Does the Paradox of Fiction Exist?

Katherine Tullmann · Wesley Buckwalter

Abstract   Many philosophers have attempted to provide a solution to the paradox of fiction, a triad of sentences that lead to the conclusion that genuine emotional responses to fiction are irrational. We suggest that disagreement over the best response to this paradox stems directly from the formulation of the paradox itself. Our main goal is to show that there is an ambiguity regarding the word 'exist' throughout the premises of the paradox. To reveal this ambiguity, we display the diverse existential commitments of several leading theories of emotion, and argue that none of the theories we consider are committed to notions of 'exist' employed by the paradox. We conclude that it is unclear whether or not there remains a paradox of fiction to be solved—rather than to be argued for—once this ambiguity is addressed.

1 An Emotional Introduction

Suppose you are watching a film and are moved to tears by a protagonist's horrible plight. Imagine sitting around a campfire, shaking in fear from a scary ghost story, or laughing uproariously at the jokes of a Saturday morning cartoon. It seems as though we frequently have genuine emotional responses to fictions in precisely these ways. But now examine your beliefs about the very same protagonist whose situation you previously found so heartrending, or the ghost whose aura you found

K. Tullmann
Department of Philosophy, Graduate Center, City University of New York, New York, NY, USA
e-mail: katherine.tullmann@gmail.com

W. Buckwalter (✉)
Department of Philosophy, University of Waterloo, 200 University Avenue West, Waterloo, ON
N2L 3G1, Canada
e-mail: wesleybuckwalter@gmail.com

so frightening. While the goose bumps are real, their elicitors are not. We don't believe that the characters of fictions actually *exist*.

These purported facts about human behavior give rise to what is known as the paradox of fiction. This paradox, initially formulated by Radford (1975), is comprised of a triad of inconsistent, yet highly intuitive propositions. First, some emotional responses to fiction count as genuine emotions. Second, we don't believe that the objects of those genuine emotions—the fictional entities—actually exist. And third, we are only rationally moved by what we believe actually exists. Or in other words, having a genuine emotional response implies a belief in the existence of a fictional thing. Radford argues that all three of these premises are true. He concludes that emotional responses, such as sadness for our protagonist or fear on account of our ghost, are irrational through and through.

Aestheticians have pursued a number of approaches to rescue people's ordinary emotional responses to fiction from this threat of irrationality. Quite often, these rival approaches reflect substantive philosophical differences regarding the nature and cognitive science of emotion. And debate among researchers regarding the nature and structure of emotions is certain to continue. So while defenses and critiques of the paradox have been numerous, the correct response to Radford's triad remains the source of continued and ever broadening controversy. And though in recent years certain neo-Jamesian theories of emotion have become popular, no one theory or perspective has emerged the clear victor.

We suggest that the correct response to the paradox of fiction need not reflect any substantive philosophical differences regarding the nature or cognitive science of emotion. Nor do we think the correct response need involve defending the controversial denial of one or more of Radford's highly intuitive premises. Instead, the main goal of our paper is to show that current disagreement among philosophers about the proper response to this paradox stems from the formulation of the paradox itself. In short, we will argue that despite one's theoretical commitments, there is good reason to think there is *no such thing as the paradox of fiction*.

In the remainder of the paper we defend this bold claim. In Sect. 2 we will display the purported paradox of fiction, and briefly review a number of influential responses to it. In Sect. 3, we will identify what we believe to be the central mistake in the formulation of the paradox, an ambiguity regarding the word 'exist' in the following two premises:

(1) One does not believe that fictional entities *exist*.
(2) A genuine emotion in response to fictional entities implies that one believes that fictional entities *exist*.

In Sect. 4, the longest and most important section, we review the commitments of several leading theories of emotion, and argue that according to these theories, a uniform notion of 'exist' has yet to be deployed in (1) and (2). We conclude in Sect. 5 with a discussion of the implications that this finding has for future debates concerning the paradox of fiction. We argue that it is unclear whether or not there remains a paradox of fiction to be solved—rather than to be argued for—once the ambiguity we point out is addressed. Consequently, we suggest that if we are correct then attending to Radford's premises may not be a particularly fruitful way to

answer the central philosophical questions typically of interest to philosophers in this domain: the ways in which fictions and fictional objects exist, and what emotions are able to tell us about them.

2 The Paradox and Responses

The paradox of emotional responses to fiction stems from an inconsistent triad comprised of Radford's three premises (adapted from Radford 1975):

> (PF1) We are genuinely moved by fictions. Emotions in response to fictional things can be genuine.
> (PF2) We do not believe that fictional entities exist. We believe that what is portrayed in fictions is non-actual.
> (PF3) A genuine emotion in response to fictional entities implies that one believes that fictional entities exist. We are only genuinely moved by what we believe is actual.

For example, when Frodo and Sam finally reach Mt. Doom in Tolkien's *Lord of the Rings*, one may feel relief, exultation, or pride at their accomplishment. Furthermore, suppose it's indeed possible that these emotions are genuine tokens of relief, exultation or pride, and not, quasi or pretend emotions. Even though we are affected by Frodo and Sam's story, we know that these characters are not real—they cannot be found in the actual world. Their journey stems only from the clever imagination of the author of the fiction. So, if we believe that Frodo and Sam are not real, why do they affect us *at all*? The answer, Radford suggests, is that these kinds of emotional responses to fictional characters are irrational.

Most theorists reject Radford's conclusion that these kinds of responses are somehow irrational or inappropriate. Instead, traditional solutions to the paradox of fiction have proceeded by rejecting one or more of the propositions above. Many theorists have rejected (PF1) by denying that emotional responses to fictions are *genuine* emotions. Perhaps the most well known response along these lines is the make-believe theory proposed by Walton (1978). In "Fearing Fictions," Walton utilizes a complex analogy of a child's game of make-believe or pretend-play. When children play 'house' for instance, they generally designate certain objects as something else. And similarly, when engaging in fictions, it might be appropriate to say that someone "pities" Anna Karenina or "admires" Superman even if what they say is not literally true ((1978), 9). Rather, it is only a matter of *make-believe* within the context of the story.

