Aporia vol. 35 no. 2–2025

The Tall Tale of the Paradox of Fiction: Diagnosing and Dissolving the Problem of Emotional Responses to Fiction

Conor Jedam

In this paper, I explore a metaphysical assumption underlying the paradox of fiction and its solutions. Furthermore, I show this assumption, namely that there is a distinction between mind-dependent, fictional states of affairs, and mind-independent reality, to be a problem which undermines the paradox. In the first section, I provide a brief history of the philosophical discussion regarding the relationship between art and emotion before describing the paradox itself, the standard moves philosophers have suggested to resolve the paradox, and sketch some of the most notable constructive solutions. Importantly, I also draw out the metaphysics suggested by the paradox's requirement that emotion about some entity implies belief in that entity's existence. In the second section, I undermine the distinction between fictional and non-fictional entities, and the related Augustinian picture of language, to show that it is ill-equipped to deal with the question of fiction. I argue that instead, Wittgenstein's use-theory of meaning and language-games, which do not make metaphysical commitments, make sense of the paradox of fiction by accounting for the various contexts in which emotional responses occur. If we accept Wittgenstein's account of language use, then the paradox dissolves, in part because

Conor Jedam recently graduated from The University of Queensland with a Bachelor of Arts with Honours Class I in the Field of Philosophy. His thesis explores how political agency and ethical subjectivity are promoted by imagining the future, particularly in the form of dystopian cinema, through proleptic engagement with fear. He is looking forward to working and researching at UQ in 2026 and pursuing a PhD in the future.

Radford does not consider the importance of language-games in mediating the relationship between the fictional and non-fictional. Additionally, by uncoupling meaning and metaphysics, I cast doubt upon the requirement that the paradox of fiction places on existence for genuine emotional responses. The paper follows a distinctly quietist approach by first diagnosing the problem and coming to an understanding of the terms of the debate, followed by the dissolution of the problem by calling into question the grounds upon which the debate sits.[1]

1.0 Diagnosis

1.1 The Philosophical History of the Relationship between Art and Emotion

Philosophical discussion of the relationship between art and emotion has a long history, stretching back to Aristotle. In the *Poetics*, Aristotle claimed that in order to write a tragedy or epic poem which truly elicits emotional responses such as fear and pity in its audience, the author must provide a *mimesis*, or imitation, of a situation with serious ethical stakes. In the context of these artistic forms, the audience not only experiences fear and pity, but also a *katharsis* of both (Freeman 249). In other words, one function of tragedy and epic poetry is to allow the audience to engage with emotions and experience a sense of relief from them, without being placed in harm's way (Graham 36). More than two-thousand years later, David Hume also considered why we enjoy tragic fiction and developed a psychological notion of *katharsis*. Hume suggested that we derive pleasure from experiencing emotions like fear and pity in response to fiction (Freeman 249). In 1975, Colin Radford considered this puzzling relationship and proposed the paradox of fiction.

1.2 The Paradox of Fiction

In, "How Can We Be Moved by the Fate of Anna Karenina?," Radford posits three highly intuitive and seemingly plausible premises. According to Radford, when these premises are taken together, they demonstrate that emotional responses to fiction are irrational (Freeman 249). That is,

1 Thomas J. Spiegel, "What is Philosophical Quietism (Wittgensteinian or Otherwise)?," in *Quietism, Agnosticism and Mysticism: Mapping the Philosophical Discourse of the East and the West*, ed. Krishna Mani Pathak, Springer, 2021, p. 22.

having an emotional response to a fictional entity when one is aware of their status as fictional, "involves us in inconsistency and so incoherence," because the emotional response itself implies belief that the entity exists non-fictionally (Radford and Weston 78). While the rational actor would cease their emotional response upon learning that the entity is fictional, many of us do not, and even have emotional responses to entities we know are fictional from the outset (68). The first premise of the paradox states that we do, in fact, have emotional responses to fictional entities (71). We weep at tragic deaths, cheer at moments of victory, and shudder in fright when we see the killer approaching an unwitting suspect. The second premise raises the point that we do not believe that fictional entities exist (70–1). We know that nobody really suffered a tragic death, we know the victory is made up, and we know that the killer poses no real harm. The third premise states that genuine emotional responses to fictional entities imply that we believe such entities exist (68).

