Is the Paradox of Fiction Soluble in Psychology?

(penultimate draft, to appear in *Philosophical Psychology*)

Abstract

If feeling a genuine emotion requires believing that its object actually exists, and if this is a belief we are unlikely to have about fictional entities, then how could we feel genuine emotions towards these entities? This question lies at the core of the paradox of fiction. Since its original formulation, this paradox has generated a substantial literature. Until recently, the dominant strategy had consisted in trying to solve it. Yet, it is more and more frequent for scholars to try to dismiss it using data and theories coming from psychology. In opposition to this trend, the present paper argues that the paradox of fiction cannot be dissolved in the ways recommended by the recent literature. We start by showing how contemporary attempts at dissolving the paradox assume that it emerges from theoretical commitments regarding the nature of emotions. Next, we argue that the paradox of fiction rather emerges from everyday observations, the validity of which is independent from any such commitment. This is why we then go on to claim that a mere appeal to psychology in order to discredit these theoretical commitments cannot dissolve the paradox. We bring our discussion to a close on a more positive note, by exploring how the paradox could in fact be solved by an adequate theory of the emotions.

1. Dissolving the Paradox of Fiction

Since its original formulation by Colin Radford (1975), the paradox of fiction has generated a substantial literature. It arises when one attempts to combine the following three propositions:

> (PF1) We can feel genuine emotions for fictional characters. (To take a classical example, we can feel genuine sadness for Anna Karenina when reading about her tragic fate.)
>
> (PF2) We do not believe that fictional characters exist. (We do not believe Anna Karenina to really exist in the actual world.)
>
> (PF3) To feel genuine emotions, we must believe that these emotions are directed to actually existing objects.[1] (Feeling genuinely sad for a person requires believing that this person actually exists.)

It is easy to realize that these propositions seem to yield a paradox. If feeling a genuine emotion requires believing that its object actually exists, and if this is a belief we are unlikely to have about fictional entities, then how could we feel genuine emotions towards these entities?

So far, the prevalent attitude has been to try to *solve* the paradox of fiction by determining which proposition amongst (PF1)-(PF3) is false. This may look like the only reasonable reaction. Yet, this attitude is of course premised on the idea that there is a paradox

[1] More precisely, (PF3) should be rephrased in the following way: "To feel genuine emotions, we must believe that these emotions are directed to objects which actually exist, existed or have a chance of coming into existence." Indeed, we feel emotions for dead people (as when one pities Caesar) and for people who do not exist yet (one is, say, afraid of the fate of future generations). Given that these further issues are not relevant for our discussion, we will stick to the less rigorous formulation.

to begin with. And a paradox is certainly more than a set of incompatible propositions. This is why one does not generate a paradox by laying down any triad of propositions such as (i) the moon is a huge watermelon, (ii) the moon is white, and (iii) watermelons are green. The incompatible propositions at stake must in addition be *cognitively attractive*. Thus, trying to solve the paradox presupposes that one admits that (PF1)-(PF3) are all cognitively attractive and that their conjunction yields a contradiction. A solution has thus to meet two requirements: it should determine which premise is to be rejected and explain its initial plausibility.

Let us illustrate this point. One popular solution has been to reject (PF1) in arguing that we do not feel genuine emotions for fictional characters (e.g. Walton 1978; Currie, 1990). What happens instead is that interacting with works of fiction produces psychological states that fail to qualify as genuine or full-fledged emotions even though they have similar or perhaps even identical phenomenal aspects (Walton, 1978). This is why some philosophers have christened these psychological states "quasi-emotions". It is next argued that, while (PF3) is true of emotions, it is false of quasi-emotions. In the latter case, believing that *p* should be replaced by make-believing that *p*, this explaining why quasi-emotions are so typical of our interactions with works of fiction (Currie, 1990). This solution constitutes a genuine attempt at solving the paradox of fiction: it recognizes that the three propositions (PF1)-(PF3) are incompatible, rejects (PF1) and explains why it seems attractive. Quasi-emotions being phenomenologically indistinguishable from genuine emotions, we are unsurprisingly prone to judge that we feel genuine emotions for fictional characters.

However, it is more and more frequent for scholars to *dismiss* the paradox of fiction rather than to try to *solve* it. The claim is that there is no real paradox to begin with, since (PF3) is obviously false, or at least unmotivated. The pervasiveness of this attitude is made

manifest in Robert Stecker's (2011) recent survey of the literature. Stecker wonders whether we should still care about the paradox of fiction:

> Here is the problem. The paradox was formulated during the heyday of the cognitive theory of the emotions when there was a lot of theoretical commitment to [(PF3)] or a variant of it. But now virtually no one accepts [(PF3)]. (Stecker, 2011: 295)

On the next page, he adds that

> To solve the paradox, we just have to find a way to reject one of the inconsistent statements. It appears easy to reject [(PF3)]. So why do even some of those who do reject [(PF3)] not leave matters there? Is there any reason not simply to reject [(PF3)], pronounce that the paradox solved, and stop caring about it? Many now think this is precisely the way to go. (Stecker, 2011: 296)

Ultimately, the author grants the paradox a certain heuristic value: attempts at solving it have yielded valuable advances in our thinking about emotional reactions to fiction. Yet, he concludes that we might no longer need the paradox to make progress on these matters; it may be time to "kick away the ladder, and continue to explore these responses in their own right." Looking at the way Stecker construes the dialectical situation, it looks as if we were once confronted with a paradox of fiction, which was generated by the then dominant cognitive theories of emotions and their commitment to (PF3). Since these theories have in the meantime fallen into disgrace, no reason remains to endorse (PF3) and the paradox of fiction has vanished.