But theories of make-believe have been heavily criticized because of the important differences between prototypical instances of pretend play, on the one hand, and the character of our everyday emotional responses to fiction, on the other. For instance, Carroll (1990) argues that unlike the make-believe children often engage in, our emotions are not easily turned on and off at will. When we are emotionally engaged in a particular work of fiction like a horror film, it seems we often have very little control over the fear or terror that we feel. And similarly, that experience has a very different phenomenology than pretend-play (Hartz 1999).

While the latter notion of pretence presupposes some kind of awareness of a game being played, our emotional responses often do not. So the burden of the pretense theory is to somehow explain these differences, as well as the possibility of severe and lasting emotions in response to fictions that *feel* perfectly real.

Another well-traveled option for dismantling the paradox involves denying the second proposition (PF2). Often theorists who endorse this approach advocate illusion, or "suspension of disbelief" arguments (see Coleridge 1985/1817; Hurka 2001). The basic idea is that for the time in which we engage with a particular fiction we *do* believe that the fictional entities are real. While reading *Lord of the Rings* for instance, illusion theories hold that we temporarily believe that Frodo's world exists, and this illusion is responsible for our genuine emotional responses to fiction. But of course, one straightforward objection to this view is that no matter how caught up we are in fiction, it seems that we never *actually believe* that what we experience is real. If asked about a film later, we never claim that there was some point at which we stopped believing that the story was fiction, or that the monsters of a horror film might actually be able to harm us. And at no point were we inclined to act in the way appropriate for an encounter with a harmful object. We didn't run from the theatre or call the police. Nor did we warn our beloved that they were in danger. As Katherine Thomson-Jones points out:

> I am able to appreciate the vivid depiction of an army of zombies surging forward with arms outstretched, the use of special effects or highly emotive music, the importance of the scene for the narrative, and so on. Surely, if I had suspended my belief that the zombies are fictional, I would be too frightened to appreciate film in this way (2008, 107).

Thus theorists who support illusion approaches must somehow square suspension of disbelief arguments with the contrary beliefs and actions people seem to have in response to fictional characters and situations.[1]

Based on the challenges faced by these early attempts rejecting (PF1) and (PF2), most philosophers now attempt to solve the paradox by rejecting the last premise. In fact, it seems that most of the contention surrounding the paradox arises from accepting the following trenchant requirement of emotions: we must believe that the objects of our emotions exist, and that they can affect us in some way in order to be genuine. And of course, one might think that this is not the case with fictions. Thus many philosophers have tried to obviate this intuitive requirement of emotions by arguing—in denial of the third premise—that beliefs of this sort are not necessary for emotion.

There are two basic approaches for denying that genuine emotional responses require belief in the existence of fictional things in (PF3). First, non-cognitivists about emotion—or theorists who deny that cognitive content is a necessary component of emotions—can simply reject the claim that emotion requires beliefs. This is the route taken by philosophers such as Robinson (2005) and Prinz (2004a),

---

[1] For more on motivation and the role of behavioral circumscription in theories of belief see Rose et al. (under review) as well as Tullmann (in prep), Buckwalter et al. (2013) and Buckwalter and Turri (under review).

who propose non-cognitivist theories of emotions.[2] The other approach accepts that emotions require a cognitive component, but rejects that this component must involve beliefs. Carroll (1990) favors this solution, arguing that our emotional responses to fictions presuppose our *thought contents* about the fiction, rather than our beliefs about the fiction.

Proper discussion of these responses to (PF3) is beyond the scope of this article. Nonetheless, one should note that these approaches require one to accept accounts of emotion that are not universally accepted.[3] There is little doubt that a correct analysis of emotion in modern cognitive science and aesthetics is less forthcoming than a solution to the paradox itself. Since philosophers are roughly split on the question supporting cognitive or non-cognitive theories of emotion, methods involving the denial of (PF3) are even less likely to reach theoretical consensus in debates about the paradox of fiction.

Lack of consensus above might tempt us to bite the bullet. Radford's response to the puzzle was to simply accept the irrationality of our emotional responses towards fictions. As Radford concludes:

> I am left with the conclusion that our being moved in certain ways by works of art, though very 'natural' to us and in that way only too intelligible, involves us in inconsistency and so incoherence (1975, 78).

Radford likens this to other everyday situations such as when "a tennis player who sees his shot going into the net will often give an involuntary jump to lift it over." Like the tennis player, we often encounter situations in which we are aware that our beliefs and actions do not match up to reality in the right way. According to Radford, our emotions concerning fictional characters are one of these instances, reflecting an important aspect of our cognitive and behavioral makeup.

But we are far from convinced that the appearance and persistence of the paradox arises as a result of these sorts of behavioral phenomena. Nor do we think that the solution rests in the responses we have reviewed above. Instead, we think the paradoxical nature of the premises arises due to an ambiguity in the language philosophers have used to formulate and respond to this purported paradox. In the next section, we attempt to reveal this ambiguity.

3 The Ambiguity of "Existence"

Our focus is on the use of 'exist' throughout the triad. It appears in two premises. Specifically, (PF2) holds that one does not believe that the fictional objects of our emotions actually *exist*. On the other hand, (PF3) stipulates that having a genuine emotional response implies that one believes the object of that emotion *exists*. But in

---

[2] Throughout the remainder of the paper we use 'non-cognitivism' to label those theories that reject cognitive content strictly as a necessary condition of emotion. We note however that Prinz and Robinson do *allow* for emotions that are evaluative and intentional—and are sometimes thought of as 'cognitive' on this basis.

