

Make-believe and the paradox of fiction

Simon Marcus

April 2010

There are several well-known paradoxes which concern the relationship between our emotions and art: how do we explain the emotions felt towards non-representational art like music, and what is the object of our emotions in such cases? How do we explain our simultaneous enjoyment and despair towards fictional tragedies and horrors? How do we explain the depth of feelings we experience towards characters we know to be fictional? It is with this last question, “the paradox of fiction,”¹ that I will be concerned in this paper. I will begin by explicating the problem roughly as presented in Radford’s seminal paper,² as it functions as a neat rubric through which to assess the proposed solutions. Thereafter I will provide reasons to endorse a solution which favors a possible-worlds analysis of fiction, and which analyzes the cognitive component to our response in terms of make-believe.

The paradox of fiction

Suppose Smith tells you about his sister, who, according to Smith, was recently brutally attacked in her home and kidnapped over the course of a week. You will presumably respond affectively; you will pity the woman for the ordeal she suffered, feel resentment towards her captors, and some relief that she is now safe. If Smith then revealed that he was having you on, that the story was false, your feelings for the woman would vanish – all you would feel would be some outrage towards Smith for presenting this disturbing tale as if it were true. This scenario reveals certain things about the nature of emotions. First, emotions are characteristically felt states – they have a phenomenology. Second, emotions have an object at which they are directed, and a corresponding belief about the existence and nature of that object. We are not harrowed by Smith’s story if we know that the woman doesn’t exist, and doesn’t really suffer. It might be objected that we could be moved even if Smith told us ahead of time that the story was fictional. If we reflect on the plight of the character, we might be moved. However, if we examine that case more closely, we see that we bear in mind

¹ (Levinson 1997, 21)

² I am indebted mainly for the structure (though not entirely for the content) of my explication to Radford (Radford 1975).

actual suffering: we are considering suffering that we *believe to be actual* when we pity her.

The suggestion is that the emotion of pity seems appropriate only in circumstances of actual suffering. To reinforce this point, suppose you and I attend a theatre performance in which an actor performed a sequence of a woman suffering. If I were to lean over to you and say, 'You know, I do pity that woman,' it seems I might plausibly mean one of two things. (1) I was referring to the fictional character. In that case you would probably consider my sentence unintelligible, and point out that I shouldn't worry since it's only a show, that it's not real. You might think me a little unbalanced (or under-distanced), and be concerned that I had not grasped how fiction diverges from reality. (2) I was referring to someone real who shares many qualities with the fictional character. If the character played, say, a Jew in Nazi Germany, you would understand me to be making an indirect statement about the plight of that Jew and others who *actually* suffered. Unlike (1), my remark would be intelligible in (2), since there was a suitable object for my pity – some individual who is really pitiable, of which the fictional character is a powerful reminder. (For this reason, historical fictions do not generally fall prey to the paradox of fiction.)

However, the paradox of fiction is that we frequently *do* seem to pity characters whom we believe to be fictional, *while* we believe them to be fictional, who bear no obvious similarity to historical characters, and where the emotion seems directed *at* the fictional character (that is, not at some other person who the fictional character represents). We lament that Mercutio is so mercurial, and we wish that he did not act so rashly, even when we have seen the performance a hundred times before. Our emotions are *for him*: we wish that *he* did not bait Tybalt so, and we weep for *his* death – and not for anyone else's. Generally speaking, our emotional response is directed at the fictional character, and not some actual individual of whom the fictional character reminds us: "despising a fictional character, for instance, is not simply reducible to despising people of that sort generally, or to despising some actual similar individual of one's acquaintance."³ Thus, the paradox of fiction is made most explicit as an inconsistent triad of propositions:

³ (Levinson 1997, 23)

Where Jones is an actual person, and X is a fictional character,

- a) To feel pity for X, Jones must believe that X suffers
- b) Jones feels pity for X
- c) Jones doesn't believe that X suffers, since Jones knows that X doesn't exist

Though it comes to the same thing, it is helpful to see (a) as the expression of a conditional sentence, 'Jones pities X \rightarrow Jones believes that X suffers.' Then we have the affirmation of the antecedent of the conditional in (b), and the denial of the consequent in (c). Obviously then, at least one of the propositions (as presently formulated) has to go.