We find a similar argument and conclusion under the pen of Derek Matravers (2014), who goes even further than Stecker in claiming that "paradox of fiction" may be a misnomer:

> There has always been an air of artificiality about the problem (…) The phrase 'the paradox of fiction' (…) is an unfortunate misnomer, as the problem is neither a paradox nor is to do with fiction. (Matravers, 2014: 102).

Here again, the idea is that the paradox is traceable to an inadequate cognitive theory of the emotions, which commits one to (PF3). However, given his contention that cognitive theories of emotions do not imply (PF3), Matravers' take on the issue turns out to be more sweeping than Stecker's. He observes that

> a cognitive theory (…) does not require belief: it only requires some sort of pro-attitude or other (which could be a belief, an evaluation, a positive appraisal, or even an imagined state). (…) If we cannot find support for [(PF3)] in philosophers writing at a time in which the cognitive theory was dominant, still less can we find it in philosophers since. At least since Patricia Greenspan's *Emotions and Reasons* philosophers writing on emotions regularly disavow the claim that beliefs are either causally or conceptually necessary for the emotions. As usually stated, the paradox of fiction is as made from straw as a man can be. (Matravers, 2014: 106)

So, according to Matravers, we never had any reason to endorse (PF3).

Katherine Tullmann and Wesley Buckwalter (2014) offer the most recent version of this sceptical attitude towards the paradox of fiction. These authors actually go further than Stecker and Matravers in claiming that (PF1)-(PF3) are compatible. According to them, these propositions only *seem* to be incompatible because we fail to appreciate that "exist" has one meaning in (PF2) and another one in (PF3). They observe that "exist" may have, in (PF2), one of three distinct meanings. It may mean either (i) to be a concrete entity, or (ii) to possibly be a concrete entity, or (iii) not to be an imaginary (or a merely imaginary) object. Consequently, to say that fictional objects are not believed to exist may mean either (i) that they are not

believed to be concrete objects, or (ii) that they are not believed to possibly be concrete objects, or (iii) that they are not believed to be more than merely imaginary objects.

Having identified these three potential meanings of "exist" in (PF2), Tullmann and Buckwalter turn their attention to (PF3). They want to assess whether genuine emotions demand that the subject believe the relevant objects to exist in any of these three senses. Given that (PF3) refers to genuine emotions, they undertake to review what they perceive as the leading theories of emotions—which they group in three categories respectively labelled "feeling theories", "judgement theories" and "pure-cognitive belief theories"—in order to determine whether any one of them actually entails such a claim. They come to a negative conclusion: none of these theories supports the idea that genuine emotions demand existential belief understood in any of the senses of "exist" identified in relation to (PF2).[2] Tullmann and Buckwalter conclude that there is no reason to endorse an interpretation of (PF3) that conflicts with (PF2). This in turn means that we can reconcile (PF1), (PF2) and (PF3) and that the impression of there being a paradox of fiction should disappear in the light of the foregoing considerations. The authors conclude that

> if the paradox of fiction does exist, it remains to be discovered or argued for—not solved. (Tullman and Buckwalter, 2014: 793)

Although they differ in the details they include, the three takes on the paradox of fiction that we have reviewed reveal the same attitude. Stecker, Matravers, Tullman and Buckwalter are not in the business of trying to *solve* the paradox, they rather want to *dissolve*

[2] This is not to say that no conceivable theory of the emotions could confer to "exist" as it appears in (PF3) one of the meanings that it potentially has in (PF2), and would in this way really generate a paradox. Tullmann and Buckwalter only contend that we have no obvious reason to endorse such a theory.

it. They do so either by claiming that there are no reasons to endorse (PF3), or that the reasons there are support a modified version of (PF3) that yields no contradiction. Either way, the claim is that the paradox vanishes because (PF3) is only supported by outdated and unconvincing emotion theories.

Can the paradox of fiction be dissolved in this way? We will argue that it cannot. In sections 2 and 3, we explain that the paradox of fiction primarily emerges from reasons to endorse (PF3) that have nothing to do with theories of emotions. Moreover, we contend that the philosophical arguments that support (PF3) are grounded in contemporary research on emotion regulation and affective reactions to fiction. Next, in section 4, we argue that the fact that (PF3) receives this kind of support reflects badly on contemporary attempts to dissolve the paradox. Section 5 constitutes the positive part of our discussion; we explain in what sense we agree with critics of the paradox of fiction that this paradox will be solved by an adequate theory of emotions. We will conclude with some brief comments on a normative version of the paradox.

2. Is (PF3) Intuitively Attractive?

As we have just seen, attempts at *dissolving* the paradox of fiction rest upon the claim that contemporary emotion theories do not provide any reason to endorse (PF3). One may want to confront that claim head-on by arguing that we indeed have theoretical reasons to endorse (PF3).

This strategy is difficult to implement, however, since only two theoretical explanations of (PF3)'s attractiveness suggest themselves. Those who have endorsed (PF3), one may insist, did so because they were convinced that it follows from one of the families of emotion theories Tullman and Buckwalter describe as standard. What should give us pause is that it is

reasonably clear that this conviction is mistaken.[3] Alternatively, one may explain (PF3)'s attractiveness as resulting from the adoption of a non-standard approach to the emotions. Yet, given the three families of theories described by Tullman and Buckwalter, one will be hard-pressed to name one. So, no attractive theoretical reason for explaining the prevalent outlook that there is a paradox of fiction seems to be forthcoming. We should try to discover other, non-theoretical reasons for endorsing (PF3).