[3] For instance many allege that thought-based theories are *overly cognitive* theories of emotion, while others argue that neo-Jamesian accounts are *not cognitive enough* (see Turvey 1997).

what way do these fictional things *exist*, and is the same notion of existence being employed across both of these premises?

From Aristotle to Frege and Kripke, the history of philosophy is shot through with discussion of existence. And especially since Frege, a considerable amount of work in the philosophy of language has gone into giving an analysis of existence, its multiple senses, and the similarities and differences it bears to what is expressed by the word 'is' (see Williams 1981; Miller 1986; Zalta 1988). Using examples like Sherlock Holmes or Spock, a number of influential philosophers have even argued that fictional characters do constitute *existing entities* in the sense that they are abstract artifacts (Salmon 2002; van Inwagen 1977; Thomasson 1999). But despite this tradition, such scholarship is typically not reflected in contemporary discussions in the philosophy of art concerning the paradox of fiction, or in cognitive science regarding the objects of emotion. Instead, existence (and 'existence' for that matter) is usually left unspecified in these discussions. In fact, it seems little has been said about the existential nature of the objects of our emotions, both in the literatures on fiction and emotions.

Perhaps one reason the salient sense of existence being employed in the paradox of fiction has gone underspecified is because it is meant to capture an ordinary or non-technical notion. Consider for instance the proposition "One does not believe that fictional entities exist" in (PF2). As far as we know, very few who have written on the paradox have explained precisely what they mean when they use the word 'exist'. But we are left with a few obvious candidates. For instance, on an initial, commonsense view, it could be that 'exist' means something like *is a concrete object*. We say that people, trees, paintings, etc. exist because they are currently composed of matter. Of course, some cases are not so straightforward: consider electrons, quarks, and musical works, not to mention fictional characters. It is an obvious point that fictional entities lack a certain kind of physical body. If that's so, then it's probably also the case that this is not the sense in which 'exist' is used in (PF3). But this isn't a trivial point. Consider how many philosophers handle (PF3). They reject that audiences suspend their disbelief concerning fictions because if they did, the audience members would act differently towards fictional objects—they would run screaming out of the theatre upon discovering that zombies or monsters occupy their town, for example. A suspension of disbelief in the existence of zombies implies that we pretend or imagine that zombies exist in the sense discussed here—concretely. The fact that this view is taken seriously (for example by such luminaries as Coleridge and Hurka) means that it is worth considering as an option in the paradox.

Alternatively, 'exist' could mean *might have existed concretely*. This is, in fact, a variation on the position held most notably by Lewis (1978) and Plantinga (1974)—fictional entities are possible objects that do not exist in our world, but do exist in another possible world. In that sense Emma Woodhouse might have existed in the actual world had the contingent historical events turned out differently. Similarly, one might think that for all we know Sherlock Holmes could exist, just like black swans and white crows.

Needless to say, this view has been widely critiqued in the literature on the ontology of fictional entities (see especially Quine 1953; Thomasson 1999 and

Sainsbury 2010). It is generally accepted by those who think that fictional entities exist at all (perhaps except for Meinongians) that they are created and causally dependent. There is no Sherlock Holmes in the actual world, and even if there were a Sherlock Holmes in a possible world that matched each of the detective's known properties, he would not be the same Sherlock Holmes as the one in Sir Arthur Conan Doyle's stories. Because of such considerations, we may be tempted to reject this sense of 'exist' out of hand. Nevertheless, we encourage the reader to keep in mind that this is a plausible notion of how various things exist, and a viable option in the paradox.

Lastly, we get the sense that when theorists speak of fictional objects in the paradox, they are suggesting that to be fictional is to be an *imaginary* object. What we mean by 'imaginary' is simply the inverse of the first two candidates–an object that is not a concrete particular or even possibly so, in the two senses described above. This is the most plausible notion of 'exist' used in reference to fictional entities, but also the vaguest. After all, what does it mean for a fictional character to be imaginary? When we think of an imaginary friend, for instance, we mean something like "not having a physical body" and "made up by me". An imaginary friend exists in the sense that it is created by a child in the actual world, is not corporeal, nor merely a concrete particular in a possible world.

Of course, there are a number of positions philosophers might take on imaginary objects: whether or not fictional character are real in *any* sense, whether they are created or have always existed, and how we use them. While we do not want to endorse any one particular theory of the ontological status of fictional characters, we do wish to mention that one major component typically included by theories of imaginary objects is an account of how they bear meaningful causal relations to their imaginees—or real human beings. Often, this is said to work through some kind of simulation theory (Currie 1989) or theory of make-believe (Walton 1993; see also Kieran 2010) as described above. But one need not posit a theory of make-believe. As Carroll argues, it seems we can engage in and justify actions, as well as form propositional attitudes, in regards to non-assertive thoughts about fictional entities (1990). Alternatively, another possibility is that our responses can result from unconscious or non-cognitive triggers in our environment, including fictions, as argued by non-cognitive theories of emotion (see Sect. 4). So a final way to think about 'existence' in the triad is in terms of 'imaginary', together with some account of the causal relationship between imagination on the one hand, and our thoughts and actions on the other.

Taking stock, we certainly don't think that any of these suggestions will yield particularly full-fledged metaphysical theories of existence.[4] We're simply trying to understand what philosophers of art typically have in mind when they commit to (PF2). And on that score, we think we've identified three relatively good, albeit basic, candidates:

(3)   One does not believe that fictional entities are *concrete particulars.*
(4)   One does not believe that fictional entities *might exist concretely.*

---

[4] To be clear, we remain theory-neutral on the "reality" of fictional entities and reference to them, except insofar as they relate to objects of emotions.

