Why is it rational to experience emotions towards characters and situations we believe to be fictional?

1- In his 1978 paper, *Fearing Fictions*[1], Kendall Walton outlines a situation in which a man, Charles, is 'watching a horror movie about a terrible green slime'[2]. Walton describes Charles' reactions to the movie, and states that afterwards 'Charles confesses that he was "terrified" of the slime'[3]. Walton objects to this, on the grounds that 'Charles knows perfectly well that the slime is not real, and that he is in no danger'[4] and in doing so raises a paradox (sometimes called the paradox of fiction) which has been approached in various ways over the last thirty years. Non-specifically, it can be outlined as follows:

(1):   We have emotional responses to fictional characters and situations.

(2):   We believe fictional characters and situations are not real.

(3):   A necessary condition for an emotional response is a corresponding belief in the reality of the object(s) of that emotional response.

This essay will briefly consider two of the main responses to the paradox, and explain why each of them is unsatisfactory. I will then discuss how an intuition

---

[1] Kendall Walton, 'Fearing Fictions', 1978 in Lamarque & Olsen eds, *Aesthetics and the Philosophy of Art*. (Oxford: Blackwell, 2004), 307-319.
[2] Walton, *Fearing Fictions*, 307.
[3] *ibid.*
[4] Walton, *Fearing Fictions*, 308.

from one of Walton's earlier essays (*Categories of Art*, 1970)[5], when properly phrased, might serve as a tool to enable us to propose an alternative to (3) which is not inconsistent with (1) and (2), but also helps to explain why the paradox occurs at all.

2- The first response to the paradox I will consider is Walton's own. He held that (1) must be false: 'The fact that Charles is fully aware that the slime is fictional is, I think, a good reason to deny that what he feels is real fear'[6]. He describes the emotion that Charles feels as quasi-fear, and develops a theory whereby Charles is playing a game of make-believe, with the objects on the screen as props. Propositions that are true in a game of make-believe are fictionally true, and sets of fictional truths make up fictional worlds[7]. In our games of make-believe, Walton thinks, we enter these fictional worlds. Thus, if we take the problem of the spectator of a play who dislikes happy endings but sympathises with the heroine, Walton thinks we can explain the apparent contradiction of his beliefs as follows: 'It is merely make-believe that the spectator sympathises with the heroine and wants her to escape. And he (really) wants it to be make-believe that she suffers a cruel end'[8]. Walton places the spectator's sympathy in the fictional world (and thus makes it quasi-sympathy), and the spectator's desire for an unhappy ending in the real world.

---

[5] Kendall Walton, 'Categories of Art', 1970 in Lamarque & Olsen eds, *Aesthetics and the Philosophy of Art*. (Oxford: Blackwell, 2004),142-157.
[6] Walton, *Fearing Fictions*, 308.
[7] Walton, *Fearing Fictions*, 310.
[8] Walton, *Fearing Fictions*, 317.

I disagree with Walton's fundamental premise; that the emotions we feel about fictional characters and situations are not real emotions. Such an assertion is simply too counter-intuitive. To claim that there is a set of cases wherein I can *fail* to correctly describe my own emotional response, despite these cases being to me phenomenologically indistinguishable from other cases where I could *succeed* at correctly describing my emotional response, is something I would only wish to do if all other possible explanatory options had been exhausted. I take it to be the case that the emotion felt by a person in response to works of fiction just is (cases of confusion aside) the emotion that the person would describe it to be. It seems to me that the philosopher's job is to explain *why* Charles confessed 'that he was "terrified" of the slime'[9], not to tell him that he is mistaken.