To that effect, Tullmann and Buckwalter defer to the fact that (PF3) is intuitively attractive, observing for instance that Radford's initial formulation of the paradox of fiction "is comprised of a triad of inconsistent, yet highly intuitive propositions" (2014: 780). This is in line with other suggestions scattered throughout the literature. For example, Jerrold Levinson describes the paradox of fiction as "a set of three propositions, to each of which we seem to have strong allegiance" (1997: 22). The alleged intuitive character of (PF3) should be called into question, however. In our personal experience, asking anyone who is unacquainted with the paradox whether feeling a genuine emotion demands that one believe its object to exist typically triggers amazement at learning that some are even tempted to accept this claim.

These informal observations match more systematically collected data. In an attempt at probing laypeople's intuitions about (PF3), we ran an online survey on this particular topic. Participants were 200 men and women subscribed as workers on Amazon Mechanical Turk and located in the United States (M_{age}=32.71, SD_{age}=9.88; 39.0% female and 61.0% male).

[3] This is obvious as regards feeling theories and quite transparent in the case of approaches to the emotions in terms of evaluative judgements. In addition, we find Tullmann and Buckwalter's observations regarding what they describe as pure-cognitive belief theories straightforward and convincing. Acceptance of (PF3) has never been part of any such approach.

They were recruited online and paid $0.2 for their participation in the survey. Participants indicated their gender, age and residence country, and then read the following introduction:

> There is a disagreement among philosophers about whether we can feel genuine emotions for persons who do not exist. Some philosophers think that we can feel genuine emotions (such as compassion or fear) for persons who do not exist, even if we know and are fully aware that they do not exist. Others think that it is impossible to feel genuine emotions for persons we know not to exist, though we might eventually experience something that resembles genuine emotions when learning about their fate.

Participants in the *Positive Formulation* condition then received the following statement:

> Thus, these philosophers disagree about the truth of the following statement: "For one person to feel genuine emotions for another person, the first person must believe that the second person really exists."

As for the participants in the *Negative Formulation* condition, they received the following statement:

> Thus, these philosophers disagree about the truth of the following statement: "One cannot be moved by the fate of another person, if one does not believe that this person really exists."

Finally, participants had to answer the two following questions:

1. According to you, is this statement TRUE or FALSE?
    - TRUE
    - FALSE

Please, justify in one or two sentences your answer to the previous question.

2. How intuitive do you find the idea expressed by this statement? *(on a scale from 0 to 6, 0 being "Not intuitive at all" and 6 being "Very intuitive")*

Answers to both questions across the two conditions are presented in Table 1. Overall, answers to the intuitiveness questions were significantly above the midpoint (N=200, t=6.34, df=199, p<.001), meaning that participants found (PF3) somewhat on the intuitive side. Nevertheless, in spite of its intuitiveness, most participants ultimately rejected (PF3). Unsurprisingly, rejection was most of the time justified by reference to the emotions we feel for fictional characters. For example, a participant writes "I believe this is false, because people can feel emotions for fictional characters while realizing they don't actually exist. Otherwise people wouldn't care about the characters in books and movies."

	Positive Formulation	*Negative Formulation*
Truth	TRUE: 37%	TRUE: 28%
Intuitiveness	3.94 (1.44)	3.45 (1.62)

Table 1. Participants' endorsement of the proposition and intuitiveness ratings, in function of condition (*Positive/Negative*)

It thus seems that, although participants consider (PF3) as being somewhat intuitive, they are also prone to reject it. This sheds doubt on attempts at defending the existence of a paradox on the ground that (PF3) is intuitive. Indeed, interactions with works of fiction seem to constitute as many counterexamples to (PF3) and it is indeed bound to appear counterintuitive if one takes for granted (PF1), i.e. the claim that we are genuinely moved by works of fiction, as most participants seem to do.

3. (PF3) as an Inference to the Best Explanation

Making (PF3) sufficiently attractive to support the claim that there is a paradox of fiction can consist neither in drawing attention to theoretical reasons in its favour, nor to its intuitive character. One may then legitimately wonder whether this proposition can be made to look

attractive at all. Actually, realizing that it can only requires that we go back to Radford's initial formulation of the paradox (Radford, 1975).

Radford does not regard the claim that genuine emotional responses demand existential beliefs as being obvious. Quite the opposite, in fact, since he devotes a substantial part of his seminal discussion to argue in favour of (PF3), or the slightly different proposition he uses in his own formulation of the paradox. So, what is his argument? It takes the shape of an inference to the best explanation for a range of everyday phenomena. According to Radford, (PF3)'s attractiveness is neither due to its being intuitive nor to its being the consequence of a theory. To see what is at stake here, let us revisit his own example:

> Suppose that you have a drink with a man who proceeds to tell you a harrowing story about his sister and you are harrowed. After enjoying your reaction he then tells you that he doesn't have a sister, that he has invented the story. In his case, unlike the previous one, we might say that the 'heroine' of the account is fictitious. Nonetheless, and again, once you have been told this you can no longer feel harrowed. Indeed it is possible that you may be embarrassed by your reaction precisely because it so clearly indicates that you were taken in—and you may also feel embarrassed for the storyteller that he could behave in such a way. But the possibility of your being harrowed again seems to require that you believe that someone suffered. (Radford, 1975: 68)

Radford offers here a faithful description of a quite common situation. It is indisputable that when we appreciate that the situations to which our emotions are directed do not involve something that did *really* happen to *real* individuals (i.e. concrete objects in the actual world), these emotions typically disappear and their occurrence is seen in retrospect in a quite negative light. On this basis, let us try to lay down the structure of Radford's argument for (PF3).

The argument starts off with armchair observations describing a pattern that is quite conspicuous in our emotional life.