(5) One does not believe that fictional entities are *non-imaginary* (that they exist concretely, might exist concretely) or are other than imaginary objects.

Our next task is to take these candidates for 'exist' and deploy them uniformly for occurrences of 'exist' throughout the paradox. Specifically, we would like to know how these candidates affect the conditional premise that a genuine emotion in response to fictional entities implies that one believes that fictional entities exist in (PF3).

4 The Objects of Theories of Emotion

In the previous section, we've proposed three candidates for understanding 'exist' in (PF2). To clarify the sense of 'exist' in (PF3) however, we need to know more about what it means to undergo an emotion. While there is a wide spectrum of emotion theories advocated in the literature, these accounts are typically classified into three groups. On one end of this spectrum are *pure-cognitive* theories. As the name suggests, these theories maintain that having a belief, judgment, or thought is both necessary and sufficient for genuine emotion.[5] At the other extreme of the emotional spectrum are the *non-cognitive* theories. Such theories hold that feelings, bodily changes, or perceptions of those changes are necessary and sufficient for genuine emotions. Finally, somewhere in the middle lie *hybrid* theories. These theories hold that both feelings of some kind as well as some kind of intellectual, or cognitive state are necessary for a genuine emotion.[6]

Our task is to specify the notion of existence at work in premise (PF3). In order to understand this, we must now determine which sense of 'exist' is presupposed by each type of theory on the emotional spectrum. We will focus on one leading theory from each general type: feeling (non-cognitivist), judgment (hybrid) and belief-based (pure-cognitive) theories and examine their commitments about emotional objects in this debate. In many cases, the same conclusions can be drawn across the general type of theories as well.

Before proceeding, we should point out that it is actually no easy matter determining how an object of an emotion must exist for each type of theory. The reason is that authors rarely spell out the existential commitments of the objects of their emotive theories. And when they do, it's not clear in what sense of 'exist' they employ. Nevertheless, what we will attempt to show is that, for each of these theories, a consistent use of 'exist' has yet to be applied across the following two statements:

(6) One does not believe that fictional entities φs.
(7) A genuine emotion in response to fictional entities implies that one believes that fictional entities φs.

---

[5] Interestingly, it is this kind of theory that the paradox of fiction was traditionally construed to address—emotions are caused or constituted by beliefs. See also footnote 2.

[6] This terminology is based on Prinz's characterizations of theories of emotions in *Gut Reactions* (2004a).

In short, we think it is unlikely that any of the types of theories we consider below apply the likely candidates we have proposed for 'exist' in (3–5) uniformly in (6–7).

4.1 Feeling Theories

As the name suggests, feeling theories describe a type of non-cognitive position that emphasizes both the conscious feel of emotion, and often also includes the conditions of bodies that undergo them. William James, one of the first prominent proponents of a feeling theory, famously stated that, "our feeling of the same changes as they occur *IS* the emotion" (1884). In other words, the felt bodily changes that we experience in response to various stimuli in our environment constitute an emotion, rather than any thought or belief about that object. We experience fear, for example, when we feel the hair stand up on the back of our neck, our heart pounding in our chests and our hands clutched beside us. We experience joy by a feeling of lightheartedness and buoyancy of gate. Things like beliefs, thoughts, or other propositional judgments are not required. Emotions are primitive in this way—they will occur without reference to conscious, cognitive mental states (see De Sousa 1987).[7] Such theories are associated with the philosophers Goldie (2004) and Stocker (1983), as well as the neuroscientist LeDoux (1996). We will focus here on Goldie's rendition.

Even though emotions may be characterized as "primitive," most feeling-theorists also maintain that they are not "brute". Emotions don't simply occur randomly from any stimuli in one's environment. Rather, emotions "tell us something about the world and how to act in the world" (Goldie, 92). Indeed, this is a basic theme across most theories, cognitive and non-cognitive. Emotions are intentional or about the world and value-laden. It's a common misconception that non-cognitive theories of emotions, including feeling theories, are not intentional. In contrast, Goldie maintains that the epistemic role of emotions is to reveal things about the world, provide a subject with relevant information concerning its well-being, and gain insight on how to act.

Goldie calls this "extraspective" knowledge-information about the world around us. Many objects have "emotion proper-properties"—the property of being able to cause a certain bodily state associated with an emotion (94). Some feeling theorists define these properties in terms of "core relational themes" (Prinz 2004a, b). For example, a grizzly bear is associated with the core relational theme of fear because it can cause us harm. Importantly, for all non-cognitive theories of emotions, the cause of the emotion need not be cognitive. We may only need to perceive an object in our environment in order for it to trigger an appropriate emotion (a flight response, for example).

However, emotions can also result in thoughts and introspective knowledge about our bodily condition. I learn about my own bodily state by the way emotions feel (93). This can give us information about how we personally respond to various

---

[7] Many non-cognitivist theorists maintain that the associated bodily changes need not be conscious. See Prinz 2004a.

situations, as well as clue us in on which objects of situations to pursue or avoid. Importantly for our purposes, neither introspective nor extraspective knowledge need result in an assertive belief concerning the object, situation or state of affairs that caused the emotion. We can be seriously wrong about what caused our fear, for example. We might think, alone in our dark bedroom at night, that a quick-moving shadow caused by a car passing outside of our window is actually a burglar attempting to break in. Just the perception of this shadow is enough to cause an emotion. But we need not have any thoughts or beliefs about what this shadow actually is or what harm it can cause us in order for us to we feel fear.