In his 1981 paper, *How Can We Fear and Pity Fictions?*[10], Peter Lamarque takes a very different approach to solving the paradox. He takes the central question to be 'What are we responding *to* when we fear Othello and pity Desdemona?'[11] and suggests that fictional characters enter our world, rather than, as Walton suggests, us entering theirs. He holds that this allows us to be 'really afraid and really moved'[12]. He expands this to explain that 'fictional characters enter our world in the mundane guise of descriptions … and become the objects of our emotional responses as mental representations …

[9] Walton, *Fearing Fictions*, 307.
[10] Peter Lamarque, 'How Can We Fear and Pity Fictions?', 1981 in Lamarque & Olsen eds, *Aesthetics and the Philosophy of Art*. (Oxford: Blackwell, 2004), 328-336.
[11] Lamarque, *Fear and Pity Fictions?*, 328.
[12] Lamarque, *Fear and Pity Fictions?*, 329.

characterized by those descriptions'[13]. As mental representations or thought contents are quite clearly real, the emotions we have relating to them will also be real.

In one sense, I agree with Lamarque. He provides a plausible account for why we can feel emotions towards fictional characters, and why it is sensible to talk of those emotions as real. He argues 'we need to distinguish between Charles's being frightened *by the slime* and his being frightened *by the thought of the slime*'[14]. However, he does not give an account of why it is the case that I would describe myself as frightened just by the slime. It may be the case that these objects are represented at the level of the brain by thought clusters which *cause* the emotions I am feeling, but I think a full response to the paradox should be able to explain why, if the pity that I feel I believe to be directed at Desdemona, and I also believe Desdemona does not exist, I am not irrational to hold both of those beliefs. The issue is not the actual ontological status of the objects of our beliefs, but the perceived ontological status of the objects of our beliefs. I think a full solution of the paradox should be couched in terms not only of the emotions we would describe ourselves as having, but also in terms of the objects we would describe them as being directed upon.

3- I have outlined the paradox, and several responses to it. I hold, as with almost all of the commentators on the issue (Colin Radford being a notable exception),

---

[13] *ibid*.
[14] Lamarque, *Fear and Pity Fictions?*, 331.

that I do not think that the inconsistency properly represents the ways in which we engage with, and feel emotions towards, fictional characters and situations. I do not think, with Radford[15], that our engagement with fictions can simply be classed as irrational. I think that (1) and (2) are true, and so want to find an alternative to (3) that is not inconsistent with them and helps to explain why our responses to fiction are not irrational.

To help, I want to turn to an older paper of Walton's: 'Categories of Art'. In it, he argues that the aesthetic properties a work of art 'seems to have for us depends not only on what nonaesthetic features we perceive in it, but also on which of them are standard, which variable, and which contra-standard *for us*'[16]. To perceive a work in a certain category, 'is to perceive the "*Gestalt*" of that category in the work'[17]. In short, with respect to each category in which you perceive it, the features of, say, a painting, can be:

(i): Standard: those features which warrant the inclusion of the work in that category, and are standard to it.

(ii): Variable: those features which have no bearing upon the inclusion of the work in that category.

[15] See, for example, Colin Radford, 'How Can We Be Moved By The Fate Of Anna Karenina?', *Proceedings of the Aristotelian Society* supp. vol. 49 (1975).
[16] Walton, *Categories of Art*, 144.
[17] *ibid.*

(iii):   Contra-standard: Those features which would preclude the inclusion of the work in that category.

Walton uses this distinction in several enlightening ways, but the intuition that is useful for our purposes is that this is why it is sensible for us to talk of a painting of a person as looking like a person. When we look at a painting as a painting, all of the features which would cause us to discount it as looking like a person (such as its flatness, the fact it is made of canvas etc) are made standard, and so these features 'just do not count with regard to what (or whom) it looks like'[18]. It is the variable features; 'the colours and shapes on the work's surface, that make it look to us like what it does ... these are the ones which are taken as relevant in determining what (if anything) the work represents'[19]. Hence we can sensibly make statements such as 'this is an exact likeness'.

Very simply, then, when viewing something like a novel as a novel, I think that it might help us get a handle on the original paradox to say, in Walton's terms, that with respect to the category of 'a novel', the feature 'is fictional' is made standard, and so simply does not count when talking about the characters and situations in the novel; just as the flatness of the painting does not count when talking of its likeness to a real person, or what sort of landscape it shows. In terms of emotions, if the variable features of a painting were to disturb or upset me, it would seem irrelevant to remark that the painting was flat. Can we

---

[18] Walton, *Categories of Art*, 146.
[19] *ibid.*

say that, if the characters or situations portrayed in a novel evoke emotions in me, it is similarly irrelevant to remark that they are fictional?