Resolving the paradox: some options

Now, it is clear that our responses to fiction are not entirely identical to our responses to fact. When we 'grieve' Mercutio's death, we might at the same time scoff popcorn or remark, 'My, he is brilliant.' This certainly diverges from how we grieve a real person's death, and diverges to a degree that might make us suppose that this we do not grieve at all. This might appear to dissolve the paradox by denying (b). However, Radford notes that this merely defers the problem, as we still need to account for whatever it is we do feel, which appears to be *in response to*, and *directed at*, fictional characters whose suffering brought not a drop of pain into the world. The emotion might not be grief, but there is some phenomenon – one which seems just like emotion – which requires explanation. Radford examines several unsuccessful ways of explaining this phenomenon, which I will address in turn.

1. We are temporarily deluded. *Perhaps we are temporarily deluded into forgetting that the story is fictional, so we respond as if to real characters (since that is what we, in our deluded state, take them to be). In other words, we get so 'caught up' in the action, that we forget that the action is fictional.* This is, however, deeply unsatisfactory since it is inaccurate in many respects. Firstly, it does not properly respect the fact that we are just never so deluded. While watching Mercutio die, we are never under the *illusion* that he has actually died. Indeed, if we really believed him to be dying, we would be guilty of a moral failing when we did not rush to assist him. Indeed, we have a much more robust sense of what is real and what is fictional, and we retain this distinction even when we respond affectively to fictions.

2. We suspend our disbelief. *In theatre and cinema, props and technical devices are employed to draw the viewer into the story. Actors and scripts are evaluated according to how realistic they are. Perhaps we are not deluded wholesale into believing that the events in the fiction really occur, but maybe we suspend our disbelief in their falsity.* However, like the illusion explanation, this explanation errs in inadequately characterizing the phenomenon. As Levinson puts it, the notion of suspension of disbelief “unacceptably depicts consumers of fiction as having both a rather tenuous grip on reality and an amazing ability to manipulate their beliefs at will.”⁴ We are generally *never* unconvinced of the fictional nature of the story – often we respond emotionally to subject matter which wears its fictional nature on its sleeve, like the Simpsons or Star Wars. Neither do we have the power over our belief-forming mechanisms to switch from disbelief to belief in the way that this explanation requires. Moreover, this explanation does not account for how it is we may respond emotionally to content which contains metaphysical or logical falsehoods (i.e., things which we could not or would not know how to believe), such as time-travel or magic.

3. It’s just a brute fact. *Perhaps it is just a brute fact that people can be moved by fictional stories and characters.* This response does no theoretical work, and doesn’t respect the depth of the difficulty of the paradox. No one denies that it *seems* to us that we respond emotionally to fiction; the question is whether, given the absence of belief in the existence of the fictional characters, our emotions can be made intelligible.

4. We are not merely moved by actual occurrences, but by the prospect of such occurrences. Suppose that while in class you daydream a little, and get to thinking about some friend of yours, Susan, who lives abroad and who is coming to visit you soon. Since you have a propensity to fear the worst in even the most benign of circumstances, you proceed to think about how your friend’s plane might crash. You think then of the child that Susan would leave orphaned if she died in the crash. This is too much for you, and your eyes well with tears. In this case, at least, such circumstances are *intelligible*. We can explain how it is you have had this emotional response by making reference not to actual suffering, but to the prospect of actual suffering. Furthermore, it seems that your response is more plausible if the case is more likely. If you were daydreaming about Susan, and started to worry that her flight would be hijacked by Martian separatists who crave the tasty flesh of young mothers, and you were moved to tears when

⁴ (Levinson 1997, 23)

considering her orphaned child, we would think you increasingly deluded, and your reaction would be increasingly puzzling. Perhaps something like this goes on with fiction: we have in mind the prospect of the suffering without the suffering itself, and it is at this prospect that our response is directed. However, this involves the belief that the suffering is *actual* – it is the prospect of *actual* suffering – and so it does not run afoul of the paradox of fiction. The paradox remains since we pity those fictional characters whose suffering we believe to be wholly fictional.

5. When we weep for Anna Karenina, we weep for the pain and anguish that a real person might suffer and which real persons have suffered. Perhaps our reactions to fiction are in this way derivative on our reactions to reality, so that when we appear to respond to the pain in a fiction, we are in fact responding to the actual pain in the world, of which the fictional case is a reminder. However, this is simply not true of our experience. When we weep for Anna, we do not weep for those in her circumstances - we weep for *her*.