This is the key to understanding how we can have genuine emotions about fictions. We perceive objects on a screen, or in a play, or even in our imagination (as when reading a novel) as possessing an emotion-proper property. The result is a felt bodily change—an emotion. The feeling is real even if the object that caused it is not, and so the emotion is genuine. This gives us an indication of how feeling theories typically think about existence of emotional objects. Because emotions are primitive in the way described, the object of an emotion need not be concrete particulars, possibly concrete particulars or non-imaginary. In other words, a non-concrete, possibly concrete, or imaginary object may cause our emotion by possessing the right kind of emotion-proper property.[8]

What does this mean for the paradox? We gave three candidates in (3–5) about what existence might mean in premise (PF3). And if those candidates are applied consistently throughout the paradox, it would imply that an object of an emotion must be either concrete or non-imaginary. But according to feeling theories, we can have genuine emotions about all of these things. This is established quite independently of the paradox of fiction. The paradox need not arise for these non-cognitivist theories because none of the common-sense uses of 'exist' in (PF2) reflect the feeling theorist's use of 'exist' in (PF3).

4.2 Judgment Theories

The judgment theory of emotions is probably most associated with Solomon (1993), (2004). Solomon, like the feeling theorists, argues that emotions are evaluations about the world. Unlike the feeling theorists however, he posits that emotions involve more than bodily reactions to stimuli, but also a cognitive judgment (or judgments) of them. The judgment serves as a kind of appraisal, evaluating objects of value to us in our environment.

While his early work reflected a pure-cognitive stance, Solomon's later work repeals that notion in favor of a hybrid view ((2004), 85). He still maintains that judgments are necessary for emotions to occur, as well as to cause an emotion. However Solomon also holds that these judgments can be *both* intellectual and bodily. And perhaps most importantly, these judgments need not be conscious, reflective or propositional. We might note that if a judgment is either bodily *or* intellectual (in some

---

[8] Of course, the question of perceiving fictions becomes a bit more difficult when talking about multiple genres, and for instance, if experiencing literature counts as direct perception (Goldie 100). We think that a case can be made that imagining an emotion-proper property of an object in such cases is reasonable, though will postpone this argument for another occasion.

way), as well as not necessarily conscious, propositional or reflective, then it's rather difficult to see what exactly a judgment *is* at all. Solomon more or less sidesteps this issue, instead focusing on an emotional judgment's relation to a subject's world. Like Goldie, Solomon argues that emotions are an "appraisal of the world and our place in it" (81). Emotions are about the world, an evaluative process that includes dispositions to respond in certain ways to various stimuli, perceptions of bodily states and, often, thoughts or beliefs about our environment.

Solomon's characterization of emotions as "appraisal[s] of the world and our place in it" seems to suggest that the relation between self and world required cannot be met by fictions. But again, it's unclear what exactly Solomon means by "world and our place in it." Can we only have emotions regarding the actual, present world? This doesn't seem completely correct, since he also claims that emotions can be appropriate appraisals of propositions, ordinary objects or perceptions, and imaginary or remembered things (82). So "the world" must be understood in a broad sense, including imaginary and plausibly, fictional entities. This understanding of "world" gives us a relatively straightforward way to analyze Solomon's existential commitments in terms of fictions. A judgment theorist may deny that emotional judgments must be about concrete or possibly concrete things [statements (3) and (4)]. Solomon explicitly says that we can have emotions about imaginary things (5).

How would an appraisal of a fictional entity work, according to this view? Consider our reaction to Anna Karenina's plight. How can we say that we genuinely experience pity for Anna when she is not allowed to visit her son, for example? Perhaps our judgment would be something like the following: we grow attached to this character as we read about Anna's trials, judge her to be a sympathetic character and formulate different opinions about the other characters in the novel based on this regard. As we "witness" Anna's misfortunes, we formulate appraisals of her situation much as we would with any friend with whom we sympathize in the actual world who undergoes the same sorts of experiences. In a way, this appraisal concerns *us* because it is a reflection of our own sentiments and tacit beliefs about how women should be treated, what constitutes unjust behavior, etc. Since Solomon maintains that emotions are also bodily, we may experience a variety of bodily changes associated with pity, sorrow or anger about Anna and her peers which, in part, constitute the emotion.

Again, Solomon never plainly states what it would mean for us to have an emotional appraisal about fictions, so it's difficult to say exactly how the judgment process would work. But we think that something like the process outlined here is consistent with Solomon's view. So as we saw in the case of feeling theories of emotion, the paradox of fiction need not arise for judgment theories similarly construed to Solomon's. We can have genuine emotions about fictions, where the emotion is some sort of appraisal about the fiction that concerns us. We do not believe that fictions are real, in either sense of concrete. We may have genuine emotions about things that we don't believe to be real because we may have genuine emotions about imaginary things. The commonsense uses of 'exist' in (PF3) are just not the same notion of 'exist' that the judgment theorists requires of emotional objects in order for one to have a genuine emotion.

4.3 Pure-cognitive Belief Theories

Kendall Walton, a pure-cognitivist, is one of the most out-spoken deniers that we have genuine emotional responses towards fiction. Pure-cognitivist theories are often expressed in terms of belief. Typically, belief-based theories of emotion hold that our emotional responses are caused by the beliefs we have concerning objects in our environment that bear on our well-being (see Nussbaum 2004). On this type of view for example, one fears a tornado because one believes that the tornado may do harm. For the belief-theorist this cognitive content is a necessary component of emotions. Subsequently, belief-based theories like Walton's robust view on fictions are perhaps the most difficult type of theory for us to explain to reach the conclusion that the paradox of fiction is really no paradox at all.

Before we begin however, we would like to point out that while Walton commits to a hard line cognitivist theory, he does not himself propose any specific theory of emotions over and above his work relating emotion to action. But since Walton provides no specific theory of emotions, one natural question to ask here is whether or not there is something about belief-based theories in general that forces one to accept Walton's claims about the existence of fictional objects *tout court*. Here, we will argue that the answer is 'no' insofar as prominent belief theories today—such as the one proposed by Martha Nussbaum—do not seem to be subject to the existential commitments we have been considering.