To substantiate this final claim Radford provides an example. Imagine you read an account of a group of people who are suffering an awful situation. Given your humanity, you will be moved in some way by learning about this horrible state of affairs (68). After learning about the plight of these people you may even begin to grieve. Then, you discover the story is a complete fabrication. Radford suggests you could no longer continue to grieve since it would be irrational (68). From this, Radford concludes that we may only be emotionally moved by the plight of others if we believe something appropriate has happened to them. In this sense, an emotional response to any entity, implies belief in said entity (68). So, for Radford it does not make sense to have emotional responses to fiction and it is a genuine paradox (Freeman 249).

1.3 Rejecting the Paradox

Of course, many philosophers in the following years have attempted to break the paradox of fiction, usually by rejecting at least one of its premises. As I am adopting a quietist approach to philosophical inquiry, it is important that prior to dissolving the paradox of fiction I am familiar with the terms of the debate which surrounds it. This allows for an understanding of the assumptions upon which the debate takes place. It is not to any particular response to the paradox that I critically respond, but these underlying assumptions which I find to be problematic, and which lead to the paradox's dissolution in section two.

Kendall Walton suggests that emotional responses to fiction are not genuine and replaces them with quasi-emotions, which are

phenomenologically identical to ordinary emotions, but do not require belief in the existence of the object of the quasi-emotion (Freeman 249). For Walton, this is what it means to engage in make-believe (Tullmann and Buckwalter 781). In order to reject the second premise, another group of thinkers have evoked Samuel Taylor Coleridge, who posited that when engaging with fiction we willingly suspend our disbelief, and in doing so can temporarily hold belief in the existence of objects we would, under regular circumstances, regard as non-existent (782).

Two branches have emerged which reject the final premise, which suggests that genuine emotional responses to fictional entities imply our belief in their existence (782–83). Those who hold a non-cognitivist theory of emotion can simply reject the notion that emotion requires belief (782). On the other hand, there are those who maintain that emotions have a cognitive aspect, but reject that this aspect involves belief (783). A string of thinkers, including Peter Lamarque, Noël Carroll and Murray Smith hold this view. While they accept that belief is required for emotional responses to non-fictional entities, they suggest that those who consider this a paradigmatic feature of emotional responses in general are mistaken (Freeman 248). According to this group, we might imaginatively propose, entertain in thought, or mentally represent the existence of fictional objects (248). The common trait among all these solutions is that they are constructive, which is to say that they build some system, or framework, which attempts to adequately capture the relationship between emotion and fiction (Spiegel 220).

1.4 Other Responses

Rejecting at least one premise is not the only route philosophers have taken in response to Radford's paradox. Michael Weston's response, originally published alongside Radford's paper, took issue with the claim that it is a brute fact about humans that we can have emotions in response to fictional entities (Radford and Weston 81). Instead, Weston suggests that emotional responses to fictional entities are actually emotional responses to works of art (81). In this sense, what we are responding to is not fictional at all, but rather an art object in the world.

Although my primary goal in this paper is to dissolve the paradox, my own solution in this tradition is to call into question Radford's appeal to the notion of irrationality. It is an unnecessarily strong claim to suggest that emotions divorced from belief are irrational, in the sense that they are not logically coherent. Given this framework, it is the case that we have irrational emotions about non-fictional entities as a matter of course. This

is what it is to have a phobia, or to fear our own death even if we believe the experience of death to be equivalent to dreamless sleep (78–9). My suggestion is that it makes sense to talk about whether emotional responses are justified, rather than rational, since the notion of justification can capture a greater degree of complexity in the way we give, compare and weigh the reasons for our emotional states.

More recently, Katherine Tullmann and Wesley Buckwalter have argued, with reference to theories of emotion, that the paradox falls apart once we begin to consider the various ways in which the word 'exist' can be used (Tullmann and Buckwalter 784-5). Since existence can be taken to mean either existing as a concrete object, having the potential to exist as a concrete object in some possible world, or existing as an imaginary object, and no theory of emotion makes the use of existence uniform between the paradox's second and third premises, Tullmann and Buckwalter conclude there is no such thing as the paradox of fiction (793). They are right to consider the limitations of the language required to make sense of and criticise the paradox. In the following sections, I delve further into this notion of existence and, without reference to theories of emotion, dissolve the paradox of fiction by calling into question the basic metaphysical picture it posits.