On the strength of these objections (and one further worry, which I will address later), Radford concludes that our apparent emotional response to fiction is unintelligible. Indeed, it must be granted that he has canvassed a wide range of seemingly plausible explanations, all of which are deficient in some or other respect. However, I wish now to present and defend an alternative explanation not addressed by Radford, according to which the phenomenon is explicable in terms of *make-believe*.⁵ Before explaining how my account is to proffer a satisfactory answer to the paradox of fiction, it is necessary first to elucidate the theoretical framework upon which the solution is based.

Make believe and fiction

Consider some schoolchildren who are playing a game during their lunch break. The game works where one designated child starts out as the wolf, who tries to catch and eat the other children, who are sheep. The sheep do not wish to be caught because they will turn into wolves too, and will then have to chase the other sheep. This game is easy enough to follow, and, important for our purposes, it is a game of *make-believe*. There are a set of norms that govern the game, based on the true propositions of the make-believe world of the game. For example, if you participate in this game you know that, according to the game,

⁵ This theory draws liberally from the work of Walton (1997) as well as Gaut (2007) and Davies (2009), though much of its formulation and defence is my own.

you are either a sheep or a wolf. As a sheep, you wish to avoid being eaten. You also know that you will transform into a wolf immediately if you are eaten. Those children who are not participating in the game are neither sheep nor wolves, but become sheep as soon as they are permitted to play. There are also truths which merge components of the real and game world. For example, those children who are fast-running children are also fast-running sheep or fast-running wolves, and only big children can be big wolves. However, certain truths in the actual world do not transcend into the game world, and vice versa. For example, those children who are excellent at math are not wolves who are excellent at math (though, cruel as schoolchildren are, it is likely that ugly children make ugly sheep too). The bounds of the game seem quite neatly demarcated. As soon as the break time is over, the game ends, everyone who was a sheep or wolf transforms back into a child. It would not 'count' if a child who was the wolf touched a child who was a sheep after break time – that touching would not make it true in the world of the game that a wolf had eaten a sheep.

The suggestion here is that games help us to understand fictions, since both trade in make-believe. In both the game and in fiction we are presented with alternative worlds which posit alternative norms, laws, and truths. We desire a rich experience of the game, as we do with fiction, and the possibility of meaningful engagement is achieved through our efforts at make-believe. As Friend observes, "works of fiction are designed to prescribe imaginings about their content, and imagining what is prescribed is participating in the game of make-believe authorized by the work."⁶

It will do to spell out the role of props which has been central to Walton's theory of make-believe. According to Walton, "pictures are props in visual games of make-believe."⁷ In this sense, paintings, films, and novels are props like a child's paper airplane, insofar as they are *tools which facilitate our make-believe in the world of the fiction*. The props may be complex or simple, though our make-believe in complex fictions usually requires sophisticated props. In the case of a novel, our capacity to be 'drawn in' to the scene will depend on the richness of the descriptions and imagery, those props in the novelist's toolkit. When watching a film, we will find make-believe easier if the cinematography 'gives away' as little of its fictional nature as possible. Our capacity for make-believe, it seems to me, is dependent largely on the plausibility of the presented fiction *as*

⁶ (Friend, *Imagining Fact and Fiction* 2008, 152)

⁷ (Walton, *Marvelous Images: On Values and the Arts* 2008, 63)

a coherent world. Now, this is different from its plausibility as *our* world, and this difference explains why we may be drawn into fictions which postulate worlds quite unlike our own. Part of what makes, say, Tolkien's trilogy so compelling is the metaphysical coherence of Middle Earth. Certainly, we do not believe the contents of the fiction to be true of our world – for starters, all sorts of difficult *local* issues arise if we tried actually to believe that, like whether wraiths are material or immaterial, or whether Mordor is to the west or east of Japan. However, notwithstanding certain trans-world incommensurables, we observe a rich *internal* coherence to the fiction, which makes it easy to make-believe in the story.

