

## Consuming Fictions Part III

### Immersion, Emotion, and the Paradox of Fiction

#### 11.1 Introduction

The previous two chapters argued that the best explanation of how we comprehend fictions, and of why we become immersed in them, will not appeal to *sui generis* imaginings. To summarize the positive picture I have so far proposed: recovering fictional content involves making a variety of different inferences, some of which require conditional reasoning, and none of which call for *sui generis* imaginings. Nor is there any obvious barrier to our immersion in fiction being explained by suitable pairs of beliefs and desires concerning the fictions themselves. The presence of ‘in the fiction’ operators within these beliefs and desires does not, I argued, put us at too great a cognitive distance from fictions to be immersed in them. Further, we’ve seen that such operators would in any case be required within *sui generis* imaginings, were they relied upon for fiction consumption instead.

There is still a missing piece, however. It can be summarized in the blunt question: why *care* about events you know to be fictional? To believe thus and such is true in a fiction is, after all, to believe that it is not *really* occurring. How could such beliefs be relevant to explaining our immersion in fictions? It might seem that these are precisely what must be “suspended” for us to become immersed in a fiction. A more intuitive picture, perhaps, is that we merely *imagine* the events of the fictions we enjoy. Then, it may seem, we are representing the events themselves and not dwelling on their fictionality—even if we still don’t believe them to be occurring in reality.

Unsurprisingly, I lack sympathy for this picture. I sought to undermine it last chapter by arguing that ‘in the fiction’ operators would be required within *sui generis* imaginings as well. But I agree that more still needs to be said in defense of the positive proposal that beliefs and desires are the states responsible for our immersion in fiction. Facing up to the blunt question above—why *care* about events you believe to be fictional?—requires addressing a perennial puzzle in aesthetics: the (so-called) *paradox of fiction*. I will explain the paradox below and spend most of this chapter developing a novel response to it—one that supports my broader project of showing how the A-imagining that occurs during the consumption of fiction can be explained in more basic folk psychological terms. Before

turning directly to the paradox, however, I want to mount an independent argument against the competing view that *sui generis* imaginative states are involved in generating emotional responses to fiction—one that does not rely upon the arguments of last chapter. Its conclusion will serve to motivate the reductive account of A-imagining to come.

## 11.2 The Emotional Irrelevance of What We Merely Imagine

As earlier discussed, it is now widely agreed that *sui generis* imaginative states are involved in the generation of fiction-directed affect (Doggett & Egan, 2012; Gendler & Kovakovich, 2006; Kind, 2011; Meskin & Weinberg, 2003; Nichols, 2006b; Schellenberg, 2013; Schroeder & Matheson, 2006; Spaulding, 2015; Weinberg & Meskin, 2006b). The idea motivating these views is that imagining that *p* generates affective responses rather like believing that *p*. Friends of the view explain:

According to the single-code hypothesis, then, the emotional systems will respond to pretense representations [*viz.*, *sui generis* imaginings] much as they do to parallel beliefs. That is, if the pretense representation that *p* gets processed by an affective mechanism, the affective outputs should parallel those of the belief that *p*. (Nichols, 2004a, p. 131)

There is no distinct anatomical region of the brain used for representing the merely imaginary... in spite of the fact that [fictional films] are all known simulacra, they are represented much as real things would be... once formed, these unimodal and multimodal representations have some of their characteristic effects including sending impulses to emotional centers.

(Schroeder & Matheson, 2006, p. 28)

When [in *Anna Karenina*] I read that some horrible misfortune has befallen Anna, that simulated belief [*viz.*, *sui generis* imagining] can activate my affect systems just as it would if I had a real belief that an actual person has suffered such a misfortune. Thus my emotional response is appropriately robust.

(Meskin & Weinberg, 2003, p. 24)

According to these views, if we imagine that lions are attacking the crowd at a circus, we will respond emotionally in much the same way we would if we believed lions were attacking the crowd at a circus. Some then offer cognitive-architectural explanations of this supposed mirroring. These theorists are not overly concerned with the fact that—by my lights, at least—our actual emotional responses to imagining are rarely *very* similar to what we would feel if we thought the imagined situation really obtained. I find that the emotions I experience in response to imagining that lions are attacking the crowd at a circus closely resemble

those triggered by the realization that the dishwasher needs emptying, and not so much the shock I would experience at judging that lions are attacking a crowd. But neither am I sure how strong these theorists intend the parallel to be.

In any case, defenders of these views can modify and qualify their general claims to accommodate exceptions. Nichols (2006b, pp. 469–72) proposes that our emotional responses to what we imagine are often not as intense as what we experience in response to believing the same propositions simply because we do not develop the scenario as richly in imagination. For instance, when asked to consider our moral duties in a world where most of mankind has been wiped out by an asteroid, we do not, typically, draw out in imagination the devastating inferences concerning our loved ones that we would if we came to believe the premise. According to Nichols, this difference in how thoroughly the scenario is developed is itself grounded in our desires and interests, which may differ depending on whether a scenario is believed or imagined. He explains:

I have strong and persistent desires for the health of my family, my friends, my colleagues. However, I often have no closely parallel desires about what happens in an imaginary scenario. Since what we desire affects recall and inference, this will have differential effects on what gets elaborated in the imagination. (2006, p. 471)

Differences in degree of “elaboration” between beliefs and imaginings, then, are called on to explain different degrees of emotional response.

I am not convinced by this response. For one, it isn’t clear that we need to richly develop a scenario when reacting—strongly—to a newly believed proposition. Emotional shock can be immediate, as when we suddenly realize that we’ve neglected an important obligation. At the same time, we often take in (and, presumably, *imagine*) very elaborate fictions—such as alien-invasion films—without much emotional response at all. True, we may lack relevant desires in such cases; but, on Nichols’s picture, desires are only relevant to explaining fiction-related affect insofar as they help to explain why our imaginings are richly developed, or not. (Otherwise, affect systems take input only from belief and imagination.) And, in the case of enjoying a fiction, we can assume the imaginings are richly developed simply as a function of our having to recover lots of fictional content. But there is another, broader problem with the idea that *sui generis* imaginings are the triggers for our immersion in a fiction.

The problem, roughly stated, is this: we can choose what to imagine; but we cannot similarly choose the affect we experience in response to a fiction. And yet it seems we *should* be able to choose the affect we experience in response to fictions if that affect really is determined by our imaginings (and all else held equal).

More formally: If our affective responses to fictions are caused in part by what we are currently imagining (in the sense that certain imaginings are causally

necessary form maintaining them), then we should be able to end such affective responses by stopping, or reversing, the relevant imaginings. But we cannot end such affective responses by stopping, or reversing, the relevant imaginings. Therefore, (by *modus tollens*) it's not the case that our affective responses to fictions are caused in part by what we are currently imagining (i.e., it is not the case that certain imaginings are causally necessary for maintaining our affective responses to fictions).<sup>1</sup>

An example: I am watching *Romeo and Juliet* and feel pity and shades of despair as the dual suicides unfold. To counteract these emotions, I should be able to imagine that Romeo and Juliet live happily ever after. For it is my imagining that they are committing suicide that, on the standard account, is essential to my experiencing negative affect.<sup>2</sup> Try as I might, though, I am unable to remove the pit in my stomach simply by imagining that Romeo and Juliet survive and live happily ever after. For I still believe that they die in the fiction! This belief easily overrides anything I might try to imagine.<sup>3</sup>

Examples are easily multiplied. We are engrossed in a heartwarming romantic comedy. Everything works out in the end and we are heart-warmed. But now, as the credits roll, we choose to imagine that the happy ending is upended—that the romantic relationship at the center of the narrative is permanently undermined by further (entirely avoidable!) misunderstandings. There is no difficulty in imagining such a thing. Yet it won't undermine our warm glow. By contrast, suppose that the film really contained the less satisfying ending in three additional minutes tacked on to the end. In that case, our emotional response really would be undermined. The reason is not that, in the second case, we are imagining that the relationship comes to a dismal end. We did that before, on our own, after the credits rolled. The reason is that we now *believe* that, in the fiction, the relationship flops.

It does not matter that when, of our own accord, we imagine the relationship ending, such an imagining would not be appropriate to the fiction, or that it is not what the fiction prescribes us to imagine. The fiction can *prescribe* whatever it

<sup>1</sup> Thanks to Shen-Yi Liao for spurring me to be more careful about the statement of this argument.

<sup>2</sup> Strictly speaking, this imagining should involve an 'in the fiction' operator, in line with last chapter's argument. I leave it off here, however, so as not to rely unnecessarily on that argument.