(1) Emotional reactions outside fiction typically recede once we learn that their objects do not exist.

If we are saddened by the alleged sufferings of a neighbour's aunt imprisoned in a far-away country, the feeling comes to a more or less abrupt halt when we learn that the neighbour never had an aunt. If we fear for a friend we believe to be threatened by danger, fear recedes as soon as we learn that he is safe. Moreover,

(2) If emotions do not vanish in such circumstances, they are held to be "unwarranted" or "unfitting".

This is indeed how we would indeed assess sadness about the aunt after having learnt that there is no aunt, i.e. in the same way as one would assess one's recalcitrant fear of a spider after having been informed that it is totally harmless. Why are emotions susceptible to such patterns of changes and assessments? Radford offers as an obvious explanation: to be saddened by the circumstances affecting a person, one must believe her to exist and to suffer. That is,

(3) These phenomena are best explained by (PF3).

Inference to the best explanation then gives us robust reasons to conclude that

(4) (PF3) is true.

This is in apparent tension with what happens when we interact with fiction, however, since

(5) Emotional reactions to fiction are (almost) always elicited despite our believing that their objects do not populate the concrete world we inhabit.

Now, (5) may certainly tempt us into concluding via *modus tollens* that

(6) (PF3) is false.

Yet, if we do so, it seems that we cannot explain the emotional patterns in the cases so nicely emphasized by Radford. We face a sort of dilemma. We seem forced either to leave unexplained the fact that our emotions are frequently sensitive to our beliefs about the actual existence of their objects, or to transform our ostensibly genuine emotional reactions to fictional entities into a perplexing phenomenon.

Thus, Radford does not recommend (PF3) in light of its intuitiveness or of a specific theory of the emotions, but because it constitutes the best explanation of a range of phenomena. In his view, the paradox of fiction is rooted in this tension between, on the one hand, what appears to constitute the best explanation of how emotions behave in many mundane cases and, on the other hand, the introspective certainty that we feel genuine emotions for fictional entities not believed to be actual. This is why he describes the phenomenon he is interested in as a sort of "incoherence" between our emotional reactions to fictional and nonfictional entities—and we shall follow his lead in our use of this term. So, a solution to the paradox demands that this incoherence be eased.

If this correctly identifies the source of the paradox of fiction, then the attempts to dissolve it presented in section 1 are under threat. Against Stecker and Matravers, there is a substantial non-theoretical reason to adopt (PF3). The same considerations also tell against Tullmann and Buckwalter, since they should lead us to adopt an interpretation of "exist" that helps explain the phenomena under discussion. Now, (PF3) contributes to this explanation only if "exists" is read as "is a concrete object" or "is an inhabitant of the actual world". And since we certainly do not believe that fictional characters inhabit our world as concrete individuals we might run into around the corner, "exists" seems to have the same meaning in (PF2) and (PF3) after all. The paradox of fiction is still with us.

That being said, one might feel uneasy about the fact that Radford's reflexions are built on thought experiments and anecdotal observations. Is there really something to explain here? There is, and the phenomenon to which Radford draws attention is commonly acknowledged within psychology. In particular, psychologists interested in emotion regulation have come to distinguish several regulation strategies. One of these strategies, "cognitive reappraisal", amounts to regulating one's emotions by changing one's understanding and evaluation of their objects. The literature suggests that one way to study the effects of cognitive reappraisal is to present the same emotional stimuli (texts, images, videos) to participants while asking them to interpret these stimuli as describing either real or fictional events and people.

For example, in a study investigating the specific neural impact of different regulation strategies, Pascal Vrticka, David Sander and Patrik Vuilleumier (2011) had participants looking at emotional pictures. In the control condition, participants were asked to watch and evaluate the depicted emotional scenarios as if they were real situations to which they were personally exposed. Participants in the cognitive reappraisal condition, on the contrary, were instructed to view the depicted emotional scenes as parts of a movie clip or TV show that displayed fake or artificially set-up situations created to give rise to emotions. Finally, participants in the "expressive suppression" condition received the same instructions as participants in the control condition, but with the important difference that they were instructed not to display any felt emotions that could be noticed (e.g., through breathing frequency, heart rate, skin conductance responses, and facial expression—which were told to the participant to be recorded via electrodes attached to the body and an eye-tracker camera). After each picture, participants were shown a rating display and asked to report the feeling state evoked by the pictures.

Overall, Vrticka and his colleagues found that participants reported significantly lower affective reactions in the "cognitive reappraisal" and "expressive suppression" conditions than in the control condition. More importantly, participants also reported much lower affective reactions in the "cognitive reappraisal" than in the "expressive suppression" condition. So, reinterpreting pictures as excerpts from fictional works seems not only to be a way to reduce affective reactions, it seems to be the most efficient way to do so. Similar results are reported throughout the literature, showing that reinterpreting a stimulus as fictional decreases subjective reports of emotional experiences as well as their physiological manifestations (e.g. Kim & Hamann, 2007; Oliveira et al., 2009; Mocaiber, 2011). Moreover, studies not directly interested in emotion regulation, but rather in our affective reactions to works of fiction have also supported the claim that affective reactions decrease when stimuli are reinterpreted as fictional (Mendelson & Papacharissi, 2010; Sennwald et al., 2015; but see Goldstein, 2009).

Thus, experimental data support Radford's claim that learning about the "unreality" of a given stimulus leads to decrease in emotional experience. Since (PF3) is a natural and straightforward explanation of this phenomenon, we have a reason to endorse this proposition. Of course, the fact that (PF3) flies in the face of our emotional reactions towards fictional entities also counts as a reason to reject it, but this is precisely the tension underlying the paradox of fiction.

