interactive Fiction'. *BJA*, 61, pp. 41–52.
- Contesi, F. (2015) 'Korsmeyer on Fiction and Disgust'. *BJA*, 55, pp. 109–16.
- Contesi, F. (2016) 'Disgust's Transparency'. *BJA*, 56, pp. 347–54.
- Cova, F. and Friend, S. (forthcoming) 'How Does Fiction Elicit Emotions?' in Scarantino, Andrea (ed.), *The Routledge Handbook of Emotion Theory*. London: Routledge.
- Cova, F. and Teroni, F. (2016) 'Is the Paradox of Fiction Soluble in Psychology?' *Philosophical Psychology*, 29, pp. 930–42.
- Currie, G. (2020) *Imagining and Knowing*. Oxford: OUP.
- Deonna, J. and Teroni, F. (2012) *The Emotions: A Philosophical Introduction*. London: Routledge.
- Dos Santos, M. (2017) 'Walton's Quasi-Emotions Do Not Go Away'. *Journal of Aesthetics and Art Criticism*, 75, pp. 265–74.
- Friend, S. (2016) 'Fiction and Emotion', in Kind, Amy (ed.), *The Routledge Handbook of Philosophy of Imagination*. London: Routledge, pp. 217–229.
- Friend, S. (2020) 'Fiction and Emotion: The Puzzle of Divergent Norms'. *BJA*, 60, pp. 403–18.
- Friend, S. (ms) *Matters of Fact and Fiction*. Oxford: OUP.
- Gendler, T.S. (2003) 'On the Relation Between Pretense and Belief', in Kieran, M. and Lopes, D.M. (eds.), *Imagination, Philosophy, and the Arts*. London: Routledge, pp. 125–41.
- Gilmore, J. (2011) 'Aptness of Emotions for Fictions and Imaginings'. *Pacific Philosophical Quarterly*, 92, pp. 468–89.
- Gilmore, J. (2020) *Apt Imaginings*. New York: OUP.
- Goffin, K. (2018) *Emotional and Affective Representation*. Thesis. Ghent: University of Ghent.
- Holliday, J. (2018) 'Emotional Intimacy in Literature', *BJA*, 58, pp. 1–16.
- Humbert-Droz, S. et al. (2020) 'Lost in Intensity: Is There an Empirical Solution to the Quasi-Emotions Debate?' *Aesthetic Investigations*, 4, pp. 460–82.
- Korsmeyer, C. (2011) *Savoring Disgust*. New York: OUP.
- Matravers, D. (2014) *Fiction and Narrative*. Oxford: OUP.
- Nichols, S. (2006) 'Just the Imagination: Why Imagining Doesn't Behave Like Believing'. *Mind & Language* 21, pp. 459–74.
- Pelletier, J. (2019) 'Being Quasi-Moved: A View from the Lab', in Cova, F. and Réhault, S. (eds.), *Advances in Experimental Philosophy of Aesthetics*. London: Bloomsbury, pp. 123–41.
- Radford, C. (1975) 'How Can We Be Moved by the Fate of Anna Karenina?' *Proceedings of the Aristotelian Society*, 49 (Suppl.), pp. 67–80.
- Radford, C. (1990) 'The Incoherence and Irrationality of Philosophers'. *Philosophy*, 65, pp. 349–54.

---

<sup>22</sup> I am grateful for comments from Alex Grzankowski, Andrew Huddleston and Kris Goffin, as well as an anonymous referee, which much improved this paper. The remaining errors are my own.

- Radford, C. (1995) 'Fiction, Pity, Fear, and Jealousy'. *Journal of Aesthetics and Art Criticism*, 53, pp. 71–75.
- Saint, M. (2014) 'The Paradox of Onstage Emotion'. *BJA*, 54, pp. 357–69.
- Sennwald, V. *et al.* (ms) 'Is What I'm Feeling Real?'
- Simecek, K. (2015) 'Beyond Narrative: Poetry, Emotion and the Perspectival View'. *BJA*, 55, pp. 497–513.
- Smith, M. (2017) *Film, Art, and the Third Culture*. Oxford: OUP.
- Song, M. (2020) 'Aptness of Fiction-Directed Emotions'. *BJA*, 60, pp. 45–59.
- Sperduti, M. *et al.* (2016) 'The Paradox of Fiction: Emotional Response toward Fiction and the Modulatory Role of Self-Relevance'. *Acta Psychologica*, 165, pp. 53–59.
- Stear, N.-H. (2017) 'Sport, Make-Believe, and Volatile Attitudes'. *Journal of Aesthetics and Art Criticism*, 75, pp. 275–88.
- Stecker, R. (2011) 'Should We Still Care About the Paradox of Fiction?' *BJA*, 51, pp. 295–308.
- Stock, K. (2014) 'Physiological Evidence and the Paradox of Fiction', in Currie, G., Kieran, M., Meskin, A. and Robson, J. (eds.), *Aesthetics and the Sciences of Mind*. Oxford: OUP, pp. 205–226.
- Teroni, F. (2019) 'Emotion, Fiction and Rationality'. *BJA*, 59, pp. 113–28.
- Tullmann, K. (2016) 'Sympathy and Fascination'. *BJA*, 56, pp. 115–29.
- Van De Mosselaer, N. (2020) 'Imaginative Desires and Interactive Fiction: On Wanting to Shoot Fictional Zombies', *BJA*, 60, pp. 241–51.
- Walton, K. (2015) 'Spelunking, Simulation, and Slime: On Being Moved by Fiction', in his *In Other Shoes*. New York: OUP, pp. 273–86.
- Walton, K. (1978) 'Fearing Fictions'. *Journal of Philosophy*, 75, pp. 5–27.
- Walton, K. (1990) *Mimesis as Make-Believe*. Cambridge, MA: Harvard University Press.
- Williams, C. (2019) 'Why Quasi-Emotions Should Go Away: A Comment on Dos Santos'. *Journal of Aesthetics and Art Criticism*, 77, pp. 79–82.
- Wilson, C. (2013) 'Grief and the Poet'. *BJA*, 53, pp. 77–91.