<sup>3</sup> One might object that, if I am right that "imagining that *p*, in response to fiction *F*" is just a matter of judging that, in the fiction *F*, *p*, then I *can't* imagine things that don't cohere with what I think is true in the fiction. The response here is that, when we imagine a fiction going in ways we believe it not to be going, this is not imagining "in response to fiction *F*" in the relevant sense. When I say that imagining that *p*, in response to fiction *F*, amounts to judging that, in *F*, *p*, this is offered as an account of the imaginings used to recover fictional content—the imaginings that authors intend for readers to undertake in response to their fictions. This leaves open the possibility of other imaginings that conflict with what we judge to be true in the fiction. For instance, I might, while viewing *Romeo and Juliet*, judge that it would be a happier ending if they both survived and raised a family together. On my view, this would be an A-imagining that Romeo and Juliet survive and live happily ever after. It is clear, on my view, why such an imagining does not change the emotions I feel: what matters (and what is beyond my control) is what I judge to be happening in the actual *Romeo and Juliet*, not what I judge to be a happier ending.

wishes; it doesn't thereby force our hand. What we imagine is still up to us. Perhaps, one might suggest, engaging with a fiction poses a special barrier on our imaginative capabilities, or makes us extremely unwilling to imagine things that conflict with the imaginings prescribed by the fiction. Could it be that we experience a kind of *imaginative resistance* to imagining propositions that conflict with what a fiction asks us to imagine?

I don't think so. It does not *seem* difficult, while enjoying a play or a film, to imagine that things are going otherwise than as the fiction suggests. We don't get the classic feeling of blockage characteristic of other cases of imaginative resistance—as explored, e.g., by Gendler (2000). Moreover, the most common explanations of imaginative experience don't fit the case: we are not imagining an impossible proposition, or a violation of supervenience relations (Weatherson, 2005); we are not imagining propositions we find morally problematic (Gendler, 2000). To simply infer from our lack of emotional change during the imaginings that we in fact fail to bring about the contrary-to-fiction imaginings is question-begging.

Another response might be that, once we have imagined something and it has triggered a certain emotional response in us, it is too late to “turn around” the emotion with a subsequent contrary imagining.<sup>4</sup> This could be so even if the initial imagining was in some sense “up to us.” By analogy: if I buy a lottery ticket and then win the lottery—the number choice having been “up to me”—I cannot undo the win by choosing to buy yet another ticket with a different number. Why, then, should I be able to undo an imagination-triggered emotion by generating new imaginings? However, the analogy is faulty. What's at issue is not whether we can change the fact that some past event occurred—be it a lottery win or the fact that an emotion was triggered. Of course we cannot. The question is what sort of causal intervention will be of the right sort to end an ongoing phenomenon. We know, for instance, that when anxiety is triggered at the recognition that the protagonist of a fiction is in serious danger, it will be relieved if we see that the protagonist is prevailing after all. Imaginings are supposed to have the power to effect that kind of reversal. And yet, from our examples, it seems clear that they do not.

We could consider other ways of preserving the idea that *sui generis* imaginings directly, or indirectly, generate fiction-related affect. But why do so when we have a promising alternative close at hand? What matters to my emotional state when enjoying a fiction is not what I am imagining, which can depart from anything the fiction prescribes, but what I am *judging* to be true in the fiction, combined with what I *desire* of the fiction. This offers a natural explanation for why I cannot adjust my emotional responses to fictions willy-nilly. For I can no more choose the judgments I make about what is true in a fiction than I can choose my judgments generally; *mutatis mutandis* for my desires about a fiction. It is my

<sup>4</sup> Thanks to Shen-Yi Liao for suggesting this worry.

beliefs (*qua* judgments) about what is true in the fiction that combine, in an unremarkable manner, with my desires about what is true in the fiction to generate fiction-related affect.

To bolster the plausibility of this answer, however, we need to understand why it is that we would ever have strong desires about what happens in a fiction, given that we never in fact enter into *sui generis* imaginative states when enjoying one. Answering requires that we confront and resolve the paradox of fiction. That is the project I turn to now. It will occupy me to the end of this chapter.

### 11.3 The Paradox of Fiction

The “paradox” of fiction concerns our emotional responses to fictions. That much, everyone agrees. Thereafter, things get murky. Some see the paradox as a puzzle about how (and if) our emotional responses to fiction can be *rational* or *warranted*. Others view it as a puzzle about how (and if) genuine emotional responses to fiction are *possible*.<sup>5</sup> Derek Matravers’ formulation of the puzzle captures this ambiguity well:

The problem is not primarily the fact that Anna [of *Anna Karenina*] does not exist but that...our regret that Anna was scorned *makes no sense* given that we believe that nobody was scorned. (Matravers, 2014, p. 112, emphasis added)

Does our regret make no sense as a matter of metaphysics—because one’s regretting that *p* implies, necessarily, that one believes that *p*? Or does it make no sense as a normative matter, in the way it “makes no sense” to go fishing in the bathtub? These distinct questions can easily be confused or considered equivalent. My focus here will be on the *normative* question. When we are puzzled about why we would ever have strong desires about what happens in a fiction—and resulting emotions when the desires are satisfied or frustrated—our puzzlement concerns how it could be at all reasonable or appropriate to *care* about events we know to be fictional. Nevertheless, we will see that the two questions—normative and metaphysical—interact in complex ways. Putting the normative question to rest requires careful consideration of the metaphysical one as well. Theorists have invoked *sui generis* imaginings in their answers to both questions; in neither case, I will argue, is it helpful to do so.

The “paradox” is often expressed as an inconsistent triad, each claim of which has independent plausibility (Cova & Teroni, 2016; Currie, 1990, p. 187; Friend, 2016;

<sup>5</sup> Friend (2016) finds the same two questions at work in discussions of the paradox of fiction—labelling them the “descriptive” and “normative” puzzles, respectively.

Gendler & Kovakovich, 2006; Kim, 2010; Nichols, 2004a; Van Leeuwen, 2016). I will examine the triadic formulation shortly. For now, the heart of the paradox can be seen in the following contrast. Typically, the reasonableness of an emotional response to a situation depends on our believing the situation to obtain. We are alarmed when we hear shattering glass in the night, fearing that someone is breaking in; our fear dissipates when we realize that the cat has simply knocked a cookie jar off the counter. It would be unreasonable to continue fearing thieves, once we are convinced there are none present. Simply knowing that thieves are not, in fact, breaking in is sufficient to quell our fears.

Not so in our engagement with films, novels, and plays. Knowing that fictional characters and situations do not exist does little to undermine the emotions we feel toward them. To address the paradox of fiction, we need to come to terms with this discrepancy. Why are *some* of our emotions tightly constrained by beliefs about the reality of their objects, while others, apparently, are not? And what does this discrepancy tell us about the nature and rationality of our emotional responses to fiction?

Before answering, two prefatory remarks. Not everyone will find it proper to speak of the *rationality* of an emotion; it is, however, common to hold that emotions are accountable to standards of appropriateness, fittingness, or aptness of some kind. *Ceteris paribus*, it is proper to fear what poses a danger, improper to resent what caused you no insult, warranted to pity undeserved suffering, unwarranted to feel gratitude to those who betrayed you, and so on. In querying the appropriateness or fittingness of our fiction-elicited emotions, we are asking whether such emotions are appropriate, given that we do not believe in the existence of the characters and situations at which they appear directed. Adopting distinctions from D'Arms & Jacobson (2000), we can further specify that the puzzle concerns the *epistemic* rationality of our emotional responses to fictions, as opposed to their prudential or moral rationality. We want to know whether the emotions we direct at fictions are warranted in the situations in which they arise, irrespective of any idiosyncratic practical goals or moral views we may have.

A further distinction to note is that between an emotion's being *warranted* given one's epistemic situation and its being *apt* (or *fitting*) given the nature of its object. Compare: it is common to speak of a belief's being warranted whenever a person has good reasons for holding it. One's evidence might make one warranted in believing that *p*, even if *p* is false. In another sense, however, the belief that *p* is infelicitous when *p* is false; it is not "apt" or "fit," given the way things really are. Similarly, we might say that an emotion is warranted given one's epistemic situation, even if it is inapt, or unfitting, because the object of the emotion lacks any feature to which the emotion is an appropriate response. For instance, feelings of resentment toward a friend might be warranted by the mere fact that you have reason to think the friend has betrayed your trust. But, if the friend did not, in fact, betray your trust, then the resentment remains inapt and unfitting of its

object. I will use the term “warrant” to capture the purely epistemic sense of appropriateness and the terms “fittingness” and “aptness” to capture the stronger sense of appropriateness that can apply to emotions, where the latter notions demand the object of the emotion to in fact have some feature that renders the emotion fitting or “correct.”