4. Can Contemporary Theories Dissolve the Paradox?

You may recall from section 1 that the main argument of those who attempt to dissolve the paradox of fiction is that we have no reason to endorse (PF3). Yet, we brought the previous section to a close concluding that such reasons exist. It may then seem that we no longer have any reason to deny that there is a paradox of fiction. This would be too quick, since one may still try to dissolve the paradox in a slightly different way: not anymore by arguing that

contemporary theories of emotions provide no reason for *endorsing* (PF3), but by arguing that these theories provide reasons to *reject* this proposition. Is this a viable strategy?

Here is a first reason to think it is not. In the light of the foregoing, such an appeal to emotion theory would no longer amount to dissolving the paradox—it would not support the conclusion that there is no paradox to begin with and that the three propositions can be maintained. The strategy would rather come down to a more traditional approach, that consisting of solving the paradox by rejecting (PF3).

A second reason is that it would be unsatisfactory to deny (PF3) on the sole basis that it is not vouchsafed by emotion theories. This constitutes a quite superficial solution, because it fails to engage with the undisputed phenomena that led Radford to formulate the paradox. In adopting it, one would fail to explain why existential belief is irrelevant for grieving Anna Karenina's sad end but of special relevance in most nonfictional contexts. As a matter of fact, Radford himself took a specific version of this solution into consideration. According to him, it amounts to viewing the emotional reactions stirred in us by fiction like so many brute psychological data, and he is keen on emphasizing how unsatisfying this is:

> [the] problem is that people can be moved by fictional suffering given their brute behaviour in other contexts where belief in the reality of the suffering described or witnessed is necessary for the response. (Radford, 1975: 72)

We concur in viewing this solution as a last resort, since it fares quite badly when compared with the other available ones. Consider the idea that, when immersed in fiction, our attitude is one of belief or at least of acceptance in the existence of fictional characters and events (e.g. Suits, 2006; Todd, 2013), or the idea that the emotions we feel towards individuals or events we believe not to exist are make-believe or non-genuine emotions (e.g. Walton, 1978; Currie, 1990). Whatever worries these solutions may raise, at least they engage with the underlying phenomena and are as a result in a position to provide two explanations of the puzzle. The

idea that our attitude towards fictional entities is one of belief reconciles the claim that genuine emotions require existential beliefs with the claim that we feel genuine emotions towards fictional entities. Alternatively, the idea that fiction elicits quasi-emotions may lead one to explain the everyday life cases by reference to (PF3), and to elucidate the mistaken impression that we feel genuine emotions about fictional entities by reference to similarities between emotions and quasi-emotions. The dissolution strategy is not of the same stripe, since it does not purport to explain the apparent incoherence of affective responses across fiction and nonfiction cases. Given that this is the source of the paradox of fiction, this strategy is not a satisfying solution to it. As we are about to see, this is the case even if (PF3) is at odds with emotion theories.

A third and final reason to reject this strategy is indeed that a version of (PF3) apt to generate a paradox need not run afoul of contemporary emotion theories. Of course, the version of (PF3) we have used until now (i.e. "To feel genuine emotions, we must believe that these emotions are directed to actually existing objects") is irreconcilable with them, as well as with introspective evidence. The mere idea of eating vomit ice cream seems after all to trigger disgust in most of us, and this independently of believing that production has already begun. Similarly, it has been shown that feces-shaped chocolates disgust people, even when they are fully aware that they are chocolates (Rozin, Millman and Nemeroff, 1986). That being said, the paradox does not presuppose that we endorse such a strong version of (PF3)—the claim that *some* of the emotions we feel towards fictional entities presuppose the relevant existential beliefs is sufficient. This is actually in tune with what Radford has to say. He does not claim that all emotions require existential beliefs, only that grieving for someone does so despite our susceptibility to grieve over fictional events. This stripped-down version of (PF3) does not appear to be in conflict with standard emotion theories. As advocates of the dissolution strategy themselves observe, these theories only accept that emotions may be

triggered absent such beliefs, which does not rule out that the occurrence of some emotions demands existential beliefs.

All in all, then, the paradox of fiction cannot be solved or dissolved by merely heralding the fact that a strong version of (PF3) is in conflict with standard emotion theories.

5. How Can Emotion Theories Solve the Paradox?

Against scholars wishing to dissolve the paradox of fiction (or claiming that it has already been dissolved), we have argued that there are reasons to endorse (PF3) and that a genuine paradox of fiction ensues. Despite these criticisms, we are sympathetic to two main tenets of the approach these same scholars favour: we agree that a satisfying solution to the paradox will come from a full-fledged theory of emotions, and that this solution will require rejecting (PF3).

That it is so can be argued from our claims regarding the phenomena underlying the paradox of fiction. Drawing inspiration from Radford, we have traced the source of the paradox to an apparent incoherence in our affective reactions. Belief in the existence of an emotion's object seems to matter in everyday life, but not anymore when we are engaged with works of fiction. We have insisted that a satisfying solution to the paradox must explain this incoherence, which requires developing a full-fledged account of the emotions and of their cognitive preconditions. Now, such an account will lead to the rejection of either (PF1) or (PF3), and (PF3) appears to be the most vulnerable proposition. As we have seen, (PF1) is intuitively much more compelling than (PF3): most participants are prone to reject (PF3) because they are convinced that they feel genuine emotions for fictional characters. Moreover, the main reason we have found for endorsing (PF3) is that it is the best explanation of the fact that, in some situations, learning that the object of one's emotion does not exist makes this emotion vanish. So, if one could find an alternative explanation of this fact that is compatible

with (PF1), one would be justified in replacing (PF3) with this alternative explanation. This might well qualify as the most satisfying solution to the paradox of fiction. It would preserve both (PF1) and (PF2), the two intuitively compelling premises, while at the same time offering an alternative explanation of the phenomenon that first motivated the adoption of (PF3).