1.5 The Metaphysical Picture Underpinning the Paradox

Although Radford does not explicitly qualify which picture, or pictures, of existence he is committed to, clarifying the distinction between fictional and non-fictional entities is crucial to making sense of his paradox. While Tullmann and Buckwalter suggest that Radford uses existence in a general rather than technical sense, the paradox itself rests upon the dichotomy between those things which really exist, the non-fictional, and those things which do not really exist, the fictional (784). Furthermore, we should not take for granted that this distinction is undisputable or necessary (Matravers 96). Let us return to the paradox itself to further tease out Radford's commitments.

For Radford, emotional responses to entities imply belief in the existence of these entities. In other words, it is only rational for me to recoil in my seat during a screening of John Carpenter's 1978 film, *Halloween*, if I believe that the masked killer, Michael Myers, poses a real threat to me. Of course, Michael Myers can only be a real threat if he is a real person that exists out there in the world. This description, extrapolated from Radford's paper, is indicative of a commitment to the notion that those things which exist, do so in a mind-independent reality. That is, whether I have it in

mind or not, Stonehenge really exists on Salisbury Plain. Stonehenge is actual and I can have rational emotional responses to it.

On the other hand, Michael Myers cannot be a real threat if he is merely made up, imagined, the stuff of fiction. There may exist pictures of Michael Myers projected onto the cinema screen, and his name may appear written in the screenplay, but these are not instances of Michael Myers himself. These concrete signifiers of Myers are the result of his first being imagined. That is to say that Michael Myers is mind-dependent, and so too is the entire fictional reality, the town of Haddonfield, Illinois, in which he goes about his murdering. It is nonsensical to say Haddonfield exists independently of whatever mind conceived of it. Furthermore, when Haddonfield is thought about, it does not suddenly appear in the physical world such that one could measure the distance between it and Stonehenge. This dualism made up of real, mind-independent, existing things on one hand, and fictional, mind-dependent, non-existent things on the other, is the basic metaphysical assumption underpinning Radford's paradox of fiction.

2.0 Dissolution

2.1 Augustinian Meaning

The metaphysical commitment implied by the distinction between the fictional and non-fictional raises questions regarding the use of language. In particular, we must investigate what it is we are referring to when we refer to either fictional or non-fictional entities. In the case of non-fictional entities, it is consistent with Radford's position, that we refer to things *out there* in the world (McNally 8). That is, there is a worldly object to which I am referring when I say, 'Stonehenge is on Salisbury Plain.' The arrangement of rocks that make up Stonehenge and the geographical location of Salisbury Plain give meaning to my statement. In another way, the statement makes sense because we know what it describes (8).

Conversely, we can question what it is we refer to when we refer to fictional entities. When I say, 'Michael Myers is murdering people in Haddonfield,' there is nothing in the world which stands for the terms 'Michael Myers' or 'Haddonfield.' So, given the picture we are working with for now, this statement does not carry meaning. Importantly, it is different from the statement, 'In the film *Halloween*, the character of Michael Myers murders people, and the film is set in a place called Haddonfield.' In this case, the object *in the world*, to which I am referring is the film,

Halloween. In other words, we can make sense of this statement, because we know what it describes. This referential way of accounting for meaning in language is attributed by Ludwig Wittgenstein to Augustine of Hippo (Wittgenstein 1). Like Wittgenstein, I find that the Augustinian picture of language lacks the tools to properly account for language use. In particular, it is ill-equipped to deal with the complex ways in which we emotionally engage with fictional entities. It is not possible to account for the grief, rage, and fear which give rise to Radford's paradox while restricting oneself to an Augustinian account of language.

2.2 Meaning as Use

For Wittgenstein, the Augustinian picture of language is simply too limited in its applicability to actual language use (McNally 13 and Addis 97). If referential language use is predicated on the notions that words name objects and sentences are made up of these words, then the meaning of a word is the object for which it stands (McNally 9). In this sense, words are connected mentally with objects, and to understand a sentence is to know what it describes (9). We can explain the meaning of words by ostensive definition, that is by gesturing towards the referent of a word as we say it (Wittgenstein 27 and McNally 13). For example, I stand on Salisbury Plain, point at the arrangement of rocks and say, 'That is Stonehenge.' For Wittgenstein, this is the historically pervasive understanding of how language works (Addis 97). For me, it is clearly an assumption in Radford's paradox since emotional responses are only rational when they are associated with non-fictional entities. In the same way that ostensive definitions are made by referring to an object in the world, for Radford emotions are only rational when they exist in response to something non-fictional.