### 11.4 Some Background on the Paradox

Radford (1975) sparked the modern literature on the paradox of fiction, concluding that our emotional responses to fictions are indeed irrational—or, in the terms above, both unwarranted and unfitting. Most others have sought to rationalize our responses to fiction in one way or another, revealing them to have a kind of epistemic warrant after all. In an early and influential answer to the paradox, Peter Lamarque (1981) proposed that we can respond emotionally to the mere “thought that  $p$ ” without believing that  $p$ —and that this happens in many contexts outside of our engagements with fiction. More recent variations on this approach have emphasized the broader benefits of being cognitively “wired” in a way that allows for us to respond emotionally to non-actual situations (Gendler & Kovakovich, 2006; Van Leeuwen, 2016; Robinson, 2005; Weinberg & Meskin, 2006b; Nichols, 2004b). Related views hold that such responses are valuable pre-cognitive reflexes of a kind (Carroll, 2003, p. 524; Matravers, 2014, p. 117). In each case, our emotional responses to fictions gain a kind of vindication through their participation in a general pattern of response that has benefits to our survival. On many of these views (excepting Matravers (2014)), our *imaginings* are responsible for the related emotions, not our beliefs about the fiction. It is the tendency of *sui generis* imaginative states to evoke emotions similar to those generated by beliefs with matching contents that is supposed to explain—or explain away—the apparent inappropriateness of responding emotionally to a fiction.

I will argue, to the contrary, that what rationalize—or fail to rationalize—our affective responses to fictions are features of the fictions themselves, in accordance with a set of norms specific to the appreciation of fiction. Evolutionary considerations are irrelevant, as are claims about *sui generis* imaginings and cognitive wiring. Others before me have also appealed to the importance of fiction-specific norms when theorizing about our emotional responses to fiction (Livingston & Mele, 1997; Currie, 1990, pp. 213–15; Friend, 2016). I will appropriate aspects of Livingston & Mele’s (1997) and Gilmore’s (2011) helpful discussions. However, the relevance of fiction-specific emotional norms to resolving the paradox of fiction has never, in my view, been properly articulated; nor have such norms been used, as I will use them, to show how irrelevant imagination is to the questions at the heart of the paradox. Currie (1990, pp. 213–15), for instance, makes note of the special norms relevant to fiction-appreciation while still maintaining that *sui*



*generis* imaginings are nevertheless central to the paradox's resolution. More recently, Friend (2016, p. 226) observes that simply noting a difference in norms relevant to fiction consumption "does not resolve the problem" at the heart of the paradox and can look like "an *ad hoc* maneuver to avoid Radford's conclusion." At a minimum, we need some further explanation for why it is appropriate that the norms relevant to fiction-appreciation differ from those in place outside of fiction. My project here is to provide that further explanation, while dispelling the appearance that *sui generis* imaginings must play some role in it.

Carrying out this project requires first untangling the metaphysical and normative dimensions of the paradox (section 11.5). This will allow us to better appreciate what would, and what would not, undermine the reasonableness of our emotional responses to fictions. In the end, there is not a deep discrepancy between our everyday emotions and the emotions we direct at fictions. Both can be supported by reasons; and both can be shown unwarranted, and unfitting, when our reasons are overturned.

### 11.5 Distinguishing the Metaphysical and Normative Puzzles

We can begin at the first pages of Radford's (1975) article that sparked this debate. Radford asks us to consider how we would feel in two conditions: first, after reading an account of terrible suffering; and, second, after learning that the account is a fiction:

Suppose then that you read an account of the terrible sufferings of a group of people. If you are at all humane, you are unlikely to be unmoved by what you read. The account is likely to awaken or reawaken feelings of anger, horror, dismay, or outrage and, if you are tender-hearted, you may well be moved to tears. You may even grieve.

But now suppose you discover that the account is false. If the account had caused you to grieve, you could not continue to grieve. If as the account sank in, you were told and believed that it was false this would make tears impossible, unless they were tears of rage. If you learned later that the account was false, you would feel that in being moved to tears you had been fooled, duped. (1975, p. 68)

This thought experiment exploits what I will call a "rug-pull" structure: first, a person has an emotional response to a situation; and, second, revision of one of the person's beliefs pulls the rug out from under that response. The rug is pulled out in two senses: first, the emotional response itself dissipates; second, the response would no longer be warranted were it to continue, given one's newly revised beliefs. In Radford's example, we are asked to suppose that we've read (what we take to be) a genuine account of people who have suffered greatly. We are upset by the account.

Then we are told that it is only a fiction. At that point we are no longer saddened by the story. We are angry at being “fooled, duped.” Sadness is unwarranted now that we know the events didn’t actually occur. Learning that the events didn’t occur pulls the rug out from the sadness itself, both bringing it to an end (at least typically) and rendering it unwarranted were it to continue.

Other rug-pull cases are easy to find. Above I gave the case of hearing glass shattering at night and fearing thieves. As soon as we conclude that thieves are not, in fact, breaking in, the rug is pulled out from under our fear in the two senses described: we are, in all likelihood, no longer afraid of thieves; and, should our fear continue, it is no longer warranted. Reflection on such cases makes it tempting to extract a general moral about emotions: reasonably maintaining an emotion concerning a situation requires us to believe the situation to have features that would make our response fitting or apt. The principle gains credence as a kind of inference to best explanation for the changes that occur in rug-pull cases. (Here I am in agreement with Cova & Teroni (2016).) Its truth would explain why our emotions dissipate when we make the relevant discoveries. This general principle indeed forms one of three claims in the inconsistent triad mentioned earlier, which many see as capturing the core puzzle generated by our emotional responses to fictions. Here is how Gregory Currie presents the triad:<sup>6</sup>

- (1) We have emotions concerning the situations of fictional characters.
- (2) To have an emotion concerning someone’s situation we must believe the propositions that describe that situation.
- (3) We do not believe the propositions that describe the situations of fictional characters. (Currie, 1990, p. 187)

The most common response to the inconsistent triad is to reject condition (2)—the condition that links emotions to beliefs in the reality of the objects of the emotions (though Walton (1990) famously rejects (1); and I will reject (3).)<sup>7</sup>

There is less consensus on *why* (2) should be rejected. This is where it becomes crucial to distinguish the metaphysical and normative versions of the paradox. To dispel the attraction of (2), one needs to be clear about which of its interpretations is in question: is the “must” in (2) the must of rational warrant, or of metaphysical necessity? As each reading has some plausibility,<sup>8</sup> both need to be

<sup>6</sup> Not everyone structures the triad in exactly this way, though Currie’s formulation is frequently cited (e.g., by Matravers (2014, p. 105) and Nichols (2004a, p. 133)). It is also substantially the same as Gendler & Kovakovich’s (2006, p. 241).

<sup>7</sup> In rejecting (3), I do not hold that we believe the explicit or implicit content of fictions. Rather, I hold that we believe the propositions that describe the situations of fictional characters (just as (3) denies). These are propositions such as: in *The Great Gatsby*, Carraway is Gatsby’s neighbor; and, in *The Big Lebowski*, The Dude drinks White Russians.

<sup>8</sup> I disagree with Matravers’ view that “it is difficult to see that C [which is (2) in its metaphysical guise] has *any* intuitive appeal” (2014, p. 104, emphasis in original). Certain emotions fit the mold

addressed if one is to put the puzzle to rest. A few words about the metaphysical reading first.

If one comes to the table with a theory of the emotions on which central emotions such as fear, pity, and the like require an evaluative belief about the object of the emotions, then the metaphysical reading will of course seem correct. If, for instance, the emotion of pity contains, as a proper part, a belief that someone is suffering undeservedly, then there can be no pity without an attending belief in the object of the emotion. An attractive response to the resulting puzzle will then be to hold that the emotional states we experience when encountering fictions are similar to, but not strictly the same as, genuine emotions such as fear and pity. The responses may, instead, be “quasi-emotions” (Walton, 1990). It is no objection to this view to hold that our emotional reactions to fictions are too strong or phenomenologically rich to be “mere” quasi-emotions. Walton, the most prominent defender of the quasi-emotion view, allows that quasi-emotions can be both phenomenologically rich and physiologically acute. As he remarks with respect to the (imagined) theater-goer Charles, who appears frightened by the slime depicted in a horror film he is watching: “our question is whether [Charles’s] experience, however intense, was one of fear of the slime.” Walton clearly accepts that Charles has “a genuinely emotional experience,” while insisting that “it was not fear of the slime” (1990, p. 197).