In his famous "Charles and the Green Slime" example (1978), Walton argues that Charles does not feel *genuine* fear about the oozing monster, even though he displayed all the typically signs of fear—clutching his armrest or covering his eyes, jumping out of his seat, etc. But if it looks like Charles fears the slime by displaying all of the relevant behaviors associated with fear, then why shouldn't we say that his fear is real? Walton's answer, of course, is that Charles does not have the right sort of belief about the green slime in order to constitute a genuine emotion. Namely, Charles does not believe that the slime can actually cause him harm. If he did, he would surely act on this belief. In other words, emotions are *motivating.* And this is the significant difference between emotions about the fictional world and those about the real world. Real-world emotions motivate us to act in response to them—in response to whatever object or state of affairs actually caused the emotion to begin with. Charles was never motivated to act on his "quasi-fear." He never runs out of the theater, putting as much ground between him and the slime as possible. He doesn't call the fire brigade or army reserve to put the slime down once and for all. These things simply do not happen in regards to fictions. Charles lacks the motivation to act, and so is missing an important element of a genuine emotion.

Nussbaum's "neo-Stoic" theory of emotions works in a similar way. Nussbaum highlights four different features of an emotion. First, emotions must have an object. Second, that object must be interpreted in a certain way. Third, emotions are based on beliefs about the world. Fourth, these beliefs are value-laden (190). The picture of emotions we get from Nussbaum places special emphasis on the idea that emotions are evaluative and "eudaimonistic"—they have to do with human flourishing (189). An emotion occurs, or will occur, when we evaluate some object

or state of affairs as having a "non-trivial" (188) bearing on ourselves or on those we care about, and believe that this is so.[9]

But alas, the ambiguity in 'exist' can be found here as well. Walton and Nussbaum do not clarify how they use this term. The general assumption in premise (PF2) seems to be in line with (3) or (4)—that the object of an emotion must be a concrete particular or possibly so.[10] However, Nussbaum explicitly claims that the object of a genuine emotion need not exist in the world, presently or *ever*. While this idea is not developed in detail, we take Nussbaum's position to be that these kinds of entities nevertheless *do* instill in us the right sort of belief or elicit the right sort of evaluation to constitute an emotion. Conversely, it seems like Walton would adamantly deny this, claiming that entities that are not concrete, possibly concrete or non-imaginary can never cause the right sort of evaluative belief in order to give rise to genuine emotion.

So the crucial question for the emotion theorist who adopts the pure cognitivist theory about belief is to decide which kinds of entities can give rise to the right sort of belief for emotion. And despite Walton's hardline stance, we think that, in principle, it is possible that imaginary objects *can* cause genuine emotions under this type of pure-cognitive theory. Importantly, as we saw in the example of Charles and the Green Slime, the right kind of belief (according to Walton) has to with *motivation*. The key is to determine what kind of existence an object must have in order to achieve a belief that inspires the right kind of value or motivation to act. This is illustrated in the following example:

> Becky has bad luck in her romantic relationships. Recently she read Jane Austen's *Pride and Prejudice* for the first time. Becky is stuck by Austen's description of the depth, intelligence, sincerity, and, let's face it, wealth and physique of this fictional character. She imagines a person just like Mr. Darcy that exists in her world, and has certain emotional responses towards him: longing, hope, etc. Becky even goes so far as to change her appearance, reading list, and locale in order to appeal to men like the ones in her fantasy (perhaps *he* is out there now).

It's not hard to imagine real situations like this. The question is, would Becky's imagined, Mr. Darcy-like entity fall under the first, second, or third notion of existence that we've discussed? Clearly not the first; much to Becky's dismay, the imagined Mr. Darcy is not concrete. Whether the imagined Mr. Darcy would be possibly concrete or strictly imaginary may depend on one's theory of fictional entities. Walton would claim that the imagined Mr. Darcy here is not imaginary in our third sense and so does not present a counterexample to his theory. Rather, Becky has genuine emotions about a possible actual person, which, he would argue, is perfectly fine for his theory (see Walton 1993).

---

[9] Note that neither Walton nor Nussbaum discuss the possibility of unconscious emotions. Perhaps we can have an unconscious belief about a fictional character that triggers an emotion. This, however, seems to go against the spirit of the pure-cognitive project.

[10] On this view one might also wonder about the status of abstract systems, mathematical and logical entities that sometimes evoke emotional responses—we find them beautiful.

This reply might be suitable for possibly concrete objects—an imagined tornado, an imagined unicorn, etc. But does it work for fictional entities? Consider the following adaptation to the example above:

> The unlucky-in-love Becky restricts her imagination to the *character* Mr. Darcy, just as he is described in Jane Austen's novel. It is this Mr. Darcy that serves as the object of her emotions (longing, wistfulness—but also, this time, a strange jealousy for Elizabeth Bennett and regret that such a man doesn't exist). These emotions may even cause Becky to engage in certain peculiar actions: she places her copy of *Pride and Prejudice* on her nightstand in an oddly affectionate manner and makes caustic remarks about Elizabeth Bennett's contrariness to her friends.

What might we say about this example? If we agree with Lewis and Plantinga's view of the nature of fictional characters, then we might be inclined to say that the two Becky cases are similar in at least one significant respect: they are both entities that exist in another possible world, but not the actual one. We hold, in contrast, that there is a marked dissimilarity in the two cases. Recall that there are many arguments against that claim that fictional entities are possible objects. We can describe two of the main critiques without delving into too far into this literature.