However, to suppose that words have meaning insofar as they correspond with objects is to impose a mistaken functional uniformity. Wittgenstein rejects this uniformity and claims instead that words have a variety of functions in use, and carry different meanings in different contexts (Wittgenstein 23). Like tools, we can use words in diverse and complicated ways for different purposes (11). So, given the complexity of language and its uses, we must examine how it is that most of the time we are not prone to errors of confusion and ambiguity. How is it we still engage in meaningful language use? According to Wittgenstein, "for a *large* class of cases… the meaning of a word is its use in language" (43). It is not the case that there is a metaphysical system tied to language use, since it is not necessary for an existing object to make sense of a word or sentence (Addis 102). Words and sentences only have meaning within certain linguistic

systems (Skelac and Jandric 45). The meaning of a word is the way in which it is used, and use is only meaningful if it abides by the rules of a system.

2.3 Language-Games

These systems in which language makes sense are called, by Wittgenstein, language-games (Wittgenstein 7). Ordinary language use is a complex network of overlapping and interconnected language games (Skelac and Jandric 45). Each language-game is related to its context of use, as well as its community of users (45 and Conant 239). Importantly, as a game, language is governed by rules (Wittgenstein 31). The relevant linguistic community is responsible for establishing the rules and conventions of the game, though not in a formal, organised sense. Rather, members agree upon meaningful use in the way they respond, both verbally and physically, to each other and their shared environment (Skelac and Jandric 45). So, there are rules which govern whether some utterance has one meaning, some other meaning, or no meaning at all in any given context. In this sense, language-games can show us the context where the use of an utterance is of significance. It rejects the referential system and its metaphysics, and situates language by reminding us that we cannot separate a statement from its speaker, audience, location or time (Conant 239). So, the context of any utterance as well as its recognition in a linguistic community is fundamental to understanding its meaning.

2.4 Making Sense in the Context of Fiction

I now turn back to the paradox of fiction. Armed with an understanding of meaning which does not have metaphysical commitments and a description of language which accounts for context, we can reject the idea that emotional response implies belief, which leads to the notion of irrational emotions. Furthermore, we can make the case that in the realm of responses to fiction, there are language-games at play which make statements and behaviours meaningful, which in other instances would cause confusion. Once we recognise that Radford's formulation of the paradox is not a problem inherent in emotional responses to fiction, but a misapplication of terms from one language-game to another, the paradox is dissolved.

To demonstrate how language-games can make sense of what Radford would regard as irrational, I return to the previous examples. While Radford would suggest that it is irrational to recoil in fear from the screen when Michael Myers appears, Wittgenstein contends that, so long as this is a generally understood behaviour within the context of seeing a

horror film, then it makes sense. If you were to lean over to your friend, and whisper into their ear something like, 'Michael is scary,' this would be a meaningful statement. If you were to whisper into their ear, 'three years ago to the day I ate oatmeal for the last time,' this would not make sense, even if it were a true statement. The metaphysical truth of the statement is irrelevant. It would be nonsense, because a remark of this sort does not cohere with the rules of the appropriate language-game. It would be reasonable to say that, among the right people, in the right location, and at the right time, a statement like, 'I fear Michael Myers ... it's so creepy the way he sneaks around Haddonfield murdering teenagers,' is not taken to mean that someone holds a belief in Myers' existence, nor that their emotions are irrational. The statement is understood by the other members of the community, and is probably taken not to entail belief in the existence of Myers or Haddonfield. So, statements like, 'I am scared by Michael Myers,' and, 'I feel sorry for his victims,' are meaningful in their use, rather than irrational in relation to their objects. It is not that irrationality is directly opposed to meaning, but rather that since language is not predicated on a particular metaphysics, an utterance's relevance is governed by the rules of the language-game at play.