Waltonian quasi-emotions differ with genuine emotions in their associated behavioral dispositions. Part of what it is to fear something, for Walton, is to be disposed to flee from it. “Fear emasculated by subtracting its distinctive motivational force,” he writes, “is not fear at all” (p. 202). Charles is not disposed to flee the slime by running out of the theater; so, however agitated he may be, what he is experiencing is not *precisely* fear of the slime—it is something else, something fear-like. To refute this view, by showing ordinary emotions like fear and pity to be wholly independent of such dispositions, one needs to mount arguments for a number of unobvious claims, such as that the behavioral dispositions that follow from believing that *p* are not essential to one’s fearing that *p*, worrying that *p*, being happy that *p*, and so on. Simply pointing out the obvious—that we get really worked up about fictions and other imagined events—does not substantively engage the theorist who takes the metaphysical reading of (2) to record an important feature of “everyday” emotions. This is why the inference from Charles’s agitated state when watching the slime to the conclusion that emotions are wholly independent of belief is “distressingly question-begging” (Walton, 1990, p. 200).

As it turns out, I will be one more who does not substantively engage the question of whether, for some emotions, (2) is correct in its metaphysical reading. Doing so would require me to defend a general theory of emotions, which is well

quite well. Regretting that *p*, or feeling ashamed that *p*, for instance, intuitively requires belief that *p*. At worst, the metaphysical version of (2) may be seen as an unwarranted overgeneralization from a few strong cases in its support.

beyond the scope of this chapter and indeed this book. But, in any case, I am not sure that Walton is wrong. Emotions are not *merely* beliefs and desires; but at least some emotions might well be complex states with beliefs and desires as essential *components* (Gordon, 1987). Such a view is a natural fit with the idea that our words for emotions gain a foothold in public discourse through their roles in folk psychological explanations—explanations that predict and explain behavior by ascribing beliefs and desires.

My interest remains in (2) considered as a *normative* principle. Yet, to ask if emotions can *reasonably* occur in the absence of relevant existence beliefs presupposes that they could *possibly* occur in the absence of such beliefs. Thus, simply moving on to ask the normative question, while leaving the metaphysical one unresolved, is risky on two counts. First, for those who think the metaphysical reading of (2) is correct, it may seem to leave the paradox of fiction unresolved; second, our answer to the normative question is left vulnerable to refutation by a theory of emotions that discovers deep connections between genuine emotions and beliefs about the objects of those emotions.

Fortunately, there is a way to table the metaphysical dispute to focus on the normative one, without begging any questions. For all sides agree that we have intense emotional reactions *of some sort* to fictions—raised heart-rates, sweaty palms, worry-like states, happiness-like states, anxiety-like states, and so on. We can then query together whether those responses to fiction are warranted and fitting, whether or not we agree that they are the very same *kind* of emotional states we experience outside of our engagement with fictions. Why, we can ask, should it be warranted and fitting to have intense worry-like, pity-like, and happiness-like states about fictional situations we *know never happened*?<sup>9</sup> After all, in the rug-pull cases that motivate (2), we don't normally retain worry-like, pity-like, or happiness-like states after learning that the objects of those states don't exist. (And, in the exceptional cases where we do, our doing so does not appear warranted.) So there remains a contrast in need of explanation, regardless of one's view on whether emotions require corresponding existence beliefs. In everyday rug-pull examples, judging that not-*p* tends to diminish, and render unwarranted, emotion-like states previously directed at the fact that *p*; not so in the case of our

<sup>9</sup> Walton thinks that once we conclude that Charles does not strictly *fear* the slime—because he is not disposed to flee it—there is no longer a need to rationalize any of his quasi-emotional responses. “One doesn't have reasons for things one doesn't *do*, like sweating, increasing one's pulse rate, involuntarily knotting one's stomach,” he writes. “So there is no need to attribute beliefs (or desires) to Charles that will render these responses reasonable” (1990, p. 199). The plausibility of this response hangs on describing Charles in purely physiological—and non-psychological—terms. Note that Charles has plenty of psychological reactions to the slime, whether or not we know how to describe them; and the warrant for these is certainly in question. Compare the case where one fears thieves breaking in, only to discover the cat having broken a cookie jar. If one remains agitated after learning that thieves are not breaking in, the agitation can no longer, on Walton's view, be called fear. Yet we can still query whether it is reasonable to remain emotionally agitated in one's new epistemic state.

engagement with fictions. Why, we can ask, is there this discrepancy? This is, I think, the key question behind the normative puzzle, and indeed behind the paradox of fiction.

The *other* puzzle still in play—which we leave unresolved—is a straightforward question about the nature of emotions. It is a question about which of the cognitive, sensory, and affective states that run together with an emotion in everyday contexts are essential to it. Our responses to fictions present dissociations among those features—e.g., we have the heart palpitations and anxiety of fear, without the tendency to run away. Fictions are, for that reason, enlightening test cases for a theory of emotions. But they are not the only such cases; similar dissociations and borderline cases can be found in the emotion-like states we experience when fantasizing, reasoning hypothetically, or reminiscing. So, to the extent that there is a non-normative puzzle afoot in the paradox of fiction—one we ignore by focusing on the normative question—it is simply the puzzle that borderline cases present for a theory of emotions. Having identified that puzzle for what it is, we can safely set it aside. Yes, something has been left unanswered, but it is nothing peculiar to our experience of fiction; nor will answering it alter our response to the normative puzzle. If our final theory of emotions tells us that the states we experience in response to fictions are not genuine emotions, we are still left with the normative question of why it is appropriate to have such strong emotion-like responses to fictions (when we generally don't in rug-pull cases). If, on the other hand, our final theory says that genuine emotions do not require beliefs in the reality of their objects, then we have rejected (2) as a metaphysical constraint on genuine emotions. But, in that case, we still have the normative puzzle before us: why are such emotions *warranted* (if they are), given what we know? In sum, no matter the formal theory of emotions we adopt, the normative puzzle needs answering independently.

### 11.6 Solving the Normative Puzzle: False Starts

Back, then, to our central question: the normative puzzle. To answer that puzzle, it obviously is not enough to point out that we *can* experience emotions (or emotion-like states) in response to fictions. This has been a recurring criticism of Lamarque's (1981) early solution. Lamarque famously rejects (2) on the grounds that: "I can be frightened by a thought or thought-cluster at a time when I am in no actual danger and do not believe myself to be in danger... I might find the thought of being stranded on a distant planet or being a monarch deposed in a military coup frightening *without supposing that this will, or even could, happen to me*" (p. 295, emphasis added). These claims are question-begging if offered as reasons for rejecting (2) as a metaphysical constraint, for the Waltonian reasons discussed above. Lamarque may only have intended them to undermine (2) in its normative guise, however. Even so, the argument still appears question-begging. Given

Lamarque's station in life—as a philosophy professor in a stable socio-political environment—it isn't at all obvious that his fear at the thought of being a monarch deposed in a military coup is reasonable. Certainly, there is nothing *prima facie* reasonable about fearing situations you think could not possibly befall you. If we were puzzled about how our emotional reactions to fictions could be warranted, reflection on Lamarque's cases will not leave us any less perplexed. This was, in effect, Radford's (1982) response: "Lamarque says... that I am frightened by my thoughts. But *they cannot hurt me*, and I know that of course... So the fear is irrational, absurd, incoherent, even though it is common place" (p. 262, emphasis in original).