This explanation also deals neatly with occasions in which make-believe proves difficult; they are those occasions where we evince criticisms such as 'I just didn't find the story believable or realistic.' How might we understand the terms 'believable' or 'realistic' in cases of science fiction, where often the beliefs in question could not possibly be held of our world, and where the subject matter makes the notion of 'realistic' representation absurd? The remarks are, I suggest, to be understood as *immanent* comments – a reflection of matters local to, or in the context of, the fiction. When someone says that a science fiction story is not really 'believable', he does not criticize according to the actual metaphysics which govern our world, but rather according to the principles and truths in the world of the fiction. For example, a competent critic will not say that Darth Vader's light sabre would be prohibited under international law; rather, he will say perhaps that Vader's modest choice of arms is at odds with his megalomaniac character *in the films*.

There is a natural correlation here between David Lewis' modal realism⁸ and how we talk about fiction: possible worlds are powerful devices which make perspicuous the complex modal features of the matter. Though I won't present his account in any great detail, it will suffice to note that Lewis maintains that each possible world is constituted by the propositions which are true in that world. More controversially, he maintains that each world is just 'as real' as ours; the terms 'real' or 'actual' are indexical, and they merely denote that world of which the speaker is part. Suppose that McCain, bitter about his electoral defeat remarks, 'I could have won.' This is to say that it is false in the actual world that McCain won, but there is some possible world in which it is true McCain won.

⁸ (Lewis 1986)

In *that* world, it is not merely *possible* that McCain won, it is *actual*, and, relative to that world, our world is a possible world in which McCain lost.

While I do not wish to defend or deny Lewis's modal realism here, I do wish to contend that this model provides a helpful method for analyzing what it is we believe when we make-believe in fiction. What do we believe when we make-believe that Anna Karenina suffers? The simple answer is that we actually believe that Anna suffers *fictionally*, that is, she suffers in the world of the fiction. We do not believe she suffers actually (in our world), since we do not think she exists actually. Yet, as the preceding discussion shows, we might indeed put this another way: instead of saying that 'I actually believe that Anna suffers fictionally,' we might just as well say that 'fictionally, I believe that Anna actually suffers.' This is to say, counterfactually, that if I were an inhabitant of the world of the fiction, I would believe that Anna suffers actually.

Now, this explanation makes clearer the differences between considering an arbitrary possible world and considering a world of fiction, and it mirrors the distinction between our imaginative response to fact and our imaginative response to fiction. A journalist wishes us to imagine the content of her report (constructing a mental image of the content), and to believe the report to be true of the actual world. A novelist, on the other hand, wishes us to imagine the content of her story, and to make-believe it. There are norms which govern the formation of our beliefs, the most natural of which is 'believe *p* if and only if *p* is true.' There is a corresponding norm in the realm of fiction: 'make believe *p* if and only if *p* is true according to the fiction.' However, it is clear that when we engage with fiction, we import a great deal of propositions which are not explicitly stated by the purveyor of that fiction. It is never said explicitly that Mercutio has two arms rather than three, but since no mention of any such peculiarity is made, and since humans typically have two arms, and *since this proposition is commensurate with and not contradicted by the fiction*, I may import it. When as a child I read *To Kill a Mockingbird* for the first time, I thought for the duration of a chapter or two that Scout Finch was a boy, since I thought the name was a boy's sort of name. However, after the other characters used the feminine pronoun directed at Scout a few times, I revised my incorrectly imported belief. I did not revise my imported belief that Scout was a bipedal human who breathed oxygen, since that was consistent with the fiction, and, indeed, a belief required of me if I were properly to understand the story. Indeed, "the metaphor of 'worlds' emphasizes the sense in which fictional truth

goes well beyond what is explicitly stated.”⁹ The intention with fiction is to take the information given about the characters and events as small portions of a much grander metaphysical structure – the fictional world – of which the characters are a part.