Such a solution may not be out of reach. We have insisted that the paradox of fiction is a genuine problem, not that it is intractable. In fact, we think that contemporary emotion theories have already made a step in the right direction by introducing the idea that emotions differ from mental states such as sensory or perceptual experiences by having *cognitive bases* (Author a). The idea is that, contrary to perceptual experiences or sensations, emotions depend on other psychological states in order both to occur and to be about something. For example, to be sad at an event, one needs first to represent that event one way or another (perhaps one is visually aware of it, or one just formed a belief about it as a result of an inference). This is why the occurrence and persistence through time of emotions depend on the occurrence and persistence through time of their cognitive bases.[4]

[4] The idea that emotions depend on the prior occurrence of other psychological states is in line with contemporary theories of emotions. One prominent way of understanding emotions in contemporary psychology is appraisal theory (Scherer, 2001; Moors et al., 2013). According to appraisal theory, all emotions involve some kind of cognitive appraisal, i.e. some kind of evaluation of the situation. Appraisal theory is very liberal when it comes to the nature of the psychological states involved in appraisal processes, which can take as input representations coming from a variety of cognitive systems such as perception, memory and imagination. All these potential inputs are what philosophers describe as the cognitive bases of emotions. This dependence of emotions on a variety of psychological states has important consequences. For

The problem highlighted by Radford can now be reformulated in the following way. How is it that the occurrence and persistence of some emotions is sensitive to beliefs in the reality of their object, while the occurrence and persistence of affective reactions to fictional entities is not? In the light of the concept of an emotion's cognitive base, it seems that this question can receive a straightforward answer. The reason is that these emotions rest upon different cognitive bases: the former rest upon cognitive bases that are sensitive to beliefs in the reality of their object, while the latter do not.

It is important to observe at this juncture that there are two main respects in which psychological states can differ from one another. On the one hand, they may involve different *psychological modes*—we readily draw this distinction by means of psychological verbs, as when we distinguish imagining, desiring and believing that something is the case. On the other hand, psychological states may differ with respect to their *contents*, i.e. what they represent. This suggests that we should distinguish two ways of solving the paradox by insisting on the kinds of cognitive bases that emotions may have.

The first solution consists in appealing to a contrast at the level of psychological modes. One maintains that emotions about nonfictional entities are (typically) based on beliefs, whereas emotions about fictional entities are based on "non-serious" psychological modes, i.e. modes whose content is supposed or accepted without being endorsed as true. This explains why emotions about fictional and nonfictional entities react, and should react, differently to information. Believing that an entity exists and supposing that it exists are subject to different norms. For instance, one norm bearing on belief is that one should

---

instance, given that perceptual states do not manifest the same kind of dependence, it threatens perceptual accounts of the emotions that are nowadays quite popular in philosophy (Author a).

abandon the belief that *p* in the light of information that not-*p*, and obviously no parallel norm bears on supposition, imagination or acceptance. So, the fact that emotions about fictional entities do not respond to information in the way those that are about nonfictional entities do is a consequence of their being based on non-serious modes. According to this solution, we should reject (PF3)—this proposition holds true, it is claimed, only for emotions whose cognitive bases are beliefs, not for emotions that depend on "non-serious" modes such as imagining.

It might be thought that this first solution comes down to the popular strategy that consists in rejecting (PF1) and claiming that affective reactions towards fictional entities are quasi-emotions that depend on imagination rather than belief. This is not the case, since the solution under discussion is committed to claiming that affective reactions towards fictional entities are typically *genuine* emotions. The difference is significant. Appeal to quasi-emotions to solve the paradox indeed means endorsing two substantial and questionable claims about the nature of affective reactions to works of fiction.

The first claim quasi-emotions theorists are committed to is that affective episodes do not count as emotions if they are not based on beliefs. This should arouse suspicion, since it seems that the very same emotions can, depending on the context, be triggered by a variety of distinct psychological states. One may for instance feel ashamed after having realized that one has done something indecent, or at the mere thought of behaving indecently. Or consider disgust. It has been shown that, on some occasions, disgust rests upon having the appropriate beliefs: whether a beverage is judged to be disgusting depends on the belief that it was in contact with a dead cockroach. Yet, on other occasions, disgust is based on low-level perceptual states, and can even go against one's beliefs (Rozin, Milliman & Nemeroff, 1986). As we have already mentioned, feces-shaped chocolates can induce disgust, even when one is fully aware that they are made of chocolate. We seem to contemplate here two instances of

one and the same emotion, disgust. More generally, emotions of many given type seem apt to take a variety of doxastic or non-doxastic cognitive bases.[5]

Should we resist this conclusion? It is fair to say that we would need a substantial reason to do so. In the specific context of the paradox of fiction, we would need a reason to think that affective reactions towards fictional entities do not differ from genuine emotions to the sole extent that they take different cognitive bases. This is the second claim to which advocates of quasi-emotions are committed. In trying to substantiate it, they often contend that quasi-emotions differ from genuine emotions insofar as they do not motivate relevant behaviour. This seems to us to constitute a further liability of this approach, since there are clear cases of affective reactions towards fictional entities that motivate us in the same way as genuine emotions. For example, the action tendency typically associated with admiration is emulation, the desire to become as good as the person who is admired. And there is empirical evidence to the effect that admiring virtuous fictional characters can motivate real-life virtuous behaviour (Thomson & Siegel, 2013). The case of disgust discussed above can be used to the same end, since Rozin, Milliman and Nemeroff's data support the conclusion that all participants were motivated to avoid the source of their disgust, whether it had been elicited by or despite their beliefs.