On the other hand, it does seem reasonable—and not at all puzzling—to fear, or respond emotionally to, situations you think *might* befall you, even if you know they are not now the case. Gendler & Kovakovich (2006) pursue this thought as a means to overturning (2) as a general constraint on the rationality of an emotional response. Results in neuropsychology, they argue, support the claim that "without the tendency to feel something relevantly akin to real emotions in the case of merely imagined situations, we would be unable to engage in practical reasoning" (2006, p. 242). In support they cite work by Antonio Damasio and colleagues (Damasio, 1999; Damasio, Everitt, & Bishop, 1996) showing correlations between emotional deficits (in people with damage to prefrontal cortex) and impairments in practical reasoning. Damasio's findings suggest that our tendency to react emotionally when considering different possibilities is an integral part of our ability to plan our actions effectively. This has become known as the *somatic marker hypothesis*. Given these "facts about our cognitive architecture," Gendler & Kovakovich propose, "it is reasonable to employ the expression 'genuine, rational emotion' in describing both actual and fictional emotions" (2006, p. 243). ("Fictional emotions," for Gendler & Kovakovich, are emotions directed at fictions.) For "it is crucial to our ability to make rational decisions about various courses of action that we respond with genuine emotions to situations that we know to be non-actual" (p. 243). This point allows them to reject (2) in its normative guise.<sup>10</sup>

Yet, by tethering the reasonableness of our emotional responses to their usefulness in practical reasoning, they open the door to a reformulation of the inconsistent triad. Unlike the possibilities we consider during action-planning, fictions do not normally present us with different possible courses of action from among which we must choose. That is precisely why it seems odd to respond emotionally to them. Thus, such responses may still be an unwarranted (and impractical) over-extension of a tendency useful in action-planning to a context where it is

<sup>10</sup> It is worth noting that Gendler & Kovakovich's argument does nothing to undermine (2) as a metaphysical constraint, as all the data they discuss is consistent with Waltonian quasi-emotions being activated during hypothetical reasoning, in lieu of genuine emotions.

inappropriate, inapt, and even hazardous. In light of this, the defender of the paradox can simply reformulate condition (2) as:

(2'): To have an emotion concerning someone's situation we must (in the normative sense of 'must') believe the propositions that describe that situation, or believe the situation is of relevance to our future decision-making.

(2') is no less plausible than (2) itself; it is, in effect, an elaboration of (2). After all, no one ever questioned the rationality of worrying that you'll get a ticket if you fail to pay the parking meter; here you were thinking about an unrealized possibility—getting a ticket—and feeling some anxiety. Yet (2') remains inconsistent with our emotional reactions to fictions. Nor do Damasio's findings give us any reason to reject (2') as a general constraint on the epistemic (or prudential) rationality of our emotional responses. (See Matravers (2006, pp. 260–2) for related criticisms.)

Perhaps what Gendler & Kovakovich really want to highlight in their appeal to the somatic marker hypothesis is that, due to our cognitive wiring, *we just can't avoid* responding emotionally to fictions and the imaginings they inspire. Further, they might continue, given that our being wired in that way is essential to our healthy functioning outside of our engagement with fictions, we are hardly to be faulted for our emotional responses to fictions. This is clearly Van Leeuwen's (2016) position; and it appears consonant with that of Meskin & Weinberg (2003) and Schroeder & Matheson (2006) as well. Van Leeuwen proposes that our responses to fictions are a natural byproduct of the fact that mental imagery “automatically” triggers affect, and that we typically form mental imagery when consuming fictions. He offers empirical evidence that imagery is processed in perceptual areas of the brain that have especially close ties to neural areas underlying affective responses. In virtue of this shared neural pathway, imagistic and perceptual states are able to automatically trigger affect, which, he argues, facilitates bodily preparedness for action, supports the evaluation of future actions (in the manner of Gendler & Kovakovich), and enables empathy-based moral appraisal (Van Leeuwen, 2016, pp. 95–101). Were we to somehow bring an end to our automatic affective responses to fictions, Van Leeuwen warns, “it would involve destroying something on which three important agential capacities depend” (p. 105). Thus, “anyone who scorns the human propensity to respond emotionally to fictions has just not understood the consequences of doing away with that propensity.”

Similar points are made by a number of other theorists who see our responses to fictions as *affective reflexes*. By “manipulating such variables as speed, scale, lighting, and sound,” Noël Carroll proposes, “the filmmaker often appears to have direct access to our nervous system, bypassing the cerebral cortex and triggering automatic affective reflexes” (Carroll, 2003, p. 524). Matravers agrees with this way of analyzing Charles's (apparent) fear of the green slime in Walton's example: “What Charles describes as ‘being terrified’ simply could be a vivid way of

expressing the fact that he was shocked (an a-rational response therefore requiring no explanation in terms of objects) by a ‘fast movement towards the camera.’” Charles is “involuntarily startled by a sudden and surprising turn of events on the screen” (Matravers, 2014, p. 117). Jenefer Robinson concurs, defending a general theory of emotions as “pre-cognitive” appraisals:

Pre-cognitive affective appraisals do not discriminate between real and imagined scenarios... It does not matter to my emotion system (fear, anger, sadness, etc.) whether I am responding to the real, the merely imagined, the possible, or the impossible... (2005, pp. 145, 149)

Like Van Leeuwen and Gendler & Kovakovich, Robinson wants to vindicate such responses by appeal to their fitting a pattern that is advantageous to our survival. In reading *Anna Karenina*, she explains:

It is adaptive to be able to sense one’s wants and wishes, interests and goals to be at stake when reading and thinking about *Anna Karenina*... When I respond compassionately to Anna, I am sympathizing with her fate in a way that is socially adaptive. (2005, p. 148)

These responses to the paradox all miss the mark, in my view. Whether Gendler & Kovakovich and Van Leeuwen are correct in their interpretations of the empirical literature, or not, and whether Robinson is right that such responses are in some sense “adaptive,” or not, these putative rationalizations *concede too much*. They offer sub-personal excuses for emotions that need no apology. Suppose—just as a thought experiment—that each of these theorists is wrong in his or her empirical claims. Suppose that Damasio’s research assistant misread the data and that there is in fact no significant difference in the practical reasoning abilities of emotionally-impaired patients and controls. Suppose also that Van Leeuwen has misinterpreted the empirical evidence and that mental imagery has no special tendency to automatically trigger affective centers of the brain. And let us imagine that Robinson is simply wrong that it’s “socially adaptive” to respond emotionally to non-existent entities; it turns out that doing so keeps us home Saturday nights watching movies, instead of out meeting new folks. Would we then be forced to conclude that our emotional responses to fictions are unreasonable and unwarranted after all?

Certainly not. What matters for their warrant is whether there are *good reasons* for our emotional responses to fictions. Whether one has good reasons, I will argue, is not a matter of one’s neural wiring or evolutionary history.<sup>11</sup> The relevant

<sup>11</sup> To be clear, I don’t deny that we *can* have reflex-like affective responses to fictions—a startle-response to an unexpected explosion, say. But this is, in general, the wrong way to think about our



factors—factors that can render a triggered emotion fitting or apt—are closer to the surface. The next section develops this point through meditation on the norms specific to fiction-appreciation. An important upshot will be that it is not (2) in the triad that needs rejecting. Whether or not (2) *is* correct, our emotional reactions to fictions give us no special reason to reject it. It can be granted that an emotion is unwarranted when we do not believe that its object has features that make it apt; for emotions we experience in response to fictions take as objects the fictions themselves. Such fictions exist and, typically, have features that render the responses apt (or so I will argue). We can instead reject (3), the proposition that we do not believe the propositions that describe the situations of fictional characters. For the propositions that describe the situations of fictional characters, it turns out, involve ‘in the fiction’ operators. In the process, further reasons will emerge for rejecting the approach taken by Gendler & Kovakovich, Van Leeuwen, Robinson, and “single code” theorists such as Meskin & Weinberg (2003) and Nichols (2004b).

### 11.7 Believing It Is a Fiction and the Norms of Immersion

What could be a good reason for responding as we do to characters and events we believe to be fictional? Ironically, the *first* reason one ought to give for such responses is that we believe the characters and situations to be fictional. We take a wrong step when we immediately go on the defensive, granting that our emotional reactions to fictions lack the kind of basis we would expect of an ideally warranted emotional response—conceding that they are, like reflexes or controlled hallucinations, in conflict with our beliefs about reality. Instead we should reply that our belief in the *unreality* of fictional characters and situations is part of what warrants the specific emotional reactions we have, and indeed coheres perfectly with them.

After all, taking in the latest alien-invasion film with only mild anxiety would be *insane* if we thought the film depicted actual events. Being amused during *The Big Lebowski* as “nihilists” urinate on The Dude’s rug and dunk his head in the toilet would be cruel if we didn’t take *The Big Lebowski* for fiction. Our disbelief in the reality of what we are reading or witnessing is an essential part of what renders the resulting emotions appropriate—not something that stands in the way of understanding them.

emotional reactions to fictions—even to glossy, megaplex, surround-sound movies, whose *modus operandi* is to shock the senses. Films that trade in razzle-dazzle can be as *boring* as any others. Whether we leap out of our seats when a character bursts out of a closet has little to do with the sound of the closet door swinging open or the shape of the knife being raised, and everything to do with whether the preceding story has engaged us. Reflexes don’t care about that history, but people do.