We are now in a position to see how this account of make-believe in the fictional world explains emotional response to fiction. In order properly to attend to fiction, we must make-believe the propositions contained in the fiction. This involves a degree of effort from us, as we galvanize those propositions presented explicitly in the fiction with those tacit assumptions that we bring to the fiction which are required in order to make it coherent. When we make-believe, we participate in the construction of the fiction. Sainsbury observes that “much of what happens in a reader’s response to a fiction constitutes an extension of that fiction, a way of making things to be so according to an extended version of the story.”¹⁰ I find the notion of *our extension* of fiction to be a highly instructive way of understanding the relation between emotion and fiction. In extending a fiction, we often must import not only truths about our world such as ‘humans are generally bipedal’ but also a system of *emotional inference*. As Walton avers, “we bring much of our actual selves, our real-life beliefs and attitudes and personalities, to our imaginative experiences.”¹¹ Part of successfully make-believing in a fiction is interpreting the psychological phenomena of the story, such as what motivates characters to act as they do, or whether a character is remorseful or vindictive. Correct interpretation is achieved through importing our actual normative emotional assessment mechanisms, as far as they cohere with the story. Indeed, this procedure is assumed by the authors of the fictions – one might be said to have missed the point of the story if one failed to make these emotional assessments. In *To Kill a Mockingbird*, we take Atticus Finch to be both pitiable and admirable, for he fights for the freedom of the innocent within an unjust legal system; we import a set of moral values and normative emotional responses consistent with the fiction which affirm his pitiable and admirable character. *Fictionally, we feel pity and admiration for Atticus*. In this sense, our emotion is fictional – *it is actual emotion within the fictional world, and it is fictional emotion within the actual world*.

⁹ (Friend, *Imagining Fact and Fiction* 2008, 155)

¹⁰ (Sainsbury 2010, 15)

¹¹ (Walton, *Spelunking, Simulation, and Slime: On Being Moved by Fiction* 1997, 38)

It is clear that fictional emotions are like emotions in some sense, but are not like emotions in others; it is my suggestion that their similarities and differences go to the origin of the paradox of fiction as well as its solution. Emotions, as mentioned, are intentional states in that they are characteristically felt states which have an object at which they are directed. Let us call fictional emotions 'F-emotions'. F-emotions differ from emotions in their cognitive component and thus in their associated actions. The cognitive component of an emotion is a belief about the object of one's emotion. In the case of fear, the required belief is that the object is dangerous and is a threat to you (and, obviously, in order for you to believe it a threat, you must believe that the object actually exists). In the case of F-fear (fictional fear), one must *make-believe* that the object is dangerous and is a threat to you – that is, in the fiction, one would believe that the object is actually dangerous and a threat to you. Emotions differ from F-emotions in their associated actions. If you believe that the slime is going to swallow you, then you will avoid the slime. However, if you make-believe that the slime is going to swallow you, then you will fictionally avoid the slime. Notwithstanding these differences, the *phenomenology* of emotions and F-emotions may coincide perfectly. One may feel from the inside quite the same towards fictional events or characters as one does towards actual events and characters.

This distinction explains and removes many of the worrisome aspects to the paradox of fiction. First, it avoids the logical contradiction of the paradox by denying the antecedent of the conditional, 'Jones pities X \rightarrow Jones believes that X suffers.' According to the theory articulated here, Jones does not actually pity X. Rather, Jones F-pities X, that is, Jones pities X in the world of the fiction, and Jones believes that X suffers in the world of the fiction. As mentioned, F-pity might *feel* just like pity. This, I argue, is how the paradox arose: if we attend *just to the phenomenology* of the matter, we might be misled into thinking that we are responding emotionally to fictional characters, of whose existence we do not have the appropriate belief. Since emotions and F-emotions may feel the same from the inside, the feeling alone cannot classify the phenomenon. In other words, the phenomenology is neutral between emotion and F-emotion. However, examination of the cognitive aspects to the phenomenon reveals its true character, and explains the divergent associated actions. If we recognize that *we cannot, as a matter of definitional fact, feel emotions for fictional characters* (though we may experience a sensation which feels internally indistinguishable), then we are no longer confused when the behavior associated with emotion diverges from behavior associated with F-emotion. If we believed that the killer slime were actual, we would avoid the slime actually. Since we do not believe that

it is actual, how may we explain the fact that we may well squirm in our seats though refrain from fleeing the cinema? According to the foregoing theory, the squirming is a kind of phenomenological residue, the evidence of having successfully make-believed in the fiction. It is not evidence that we *believe* that the slime is real, but evidence only that make-believing in fiction may produce in us phenomenologically indistinguishable internal states.