First, there are too many possible objects that are relevantly similar to a fictional character to identify them. As Thomasson notes, the descriptions of fictional characters in works of art are incomplete in detail—they do not typically describe the character's weight, blood type, or ordinary everyday activities (1999). On the other hand, possible worlds are thought of as complete (Sainsbury 2010). If we try to identify a fictional character with a possible entity, we run into trouble: the details left out by the story can be filled in an infinite number of ways by the possible entities, and we would have no way of identifying some with Sherlock Holmes, for example, but not others. Secondly, most people have the intuition that fictional characters are created by an author; they come into being via certain intentional actions by an actual person. The possibilist view has no way of accounting for this. Thomasson notes: "Even if we could find a single candidate possible detective to identify as Sherlock Holmes, this would be a possible man with the property of being born in the nineteenth century, not of being created by Arthur Conan Doyle" (18). For these reasons, as well as others, both Plantinga and Kripke eventually rejected this view (Thomasson 1999; Kripke 1980). If the objections to the possible world view hold water, then it may make more sense to say that in the second case, the object of Becky's emotions is an imaginary object rather than a possible one. Furthermore, this imaginary object has motivated her behavior in the way Walton claims that genuine emotions do: the book by her bedside, the remarks to her friends.

There may be cases, then, in which our reactions to fictional stories or characters do motivate us to act. And while we do grant that it's unclear how to best interpret the precise existential commitments of motivation in terms of imaginary things, our point is that it's perfectly possible for belief-based theorists to grant that they can generate genuine emotions.[11]

---

[11] We also retain the option of simply denying that belief theorists must be committed to Walton's notion of motivation.

If that's the case, then we have enough to extend our overall critique to these sorts of pure cognitive theories as well. In other words, it is once again false that the candidates for 'exist' provided in (3–5) are the same sense of exist required by typical belief theories in order to have genuine emotion. Of course, Waltonians are free to propose their own theory of emotion under which one is forced to accept the existential commitments we have been considering. And such a theory should be evaluated on its own independent merits. But without such a theory, and by attending to this fact, we see that even for pure cognitivist accounts, there's nothing inherently paradoxical about the paradox of fiction.

5 Requiem for a Paradox

Our discussion of the three types of theories of emotion—feeling, judgment, and belief-based theories—has revealed an inconsistent use of 'exist' in the second and third premises of the paradox of fiction. According to our argument, the only way in which the paradox arises is if 'exist' denotes the same type of existence across the second and third premises. But as we have argued in Sect. 4, it is doubtful there actually is a theory of emotions advocated in the literature today committed to the same type of existence across these two premises. Furthermore, we think it's equally implausible that any theory of emotion *would* want to grant that 'exist' must be understood consistently across these premises. Of course, it is possible that such an account may be developed in the future. But this just goes to show that if the paradox of fiction does exist, it remains to be discovered or argued for—not solved.

We conclude that the formulation of the paradox of fiction, as traditionally construed by Radford and subsequent philosophers, is largely responsible for its own persistence. That said, there is no doubt that some theorists will continue to argue for or against the basic ideas represented in Radford's three sentences—it just need not be motivated by an attempt to avoid inconsistency, or to rescue people's behavior from the threat of irrationality. The novel contribution of this paper has been to reveal a crucial ambiguity in the formulation of the paradox of fiction, and subsequently, to provide a theory-neutral and minimally controversial solution by denying that it exists as a problem to be solved in the first place.

Our solution to the paradox of fiction clears the way for a cluster of interesting psychological questions regarding the nature of fictional entities, the kinds of emotional objects they represent, and the affective responses they generate as we experience them. For instance, there is a natural analogy between the affect generated when engaging with works of art and the phenomenology or perceptual seemings generated when engaging with known illusions. To illustrate this comparison, recall the feelings that many report when they experience powerful optical illusions like the Müller-Lyer illusion. Although we know that the Müller-Lyer lines are of equal lengths, we still cannot help but feel that one line is longer than the other. Similarly, our emotional responses to characters and situations do not suddenly cease after we discover that they are purely fictional. Our experiences of fear when we see the killer enter the heroine's bedroom remain after we remind ourselves that what we are watching is just a film. Knowledge that these characters

are fictional or that Müller-Lyer is an optical illusion often makes little difference to our immediate reactions in these matters.

Returning to Radford's question of rationality, note that the affect accompanying fictional responses is no more or less "irrational" than the seemings typically accompanying the standard Müller-Lyer illusion. In the case of the Müller-Lyer illusion our perceptions appear to be largely cognitively impenetrable, and independent of deliberate control. It is simply a matter of psychology and empirical cognitive science to explain why the feelings associated with the Müller-Lyer lines persist or dissipate after the illusion is discovered and the belief that both lines are of equal lengths is formed. Similarly, it is an interesting psychological question why certain kinds of affect persist after one learns of and forms the belief that objects of an emotional response are fictional. We would only emphasize that the answer to this interesting question does not bear on the central issue taken up in this essay regarding what the paradox of fiction tells us about the rationality of emotion.[12]

But having set the question of irrationality aside, the analogy between fictions and engagements with known illusions can still be instructive. With respect to fear for our heroine for instance, some have noted that our bodies are primed to respond with reactions to perceived stimuli whether they are real or not (Harris 2000) and that these feelings are instigated before and alongside the cognitive components of an emotion, such as the belief that something in our environment can cause us harm (see Schroeder and Matheson 2006). Future research on the cognitive science of emotion might fruitfully study these kinds of bodily responses for fictional and non-fictional objects and compare them to our experiences with known illusions as one of way learning more about the affective nature of these responses.

Alternatively further work on the affective responses generated by certain kinds of artistic works over others might begin to strain the basic analogy between fictions and illusions. For instance, parables and fables often evoke very complex and sophisticated affective responses connected to motivation, action and behaviour. These affective responses seem to extend far beyond the seemings generated by Müller-Lyer or the tinge of fear we feel during a horror film. No longer clouded by Radford's threat of irrationality discharged above, comparisons between the affective responses generated by various kinds of fictions over others may again tell us more about fictional objects of emotions, their relationship to other kinds of emotions and theories of emotion. We flag this as another profitable area for future research in the cognitive science of emotion motivated by our affective responses to fiction.