Second, it is precisely because of this disbelief that there is no *mistake* underlying our emotional responses to fictions. When someone tells us what we already knew—that the fiction is indeed a mere fiction—no grounds have been removed for our reaction, no premise defeated. This marks the crucial difference with the rug-pull cases described above, where we do take the reality of the scenario to be a reason for our emotional reactions. In those cases, we have been caught in a mistake; a central reason for our reaction has been removed and the rug pulled out from under our response. By contrast, when we give the unreality of a fictional scenario as one of our reasons for reacting to a fiction as we do, our reasons cannot be undermined by pointing to that very feature.

This is not to say that there is no way of undermining our fiction-directed responses, however. This is a third important point. Indeed, were there no such way, this would itself suggest that the responses were not founded on or caused by any relevant beliefs—that they were reflexes of a kind. If, instead, we can find beliefs that, if revised, would render our responses to a fiction unwarranted (were they to continue), we have reason not to view those responses as automatic reflexes. To that end, suppose that, unwittingly, we are watching live news footage of an attack by extra-terrestrials, wrongly taking it to be a mere “mockumentary” in the style of *War of the Worlds*. We are experiencing mild anxiety, mixed with a bit of amusement, as we might be when watching an ordinary alien invasion film. If someone bursts into the room with the news that what we’re watching is no fiction at all, our mild anxiety and amusement will immediately cease or—should it somehow continue!—become entirely unwarranted. We have an exact parallel to the ordinary rug-pull examples, with the only difference being that the belief overturned was in the fictionality of the events being witnessed. We believed that, in a certain fiction, Earth was being invaded. This belief provided (partial) warrant for the emotions we were experiencing. When the belief is removed—because the footage turns out *not* to be a fiction—our emotions are no longer warranted. Other, less pleasant emotions become warranted in their stead. The fact that warrant can potentially be removed in this way, however, shows that there *are* beliefs relevant to the response’s appropriateness after all. Those who reject (2) wrongly suggest, to the contrary, that our responses “just happen” as a matter of cognitive wiring or pre-conscious, *a*-rational mechanisms—that we lack any beliefs and desires that could warrant them.

Our belief in the fictionality of what we are reading or watching is critical to explaining our affective responses for a second reason: it places us in an epistemic context where the factors relevant to explaining and justifying our emotions are different than when we are not engaging with a fiction. It opens the door to our having responses that are *aesthetically warranted*, insofar as such responses result from “the ability to recognize and respond to a work’s aesthetically relevant features” (Livingston & Mele, 1997, p. 162; see also Gilmore, 2011). Developing and amending an earlier proposal from Currie (1990), Livingston & Mele (1997) spell out in detail a

variety of norms relevant to assessing the appropriateness of a fiction-elicited emotion. More recently, Gilmore (2011) continues that project in an examination of the specific norms that apply to emotions generated in the context of fiction-appreciation, as contrasted to those that arise in everyday life. It will be useful to describe some of these norms here, as their appreciation helps to highlight the fact that our emotional engagement with fiction is typically grounded in reasons, and not reflexive, sub-personal mechanisms. (Note, however, that neither Livingston & Mele nor Gilmore offer their norms as part of a response to the paradox of fiction.)<sup>12</sup>

Livingston & Mele first articulate a norm of *emotional congruence* that is closely linked to truth in fiction:

Some emotion, *e*, is a congruent response to some feature of a work just in case either (1) *e* is warranted by the work's fictional truths, or (2) *e* is intended by the author to be the appropriate response of the target audience. (1997, p. 171)

Livingston & Mele defend an intentionalist account of truth-in-fiction, allowing that what is true in a fiction depends crucially on the intentions of the author that thus and such be “imagined and accepted for the purposes of the fiction” (p. 170). Whether this is the correct account of truth-in-a-fiction needn't concern us here. The point is simply that our emotional responses to a fiction can be warranted, they argue, if the relationship between them and what is true in the fiction is “congruent with” the relationship between ordinary emotions and the things that elicit them outside of fictions:

If anger is the appropriate response to a certain kind of unjustified aggression in reality, anger (though not necessarily of the same intensity) is also the congruent response to such events in fiction. (p. 171)

<sup>12</sup> Despite articulating a variety of distinctive norms relevant to assessing emotional responses to fictions, Gilmore does not appeal to such norms in resolving the paradox of fiction. Instead, he follows others, including Gendler & Kovakovich (2006) and Lamarque (1981), in advocating the “widely-defended” view that “our emotional response to what we take to be fictional instantiate a broadly exhibited disposition to respond behaviorally, cognitively, and affectively toward some kinds of stimuli in a way that is indifferent to their sources.” Such responses are justified “as a class” by their general ability to “promote (or at least not conflict with) an individual's wellbeing or interests” (2011, p. 471). What, then, does he aim to accomplish in noting the distinctive norms relevant to fiction-appreciation? Gilmore explains: “even if the categorical rationality of such emotions is thereby secured [i.e., by ‘source indifference’], we still need to ask under what conditions any particular emotional response fits its fictional or imagined object” (p. 471). Here appreciation of the relevant norms may help to answer those questions. However, if each *specific* emotional response is justified (or not) by appeal to a fiction-relevant norm, what need is there for some broader set of justifying criteria to provide warrant for the responses as a class? Moreover—as noted above—whether there are indeed “source indifferent” mental states, or not, and whether our general tendency to respond emotionally to what we imagine promotes our wellbeing, or not, are empirical matters. It could turn out that there are no such source-indifferent mental states, and that our imaginings do not increase our wellbeing. Even in that case, on Gilmore's own view, our *specific* emotional responses would nevertheless remain justified. Why, then, are they not the right place to look when asking for a solution to the paradox of fiction?

Their parenthetical remark—“though not necessarily of the same intensity”—is more important than they let on. The claim that they are making—or, in any case, that *should* be made—is that if anger is the appropriate response to a certain kind of unjustified aggression in reality, then a lesser degree of anger is normally the appropriate response to such events in a fiction. If it is appropriate to become extremely worried upon learning of an invasion by space aliens, then, according to the norm of congruence, it is appropriate to have a milder version of that emotion while attending to such events in a fiction.

This is not all there is to congruence, however. An emotion can also be congruent simply on the grounds of its being “intended by the author to be the appropriate response.”<sup>13</sup> This is another norm distinctive to fictional engagement. When the Killer Rabbit mutilates individuals in Monty Python’s *The Holy Grail*, laughter and amusement are appropriate responses—not because greater laughter and amusement would be appropriate if actual rabbits killed people in such a way, but because such a response was intended by the creators of the film. (I take the example from Friend (2016).) This allows for responses to “dark humor” to be appropriate—for us to appropriately laugh, or be amused by events in a fiction that we would never deem humorous if they really occurred.

A further norm of *artistic merit* also warrants mention, as it can trump considerations of congruence. Following Currie, Livingston & Mele note that “it is an aesthetic error to weep over the goings-on in sentimental trash, just as it is a mistake to howl in delight at the supposed ironies in what is in fact a botched melodrama” (1997, p. 173). Granting the inevitable difficulties in drawing bright lines in this territory—just *how bad* does a work have to be before it is inappropriate to be moved by it?—consideration of extreme cases makes clear the existence of such norms. We can easily imagine a terribly written and poorly acted tragedy—something concocted by a group of fifth graders over the course of a few hours. We ought not to be moved by it, even if it contains fictional truths to which a congruent response would be considerable sadness. Only if “the aspect of a work to which one responds is of at least moderate quality,” Livingston & Mele conclude, is one “justified in responding congruently” (p. 173). When low degrees of artistic merit make congruent emotional responses inappropriate, they add, a *contrary* response may be appropriate instead: “artistic flaws warrant any item from a collection of responses that includes boredom, laughter, and whatever feelings accompany the wincing and groans people manifest in response to fictional howlers” (p. 173). To groan and wince at a hackneyed punchline is a warranted response, just as feelings of outrage and sadness are appropriate in response to a brilliantly written scene where a sympathetic character is brutally murdered.

<sup>13</sup> Cf. Gilmore: “Something in a fiction can be fearsome . . . just because we are *caused* to feel fear toward it, if that is in accordance with the ends with which the work was designed” (2011, p. 483).

In each case, part of what warrants the response is the degree of artistic merit to which one is sensitive.<sup>14</sup>

In a similar vein, Jonathan Gilmore (2011) highlights distinctive ways that artworks elicit emotions “without supplying what would count as reasons for the emotions when they are held toward objects in real life” (p. 418). For instance:

Fear toward events in the cinema or theater can be provoked by a menacing soundtrack, while a tendency to delight is triggered through cheerful music. We are often induced to see characters as morally contemptible or threatening by their being described or shown as physically ugly or deformed...or through their being accompanied by a character-specific leitmotif.