As attractive as it might appear, the first approach and its characteristic appeal to a contrast at the level of psychological modes has to face substantial worries. The idea that supposing or imagining is the attitude typical of our engagement with fictional entities is less

[5] Many contemporary philosophers indeed insist on this latter possibility to argue that emotions cannot presuppose the relevant beliefs. See e.g. Döring 2007, Helm 2001 and Tappolet 2000. The fact that emotions can go against one's beliefs is often considered to be a reason to reject approaches to the emotions in terms of beliefs or judgements (see Author a).

straightforward than it may seem. As some have rightly emphasized, there appears to be nothing "putative, hypothetical or provisional" (Matravers 1991, see also Neil 1993) in how we relate to Anna Karenina. For instance, to describe the situation by saying that "it is as if" or that "one does as if" one were believing she is facing a dire end seems inappropriate. Yet these descriptions are fitting when we contemplate instances of imagining or supposing. Relatedly, emphasizing the contrast between believing and supposing may not be the best way to go since it applies equally to our engagement with fictional and nonfictional entities. One can after all believe *or* pretend that one's husband is facing a dire situation, as one can believe *or* pretend the same about Anna Karenina—and this difference in the psychological modes involved is likely to modulate the emotions that will ensue. So, why insist that our engagement with fiction builds exclusively upon non-serious modes? Since we seem to have in any case many serious beliefs about fictional entities, why not try to build an approach around them?

This leads to the second solution to the paradox of fiction. According to it, affective reactions to fictional characters do not differ from everyday emotions because their cognitive bases are non-serious, but rather because they are based on beliefs that have a specific kind of content. Advocates of this approach (e.g. Matravers 1991, Neill 1993) argue that emotions that target fictional entities are based on beliefs whose contents are more complex than those of the beliefs involved in emotions about nonfictional entities. More precisely, their contents are prefaced by a "fiction operator". For instance, one believes that, *in the fiction*, Anna Karenina faces a dire end. This explains why the relevant beliefs exhibit and should exhibit a different sensitivity to information about the reality of their objects. The sort of information to which a belief should be sensitive is obviously a function of its content and the belief that, in the fiction, Anna Karenina commits suicide should not be revised in the light of information that she is a fictional entity. The information it should be sensitive to concerns what happens

to her in the fiction, and it is exclusively provided by Tolstoy's novel (e.g. Livingston and Mele 1997). This generates a second explanation of the difference between our emotions towards fictional and nonfictional entities—since the former have a specific kind of content, they should, in order to be appropriate, respond differently than the latter to the information that their objects do not exist.

The main advantage of this second solution is that it avoids positing "hypothetical" attitudes towards fictional entities that seem to betray the nature of our typical engagement with them. But it raises one important worry, which regards its possible incompatibility with the sort of "immersion" that is, according to some authors, typical of our engagement with fiction (see Todd 2013). One may in this spirit insist that an advantage of the first approach is that it readily explains this phenomenon. In imagining that *p*, our attention is typically focused on what we imagine as opposed to the fact that we are imagining it, and this is why we are so often taken by or immersed in fiction. Introducing a fiction operator in the content of the relevant beliefs will make it more difficult to explain why this is so. Given the nature of this content, focusing on what we believe could not fail to make it salient that "this is only fiction", thereby blocking immersion.

6. Conclusion

The two solutions that we have briefly presented surely raise a number of worries, but they illustrate how contemporary emotion theories could be developed so as to do justice to our emotions towards fictional entities and solve the paradox of fiction. Reaching a full-fledged solution to the paradox obviously requires a lot more work, but our aim here was not to *solve* the paradox. In opposition to a recent trend in the literature, it was to explain why one cannot *dissolve* it in the light of contemporary emotion theory. The paradox resists such attempts because it gives expression to a deep psychological puzzle: why do emotions towards fictional entities behave so differently from emotions towards other entities? Insofar as this puzzle

constitutes the source of the paradox, an appealing solution or dissolution cannot fail to address it.

One may legitimately wonder whether the foregoing considerations help address the issues of rationality or appropriateness that are sometimes raised in discussions of the paradox of fiction. We shall thus conclude with some observations regarding how we think these issues should be addressed. Consider the following less popular but still widespread version of the paradox, which focuses not so much on what we feel but on what it is rational to feel.

(PF1*) It is not irrational to feel emotions for fictional characters.

(PF2*) We do not believe that fictional characters exist.

(PF3*) For an emotion to count as rational, one must believe that this emotion is directed to actually existing objects.

The question: "What is emotional appropriateness?" is challenging and we certainly cannot do justice to it here. Yet, highlighting two central ideas within recent approaches to that issue will hopefully prove sufficient to indicate the way to go.[6] The first idea is one that we have already encountered, namely that emotions have other mental states as their cognitive bases. The second is that emotions have evaluative conditions of appropriateness. To feel sad about a given event is appropriate to the extent that the event constitutes a loss, and in the same way it is appropriate to be angry at a person to the extent that she is offensive.

One appealing way to combine these two ideas consists in claiming that emotions inherit the correctness conditions of their cognitive bases and add an evaluative layer to them. In other words, if one is, say, visually aware of an event, then that experience is correct if and only if the event exemplifies the properties that one seems to be visually aware of. And if one react with sadness to what one sees, one's sadness is correct if and only if the event one sees

[6] For discussion and references to the rich literature on the underlying issues, see Author b.

constitutes a loss. If this is along the right track, then the appropriateness of emotions should be assessed in the light of, first, the appropriateness of their cognitive bases and, second, the evaluative properties exemplified by their objects.