Given these interesting psychological research questions what is next for the philosophical literature on emotions and fiction? We close with a policy recommendation. In short—forgo solutions to the paradox of fiction. If what we have said is correct, its persistence has traded on an ambiguity. But uncovering this ambiguity has been instructive. By questioning the notion of existence throughout Radford's premises, it has led us to reconsider the different ontological commitments held by several different leading theories of emotion. We suggest that researchers interested in the paradox should simply study these commitments

---

[12] We thank an anonymous reviewer for discussion on this point.

directly, or the manner of existence that objects of emotions need have according to each theory on the whole. We suspect that this new strategy will give us a better picture of what theories of emotions can actually tell us about fictions. And, perhaps with this new focus, new progress is possible in the quest to determine whether fictions really posit a unique case in the literature or whether they warrant the same consideration as other objects of emotions.[13]

References

Buckwalter, W., Rose, D., & Turri, J. (2013). Belief through thick and thin. *Noûs*. doi:10.1111/nous.12048.

Buckwalter, & Turri, J. (Under review). In the thick of moral motivation.

Carroll, N. (1990). *The philosophy of horror; or, paradoxes of the heart*. New York: Routledge.

Coleridge, S. T. (1985/1817). Biographia Literaria. In H. J. Jackson (Ed.), *Samuel Taylor Coleridge*. Oxford.

Currie, G. (1989). *The nature of fiction*. Cambridge: Cambridge University Press.

De Sousa, R. D. (1987). *The rationality of emotion*. Cambridge: MIT Press.

Goldie, P. (2004). Emotion, feeling, and knowledge of the world. In Robert. C. Solomon (Ed.), *Thinking about feeling* (pp. 91–106). Oxford: OUP.

Harris, P. (2000). *The work of the imagination*. Malden, MA: Blackwell Publishing.

Hartz, G. (1999). How we can be moved by Anna Karenina, green slime, and a red pony. *Philosophy, 74*, 557–578.

Hurka, T. (2001). *Virtue, vice, and value*. Oxford: Oxford University Press.

James, W. (1884). What is an emotion? In C. Calhoun & R. C. Solomon (Eds.), *What is an emotion?*. New York: Oxford University Press.

Kieran, M. (2010). Emotions, art, and immorality. In P. Goldie (Ed.), *The Oxford handbook of philosophy of emotion* (pp. 681–703). New York: Oxford University Press.

LeDoux, J. (1996). *The Emotional Brain*. New York: Simon & Schuster.

Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly, 15*, 37–46.

Miller, B. (1986). Exists' and 'existence. *Review of Metaphysics, 40*, 237–270.

Nussbaum, M. (2004). Emotions as judgments of value and importance. In R. C. Solomon (Ed.), *Thinking about feeling* (pp. 183–199). Oxford: OUP.

Plantinga, A. (1974). *The nature of necessity*. Oxford: Oxford University Press.

Prinz, J. (2004a). *Gut reactions: A perceptual theory of emotion*. Oxford: Oxford University Press.

Prinz, J. (2004b). Embodied appraisals. In R. C. Solomon (Ed.), *Thinking about feeling* (pp. 44–58). Oxford: OUP.

Quine, W. V. (1953). *On what there is, in from a logical point of view* (pp. 1–19). Cambridge, MA: Harvard University Press.

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

Robinson, J. (2005). *Deeper than reason: Emotion and its role in literature, music, and art*. Oxford: Oxford University Press.

Rose, D., Buckwalter, W., & Turri, J. (Under review). When words speak louder than actions: Delusion, belief and the power of assertion.

Sainsbury, R. M. (2010). *Fictional and fictionalism*. New York: Routledge.

Salmon, N. (2002). Mythical objects. In J. K. Campbell, M. O'Rourke, & D. Shier (Eds.), *Meaning and truth: Investigations in philosophical semantics* (pp. 105–123). New York: Seven Bridges Press.

Schroeder, T., & Matheson, C. (2006). Imagination and Emotion. In S. Nichols (Ed.), *The architecture of the imagination* (pp. 19–40). Oxford: Oxford University Press.

Solomon, R. (1993). *The passions*. Indianapolis: Hackett Publishing Company Inc.

---

[13] For helpful feedback and discussion we thank Berit Brogaard, Noël Carroll, Jesse Prinz, Andy Specht and John Turri.

Solomon, R. (2004). Emotions, thoughts and feelings: Emotions as engagements with the world. In R. C. Solomon (Ed.), *Thinking about feeling* (pp. 76–90). Oxford: OUP.

Stocker, M. (1983). Psychic feelings: Their importance and irreducibility. *Australian Journal of Philosophy, 61*, 5–26.

Thomasson, A. (1999). *Fiction and metaphysics*. Cambridge: Cambridge University Press.

Thomson-Jones, K. (2008). *Aesthetics and film*. London: Continuum Press.

Tullmann, K. (In preparation). *The moral experience of fiction*. The City University of New York Graduate Center.

Turvey, M. (1997). Seeing theory: On perception and emotional response in current film theory. In R. Allen & M. Smith (Eds.), *Film theory and philosophy* (pp. 431–457). Oxford: Oxford University Press.

Van Inwagen, P. (1977). Creatures of fiction. *American Philosophical Quarterly, 14*, 299–308.

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

Walton, K. (1993). *Mimesis as make-believe*. Cambridge, MA: Harvard University Press.

Williams, C. J. F. (1981). *What is existence?*. Oxford: Oxford University Press.

Zalta, E. (1988). *Intensional logic and the metaphysics of intentionality*. Cambridge, MA: MIT Press.