(Gilmore, 2011, p. 481)

Gilmore goes on to detail a range of other aesthetic features (e.g., the thickness of a brushstroke, or the fragility of a sculptural material) that become relevant to explaining our emotional responses when we take an “external stance” to the work, “a stance that identifies elements of a fiction in terms of such artefactual aspects as character, plot, style, medium, meter tone...but not as the content that is represented” (p. 482). The use of such features to elicit emotions is possible, Gilmore proposes, due to the different *ends* to which fictions answer:

Many fiction- and imagination-directed emotions...are generated in activities that are defined by ends—such as pleasure, entertainment, and absorption—that don’t require that the emotions always be rationalized by the facts of the objects to which they respond. One may feel warmly disposed toward a figure in a painting only because the beauty of the work is designed to cause one to feel that way. (pp. 484–5)

The moral Gilmore draws—correctly, in my view—is that

There are some kinds of reasons that justify an emotion felt toward a state of affairs represented in a fiction or imagining that would not justify that emotion when felt toward an analogous state of affairs in the real world. (p. 485)

The existence of emotional norms proprietary to fiction alerts us to the fact that we have at hand many different kinds of reasons we can appeal to when asked to justify our emotional reactions to fictions. We can talk about the events of the

<sup>14</sup> In addition to artistic merit, Livingston & Mele propose that *moral* norms are relevant and can trump considerations of congruence. If the content of a fiction is morally offensive—if, for instance, the author clearly intends the audience to share or adopt an offensive moral view—a contrary emotional response may be appropriate.

fiction itself as justification for why we have emotions that are milder versions of what we would experience in relation to such events were they actual. We can adduce the director's intent that the killer bunnies be seen as comedic. We can note the fiction's evocative language, its ingenious plot, the brilliance of the acting. We can observe that the characters are likeable and relatable, their predicament both familiar and compelling. We can appeal to the life-lessons the fiction teaches us by analogy and the light it sheds on people and events in our own lives. In short, we can give as *reasons for caring* the very things that film and literature critics highlight when they write a glowing review. Much of art criticism is simply this: the skillful rationalization of our emotional responses to art. When we point to such features of an artwork, we are offering evidence that our emotional reactions are apt and fitting of their objects, and therefore warranted.

### 11.8 But None of These Things Are Happening! Summary via Objection

One might fear that this response to the paradox is circular. We wanted to know how it could be reasonable (or warranted) to react emotionally to fictional characters and situations we know not to exist. And (to put it crudely) we are told that fictions have fiction-specific features that render our responses warranted. It may seem reasonable to stomp one's foot and repeat: but why are any of these emotional responses at all reasonable, given that we know the events are not occurring?

In response: they are reasonable because we can give good reasons for them—reasons that do not assume the events described by the fiction to have actually occurred. One reason we give for responding as we do is that we are engaging with a *fiction* and not, say, a documentary; another may be that the fiction is well crafted; a third may be that, in the fiction, tragic events occur to which our response is congruent. The kinds of reasons just given are the sort that people would ordinarily accept in explanation for why we responded as we did to a fiction. (See, e.g., Cova & Teroni (2016, pp. 933–4) for a survey confirming this claim.) If someone wishes to call the legitimacy of those reasons into question, they need to catch us in some kind of inconsistency, or reveal some other undermining factor.

Of course, that is precisely what Radford attempts to do with his rug-pull examples. Ordinarily, when we have an emotional response to some event, that response is undermined if we discover the event not to have occurred. Why doesn't this happen in the case of our engagement with fictions? Where is the important difference?

It is here: in the rug-pull cases Radford describes, the warrant one initially has for an emotional response is undermined when one discovers that what one *took to be* its object (e.g., a sick relative) does not exist. In the case of our response to

fictions, we make no such undermining discovery. *All along*, we take the object of our emotions to be a certain fiction, together with its characters and events. Asked why we are responding as we do, we note that it is a fiction, that it has thus and such aesthetic features, that its story details events of thus and such kind (that render our response congruent), and so on. It is only if *those* reasons are undermined that the rug is pulled out from under our response. For instance, if we discover that the fiction itself does not exist—if, for example, the narrative turns out to be a *non-fiction*, or if it turns out we are unwittingly viewing an actual argument between actors—then our emotional reactions are indeed undermined and, potentially, become both unwarranted and unfitting. But we need not discover something so surprising as that there is no such fiction as we took there to be; it is often enough to discover that the fiction does not have a specific feature we adduced in our explanation for the emotion. If, for instance, our shock is revealed to have been based on our mishearing a bit of dialogue, then we will judge it to have been an unfitting response (but perhaps still a *warranted* one, due to what we *thought* we heard).

Far from being an exception to the rule, our emotional engagement with fiction is one more instance where specific features of some existing thing (*viz.*, a fiction) elicit emotions, where specific features of the stimulus make certain emotions fitting or apt in response, and where our warrant for the emotions is undermined if we come to believe the object of the emotion lacks features that render the emotion an apt response. Provided that the fiction has the features we adduce, and provided that, in accordance with the specific norms governing fiction-consumption, those features are apt to elicit the kinds of emotions we experienced, the promulgator of the paradox has no credible means for undermining our reasons.

These points allow us to deepen the earlier criticism made of views that reject premise (2). For *even if* Damasio's somatic marker thesis is true (as adduced by Gendler & Kovakovich (2006)), and even if mental imagery shares neural processing pathways with perceptual experience (Van Leeuwen, 2016), and even if it is "adaptive" to respond emotionally to mere imaginings, our emotional reactions to fictions can *still* be unwarranted and unfitting. They will be unfitting when the fictions that evoke them lack features that render them apt; and they will be unwarranted when we do not believe the fictions that evoke them to have features that render them apt. The fact that our emotional responses to fictions *can* legitimately be undermined in these ways indicates that the solutions to the paradox that reject premise (2) prove too much, by excusing too much.

Gendler & Kovakovich and Van Leeuwen both show peripheral awareness of this problem. In the last paragraph of their paper, Gendler & Kovakovich note that, despite the cognitive architecture they posit, "there are plenty of circumstances where such emotional responses [to fictions] are very much out of place." This is because "our assessments of rationality and irrationality are, here [in engagements with fiction] as elsewhere, governed by conventional norms of appropriateness."

This is an odd last-minute concession. For understanding “conventional norms of appropriateness” and how they apply differently in our engagement with fictions than in other contexts would appear to be the entire question at issue when addressing the normative version of the paradox (which they take as their target). If the somatic marker hypothesis is not relevant to understanding these norms, how then is it relevant?

Van Leeuwen, on the other hand, qualifies his remarks by holding that the imagery that triggers emotional responses will only be generated if a fiction is “well crafted...in the relevant sense of ‘well’” (2016, p. 103). Thus, a poorly crafted fiction, he can respond, will not elicit imagery and its attending emotions. This leaves us with the question of *why* we only generate imagery when enjoying a *well-crafted* fiction. Presumably the answer is that we only take an interest in such fictions; they are the only fictions that engage us. (It is not part of his account that we need imagery simply to recover fictional content—otherwise we would form imagery and experience related emotions when consuming poorly-crafted fictions as well.) But the whole question behind the normative puzzle is why it is reasonable to *care* about events we know never happened. Once our interest in a fiction (well-crafted or not) is made reasonable, our emotional reactions will come along as reasonable for free—whether or not anything Van Leeuwen says about the connection between imagery and emotion is true.

Finally, a last objection in this vicinity was put to me by Reader Y. Reader Y suggests that my discussion of fictional norms may leave open the main question at issue. To quote from Y’s report:

It’s important to see that none of the material about whatever our warrant may be for our emotions can do any of the work that [Langland-Hassan] needs here. At best it gets him to, ‘If you feel despairingly sad while engaged with *Romeo and Juliet*, here’s why that’s normatively appropriate.’ But we lack from Langland-Hassan any sort of story about what should cause anyone who only has metafictional beliefs and desires to feel despairingly sad in the first place.

In referencing “metafictional beliefs and desires,” Reader Y has in mind my appeal to beliefs and desires with ‘in the fiction’ operators—and specifically, my claim that these pair to generate fiction-directed affect. Why, asks Reader Y, should such states ever pair to generate sadness?