I am not suggesting that there is exactly one language-game in play among theatre-goers, fiction readers, or film watchers. In reality, language-games are always overlapping each other and speakers are frequently switching between games successfully. The above examples are meant to demonstrate that in the contexts where we have emotional responses to fiction, language-games come into play which allow for meaningful verbal and physical expressions of these emotions. Interestingly, both Aristotle and Hume account for the relevance of context, and in their accounts the context of engaging with fiction performs a unique and important function (Freeman 249).

My final claim is that Radford is guilty of removing terms from their context of meaning, thereby giving rise to the confusion which allows for the paradox to take shape. It is fair enough to say we no longer grieve upon learning the story that caused our grief is false (Radford and Weston 68). However, we do not stop grieving because it would be irrational to continue. Rather, we stop grieving because we have learned that we have misapplied our language-games. Furthermore, this is the exact kind of language misuse that Wittgenstein considered responsible for all philosophical conundrums (Skelac and Jandric 46).

Conclusion

Although the relationship between emotion and fiction has a long philosophical history, and Radford's explication of the paradox is seemingly intuitive, it stands on metaphysically shaky ground. The central distinction between fictional and non-fictional entities, which gives rise to the notion that emotional response to an object implies belief in its existence, is a confusion about the way language is used, rather than a genuine problem. Moreover, the myriad constructive solutions that attempt to solve the paradox fall into the same trap of dividing the world up into the mind-independent realm, and the mind-dependent realm. This metaphysical commitment, and its associated picture of language, is ill-equipped to deal with questions regarding fictional entities, since the problem of reference will inevitably arise. Instead, Wittgenstein's use-theory of meaning, and his account of overlapping language-games capture the context-specific nature of meaning. In this essay I have followed a quietist method by first diagnosing the terms of this philosophical debate, and then dissolving the problem by calling into question its metaphysical foundations. Furthermore, I have applied Wittgenstein's account of language to the real-life context of emotional responses to fiction, and dissolved the paradox of fiction at the same time. Radford's paradox of fiction, and the debate it has spawned, do not arise if we accept Wittgenstein's view that words are granted meaning when used within specific contexts which abide by socially constructed and accepted rules, rather than rationality when used to refer to non-fictional entities. In doing so, we can reject the notion that emotional responses imply belief in the existence of the entities which elicit them. Furthermore, we can emotionally engage with our favourite fictions without fear of being charged with the crime of irrationality.

Works Cited

Addis, Mark. *Wittgenstein: A Guide for the Perplexed*. Continuum, 2006.

Conant, James "Wittgenstein on Meaning and Use." *Philosophical Investigations*, vol. 21, no. 3, 1998, 222–50.

Freeman, Damien. "The Paradox of Fiction." In *The Routledge Companion to Philosophy of Literature*, edited by Noël Carroll and John Gibson. Routledge, 2016.

Graham, Gordon. *Philosophy of the Arts: An introduction to aesthetics*. Routledge, 2005.

Matravers, Derek. "Recent philosophy and the fiction/non-fiction distinction." *Collection and Curation*, vol. 27, no. 2, 2018, pp. 93–6.

McNally, Thomas. *Wittgenstein and the Philosophy of Language: The Legacy of the Philosophical Investigations*. Cambridge University Press, 2017.

Radford, Colin and Michael Weston. "How Can We Be Moved by the Fate of Anna Karenina?" *Proceedings of the Aristotelian Society*, Supplementary Volumes 49, 1975, pp. 67–93.

Skelac, Ines and Andrej Jandric. "Meaning as Use: From Wittgenstein to Google's Word2vec." In *Guide to Deep Learning Basics: Logical Historical and Philosophical Perspectives*, edited by Sandro Skansi, Springer, 2020.

Spiegel, Thomas J. "What is Philosophical Quietism (Wittgensteinian or Otherwise)?." In *Quietism, Agnosticism and Mysticism: Mapping the Philosophical Discourse of the East and the West*, edited by Krishna Mani Pathak, Springer, 2021.

Stecker, Robert. "Should We Still Care about the Paradox of Fiction?." British Journal of Aesthetics 51, no. 3, 2011, pp. 295–308.

Tullmann, Katherine and Wesley Buckwalter. "Does the Paradox of Fiction Exist?." *Erkenntnis*, vol. 79, no. 4, 2014, pp. 779–96.

Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G. E. M. Anscombe, Macmillan Publishing, 1958.