As regards the former issue, section 5 gave us the opportunity to observe that emotions directed towards fictional entities, be they ultimately based on non-serious psychological states or on beliefs with complex contents, are subject to norms that differ from those to which other emotions are subject. So it would be wrong-headed to try to support (PF3*) by assessing emotions directed at fictional entities in light of the norms to which emotions based on simple beliefs are subject—for instance in light of the norm that they should vanish when one gains reasons to think these beliefs are unwarranted or false. This would betray a misunderstanding of the nature of emotional rationality and of the way it is modulated by the psychological states on which emotions are based.

The only way to support (PF3*) we can think of concerns the evaluative layer of emotional rationality—it consists in claiming that fictional entities are not the sort of entities that can exemplify evaluative properties. This actually allows one to recast the phenomenon Radford highlights. One may insist that, in nonfictional cases, the way our emotions vanish when we come to learn that their objects do not exist manifests our understanding that non-existent entities cannot exemplify evaluative properties, an understanding that fails to permeate our reactions to fictional entities.[7] What remains to be seen is whether this line of thought can be developed in a convincing way. It may well be the case that fictional entities cannot exemplify evaluative properties, but we surely need reasons to think that Anna Karenina does not face a dire end or that Lancelot is no courageous fellow. Additionally, we

[7] Unless, of course, such reactions are not genuine emotions.

would need to figure out why a sort of understanding that we deploy so readily in some contexts fails to permeate our attitudes towards works of fiction.

References

Author a.

Author b.

Currie, G. (1990). *The nature of fiction*. Cambridge: Cambridge University Press.

Döring, S. (2007). Seeing what to do: Affective perception and rational motivation. *Dialectica*, *61*, 363-394.

Goldstein, T. (2009). The pleasure of unadulterated sadness: Experiencing sorrow in fiction, nonfiction, and "in person". *Psychology of Aesthetics, Creativity, and the Arts, 3*, 232 – 237.

Helm, B. (2001). *Emotional reason: Deliberation, motivation, and the nature of value*. New York: Cambridge University Press.

Kim, S. & Hamann, S. (2007). Neural correlates of positive and negative emotion regulation. *Journal of Cognitive Neuroscience, 19,* 776-798.

Levinson, J. (1997). Emotion in response to art: A survey of the terrain. In M. Hjort and S. Laver (eds.), *Emotion and the Arts*. New York: Oxford University Press.

Livingston, P. & and Mele, A. (1997). Evaluating emotional responses to fiction. In M. Hjort & S. Laver (eds.), *Emotions and the Arts*. New York: Oxford University Press.

Matravers, D. (1991). Who's afraid of Virginia Woolf? *Ratio, 4*, 25-37.

Matravers, D. (2014). *Fiction and Narrative*. New York: Oxford University Press.

Mendelson, A. L., & Papacharissi, Z. (2007). How defined realness affects cognitive and emotional responses to photographs. *Visual Communication Quarterly*, *14*, 2 – 14.

Mocaiber, L., Perakakis, P., Pereira, M. G., Machado-Pinheiro, W., Volchan, E., Oliveira, L., & Vila, J. (2011). Stimulus appraisal modulates cardiac reactivity to briefly presented mutilation pictures. *International Journal of Psychophysiology*, *81*, 299 – 304.

Moors, A., Ellsworth, P. C., Scherer, K. R. & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review* 5, 119-124.

Neill, A. (1993). Fiction and the emotions. *American Philosophical Quarterly*, *30*, 1-13.

Oliveira, L. A. S., Oliveira, L., Joffily, M., Pereira-Junior, P. P., Lang, P. J., Pereira, M. G., & Volchan, E. (2009). Autonomic reactions to mutilation pictures: Positive affect facilitates safety signal processing. *Psychophysiology*, *49*, 870 – 873.

Radford, C. (1975). How can we be moved by the fate of Anna Karenina? *Proceedings of the Aristotelian Society Supplementary Volume*, *49*, 67-80.

Rozin, P., Millman, L. & Nemeroff, C. (1986). Operations of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, *50*, 703-712.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr & T. Johnston (Eds.), *Appraisal processes in emotion* (pp. 92-120). New York: Oxford University Press.

Sennwald, V., Cova, F., Garcia, A., Lombardo, P., Schwartz, S., Teroni, F., Deonna, J. & Sander, D. (2015). Is what I'm feeling real? Fiction versus reality. Unpublished manuscript, University of Geneva.

Stecker, R. (2011). Should We Still Care about the Paradox of Fiction? *British Journal of Aesthetics*, *51*, 295-308.

Suits, D. (2006). Really believing in fiction. *Pacific Philosophical Quarterly*, *87*, 369-386.

Tappolet, C. (2000). *Emotions et valeurs*. Paris: Presses universitaires de France.

Thomson, A. L., & Siegel, J. T. (2013). A moral act, elevation, and prosocial behavior: Moderators of morality. *Journal of Positive Psychology*, *8*, 50-64.

Todd, C. (2013). Attending emotionally to fiction. *Journal of Value Inquiry*, *46*, 449-465.

Tullmann, K., & Buckwalter, W. (2014). Does the Paradox of Fiction Exist? *Erkenntnis*, *79*, 779-796.

Vrticka, P. Sander, D. & Vuilleumier, P. (2011). Effects of emotion regulation strategy on brain responses to the valence and social content of visual scenes. *Neuropsychologia 49*, 1067-1082.

Walton, K. (1978). Fearing fictions. *Journal of Philosophy*, *75*, 5-27.