Well, why shouldn’t they? I take it that Reader Y is normally satisfied when we explain someone’s sadness by appeal to their strong desire that not-*p* and recent judgment that *p*. Why should it matter if the proposition *p* contains an ‘in the fiction’ operator? The form of explanation is the same. The beliefs and desires are the relevant causes.

What’s really lurking behind Reader Y’s objection, I think, is a Radfordian suspicion—still not exorcized—that there is something *inappropriate* or *unreasonable*



about reacting emotionally to what one takes to be a mere fiction. If that were the case, then propositions that include ‘in the fiction’ operators would be exceptions to the general rule where a desire that not- $p$  pairs with a judgment that  $p$  to generate negative affect. But why should we think propositions including ‘in the fiction’ operators are exceptions to the general rules governing how belief and desires cause affect? Reader Y does not say why. It is offered as obvious that someone who “only” has metafictional beliefs and desires won’t experience significant affect. My hunch is that this apparent obviousness is grounded in Radfordian reasoning of the kind we’ve already dispelled in considering rug-pull cases. Appreciating the complex norms that govern our responses to fiction helps us to see what it would *really* take to pull the rug out from under those responses and render them unreasonable—and indeed unaccountable. When we see that no such rug is pulled out in the normal case of enjoying a fiction, we see that “metafictional” beliefs and desires generate no exception to the rule where a desire that not- $p$  and a judgment that  $p$  will combine to generate negative affect.

### 11.9 Back to the Triad

My response to the paradox of fiction is now complete. But I haven’t yet explained what to make of the inconsistent triad. That omission has been purposeful. There are subtly different ways to formulate the triad (and no *standard* way); I don’t want my solution to appear hostage to one particular formulation. But it’s nevertheless instructive to see how the points already made apply to the formulation we considered earlier. Here, again, is Currie’s version of the triad:

- (1) We have emotions concerning the situations of fictional characters.
- (2) To have an emotion concerning someone’s situation we must believe the propositions that describe that situation.
- (3) We do not believe the propositions that describe the situations of fictional characters. (Currie, 1990, p. 187)

As earlier noted, the most common response is to reject (2), with the idea being that imagining that  $p$  generates emotional responses even when we do not believe that  $p$ . Now, I don’t reject the platitude that imagining that  $p$  can, at least at times, generate emotional responses of roughly the sort that would be appropriate if one believed that  $p$ . But, in line with my broader project of explaining imagination, I give an account of what it is to imagine that  $p$  (in the context of appreciating a fiction) that appeals only to more basic states, such as beliefs and desires about what is true in the fiction. With that in mind, let’s look more closely at condition

(3). It says that we do not believe the propositions that describe the situations of fictional characters. Just what *are* the propositions that describe the situations of fictional characters?

Consider an example: viewers of *The Big Lebowski* know that The Dude has strong musical preferences and, in particular, hates the Eagles. What proposition describes The Dude's situation? It is not: 'The Dude hates the Eagles.' That proposition is false. There is no actual Dude who, by hating the Eagles, could make it true. 'The Dude hates the Eagles,' is no truer than 'The Dude *loves* the Eagles.' Therefore, it cannot truly describe The Dude's situation. We will need a true proposition in order to give a true account of his or any situation. The relevant *true* proposition—the one that accurately describes the situation of this fictional character—is the proposition that, in *The Big Lebowski*, The Dude hates the Eagles. And, of course, if we have seen the movie, we *do* believe that proposition. The same points apply to other fictions we wittingly enjoy. So, contrary to proposition (3) of the triad, we *do* believe the propositions that describe the situations of fictional characters; for those propositions all involve 'in the fiction' operators. Rejecting (3), we can continue to accept (1) and (2) without conflict.

This way of defusing the triad may seem too cheap. It might seem to succeed (if it does) due to an infelicity in the way Currie formulates the triad. But, in fact, the reasons just given for rejecting (3) do not do any of the hard work in resolving the paradox. For even if one grants that the propositions that describe a fictional character's situation involve 'in the fiction' operators, this simply pushes the puzzle onto (2). For (2), in its normative reading, offers itself as a (partial) characterization of the appropriateness conditions for the emotions we have about "someone's situation." Namely, we must believe "the propositions that describe that situation," in order for the emotion to be warranted. In light of our new gloss on (3), accepting (2) requires us to accept that, when we have an emotion concerning a *fictional* character's situation, its warrant depends upon our believing that, in the fiction, thus and such. For propositions involving 'in the fiction' operators are the kinds of propositions that describe a fictional character's situation. But how could believing that a situation occurs *in a fiction* warrant an emotion like fear or anger about the situation? Now we are back to the main question that began this chapter. We can understand why the emotions are warranted if we can understand why it is warranted to have strong desires to the effect that, in the fiction, thus and such—desires which pair with beliefs that, in the fiction, thus and such to generate the emotions.

And why *shouldn't* we have such beliefs and desires? The traditional complaint against them is that it makes no sense to have strong desires about what happens in a *mere* fiction. But now we have seen that the plausibility of that complaint derives from a faulty analogy between Radfordian rug-pull cases and our engagement with fictions. As I argued last section, our emotional responses to fictions, and the desires that elicit them, are founded on reasons—reasons (and not mere

reflexes) that can indeed be undermined if certain of our beliefs turn out to be false. Happily, our reasons are *not* undermined by the observation that we are responding to a fiction. The opposite is true. The responses are warranted, in part, by the belief that what we are responding to is a fiction. Thus, no reason is undermined—no rug pulled out from under us—when we are told that the characters we took for fictional are indeed fictional. It was in making this point, and in articulating the alternative norms that govern our engagement with fiction, where the heavy lifting in my account took place. Resolving the triad with those points in hand is then straightforward.

### 11.10 Summary

The last three chapters have considered the role of imagination in the consumption of fiction from a variety of angles. I have taken it as a platitude that we engage in A-imagining—rich, elaborated, epistemically safe thought about the possible, fantastical, unreal, and so on—in response to fictions. The key question has been whether this A-imagining can be characterized in more basic folk psychological terms. I have argued that it can. We began, in Chapter 9, with the question of whether recovering the implicit and explicit content of a fiction requires *sui generis* imagining. There my answer was no. Recovering fictional content may at times require counterfactual reasoning. Yet, as I argued earlier in Chapters 5 and 6, the A-imagining that takes place during conditional reasoning can itself be reduced to more basic folk psychological states. Further, we can agree with Stock that recovering fictional content is not *simply* a matter of reasoning counterfactually about what would be true were some other set of (fictional) propositions true. The other relevant bits of reasoning—concerning, e.g., what is true in the fiction given certain genre conventions, or given likely uses of symbolism by the author—can easily be seen as resulting from abductive inferences from one's existing beliefs about the particular fiction and the practice of fiction-making generally.

Chapters 10 and 11 moved on to consider whether it is the phenomenon of immersion within fiction that more definitively calls for *sui generis* imaginings. I began Chapter 10 with the Operator Claim, arguing that our *sui generis* imaginings (were there such) would require 'in the fiction' operators in order to be directed at the fictions we take them to concern. The truth of the Operator Claim undermines the main rationale for thinking that *sui generis* imaginings facilitate immersion. If the OC is correct, *sui generis* imaginings can put us in no more direct cognitive acquaintance with fictional events than beliefs with 'in the fiction' operators; further, the way in which the events of one fiction are "cognitively quarantined" from others "in imagination" will be the same way in which our beliefs about one fiction are quarantined from beliefs about another—viz., by the use of relevant 'in the fiction' operators.

To really understand immersion, I argued, we need to grasp why it is that we form strong desires concerning fictional events. Contrary to what many have supposed, appealing to imagination and “source indifferent” affective mechanisms does not help us here. Instead we need to confront and properly resolve the paradox of fiction. That was my primary aim in this chapter. The key is to understand why we are warranted in *caring* about fictional events. This requires, first, that we see clearly why there is a presumption against caring about fictional events in the first place. The answer lies in “rug-pull” cases of the kind Radford introduced into the literature. On reflection, we can see that the rug-pull structure does not apply to our engagement with fiction. In order to appreciate that fact, it helps to highlight the rational norms specific to fiction-appreciation. This clarifies why events that pull the rug out from under our emotional reactions in everyday life do not do so in our engagement with fiction. We become immersed in fictions *because* they are fictions, not in spite of their being fictions. Our emotional responses are not “indifferent” to whether something is a fiction; they are delicately calibrated in ways only appropriate if the stimulus is a fiction. It’s not that our responses are only *truly* fitting if the fiction turns out to record actual events. Quite the opposite: only a discovery that what we took for fiction is non-fiction would reveal our responses to be unfitting.