Concluding remarks: Radford's original problem and the strength of the solution

Earlier, I presented five purported solutions discussed and rejected by Radford. His sixth, however, was closest to the mark, though the option he discusses has problems which mine avoids. He asks whether the paradox of fiction might be resolved by rejecting the same proposition, denying that Jones pities Anna. Perhaps, Radford wonders, we do not, after all, feel pity for Anna Karenina. Since the paradox reveals difficulties in our having intentional states towards fictional characters, perhaps our responses are *non-intentional*. Perhaps they are more like moods – states which have a psychological and felt character, but which do not have intentional objects. Indeed, if we pity, there must be an object which is pitiable; so maybe our response to fiction is not pity after all, and it is some mood-like state which has a qualitative character without an intentional object. Since the response is not really pity, it explains the differences in our actions: we do not grieve for Anna Karenina in quite the way we would for a friend. In these respects, my solution is very similar. Radford rejects this option, however, for reasons which illustrate the strength of my account over the one in question. He rejects it on grounds that it explains the differences in the associated actions, but cannot explain their *similarities*; if the states are not emotions, why do they *feel* to us often to be just like emotions? I hope to have shown that the make-believe solution, according to which the phenomenology of our response to fiction may coincide exactly with that of fact, explains the *appearance* of emotional response to fiction as well as the range of associated actions.

I also hope it is clear that Carroll is hasty in criticizing the make-believe solution, when he writes that the key objection to the make-believe solution “is that it relegates our emotional responses to fiction to the realm of make-believe.”¹² Indeed, the account defended in this essay has as its aim the understanding and proper classification of emotional responses to fiction. Insofar as my account might be considered skeptical about emotional responses to fiction, it is skeptical

¹² (Carroll 1990)

only as regards the *nature* of the phenomenon, not as regards its existence. It is my contention that recognizing the role of make-believe in our response to fiction dissolves the paradox in a way which satisfies our intuitions about the object-directed nature of emotions and the phenomenology of response to fictional characters. It achieves this through advancing a cogent modal distinction between those propositions which are true actually, and those which are true fictionally.

Bibliography

- Carroll, Noël. *The Philosophy of Horror*. New York: Routledge, 1990.
- Davies, Stephen. "Responding Emotionally to Fictions." *The Journal of Aesthetics and Art Criticism* 67, no. 3 (2009): 269-284.
- Feagin, Susan L. "Imagining Emotions and Appreciating Fiction." In *Emotion and the Arts*, edited by Mette Hjort and Sue Laver. Oxford: Oxford University Press, 1997.
- Friend, Stacie. "How I really feel about JFK." In *Imagination, Philosophy and the Arts*, edited by Matthew Kieran and Dominic McIver Lopes. New York: Routledge, 2005.
- Friend, Stacie. "Imagining Fact and Fiction." In *New Waves in Aesthetics*, edited by Kathleen Stock and Katherine Thomson-Jones. London: Palgrave Macmillan, 2008.
- Gaut, Berys. *Art, Emotion and Ethics*. Oxford: Oxford University Press, 2007.
- Gaut, Berys. "Reasons, emotions, and fictions." In *Imagination, Philosophy and the Arts*, edited by Matthew Kieran and Dominic McIver Lopes. New York: Routledge, 2005.
- Levinson, Jerrold. "Emotion in Response to Art: A Survey of the Terrain." In *Emotion and the Arts*, edited by Mette Hjort and Sue Laver. Oxford: Oxford University Press, 1997.
- Lewis, David. *On the Plurality of Worlds*. Oxford: Blackwell, 1986.
- Livingston, Paisley, and Alfred R. Mele. "Evaluating Emotional Responses to Fiction." In *Emotion and the Arts*, edited by Mette Hjort and Sue Laver. Oxford: Oxford University Press, 1997.
- Radford, Colin. "How Can We Be Moved By The Fate of Anna Karenina?" *Proceedings of the Aristotelian Society Suppl. Vol. 49* (1975): 67-93.
- Sainsbury, R. M. *Fiction and Fictionalism*. Oxford: Routledge, 2010.
- Walton, Kendall. *Marvelous Images: On Values and the Arts*. Oxford: Oxford University Press, 2008.
- Walton, Kendall. "Spelunking, Simulation, and Slime: On Being Moved by Fiction." In *Emotion and the Arts*, edited by Mette Hjort and Sue Laver. Oxford: Oxford University Press, 1997.