4- To make things clearer, I am going to re-express Walton's definitions in terms of the *attention* we pay to features of our experience:

(I):   Standard features are those to which, when experiencing something with respect to the relevant category, I do not attend, although they may affect the ways in which I do attend to

(II):   Variable features, which are the features to which I do attend, when experiencing something with respect to the relevant category.

Although not strictly relevant to a resolution of the paradox,

(III):   Contra-standard features would be those which, if attended to, would demand a re-ordering of my attention with respect to a category or set of categories with respect to which those features would no longer be contra-standard.

In Walton's terms, then, to percieve a work as in a particular category is to fully attend just to the set of features which are variable with respect to that category, and so to percieve the gestalt of that category in the work. I shall, however, slightly expand Walton's original definition, in that I take a 'feature' to be

anything to which one can attend in the normal course of experience. Anything to which I could fully attend (i.e. anything which I can make the object of my attention) is a feature. This includes physical features; flatness, etc and mental features; thoughts, beliefs, etc. I can focus my attention on the flatness of something. I can also focus my attention on the fact that I believe that something to be flat. Both flatness and belief in flatness are distinct features that may be attended to, but whilst it is certainly the case that attending to something as flat commits me to the belief that that something is flat, it does not commit me to simultaneously attend to that belief. I am taking my beliefs to be features *of my experience*.

Now this attentional move is tricky, and I don't have the space to fully explore it here, but I will note several further points which are of use. The first is that, if we wanted to, this is where we can connect Lamarque's theory and my own. Every feature (mental or physical) has a corresponding, in Lamarque's terms, thought, and every group of features their corresponding thought-cluster. To attend to a feature or features is therefore to have that corresponding thought or thought cluster.

Secondly, I must make clear the distinction between holding a belief and attending to that belief. In any specific moment, I hold dozens of beliefs. I believe that I am sitting in a chair, that I exist, that I am slightly hungry, that I am quite warm, etcetera. Clearly, however, I do not (and cannot) attend to all of those beliefs in the moment that I have them. I am certainly able to list most, if not all of those beliefs; but to do so requires me to attend to each of them successively.

It is not the case that those beliefs I do not attend to I do not have, because those beliefs (as standard features) can affect the ways in which I do attend to the features of my experience to which I am currently attending. This is just like the fact that I can know that a painting is flat without attending to its flatness (or my belief that I know it is flat) whilst looking at the painting; and that the fact that it is flat affects the way in which I do look at the painting. This is why I added, in my definition of (I), that Standard features 'may affect the ways in which I do attend to (II) variable features..'. I am therefore taking it to be the case that I can hold beliefs without attending to them; that my beliefs can be *standard* features of my experience.

Thirdly, I think my experience can only be described as incoherent or irrational if there is an inconsistency in the features to which I am *currently* attending. If the features of my attention are consistent, my experience is rational.

Lastly, I have said that a feature is anything which I can make the object(s) of my attention, and it seems plausible that if those objects might also be the objects of some emotion when attended to in a certain way, and I am attending to them in just such a way and *not* currently attending to any feature which might preclude me from rationally experienceing that emotion, I will experience that emotion. More succinctly, a certain set of features, when attended to in a certain manner, will result in a certain emotional response. Thus, I will rephrase (3) as:

(3'): A necessary condition for an emotional response is an attention just to the object(s) of that emotional response.

On reflection, it is easy to see why this might be the case; emotions are responses to experience, and it would have been inefficient for our ancestors to rely on anything other than the things to which they are currently attending to stimulate that response. If you have to attend to the ontological status of the charging, toothy shape of a sabre-toothed tiger, *before* you have the emotion that relates to that set of features, you're unlikely to survive to breed. If we alter the set of features we are attending to we thus alter the the emotions we are experiencing. Nowadays, for example, we might remind ourselves we are in a cinema by attending to our belief that this is the case and enough of the beliefs this entails to mitigate the emotions we felt when fully attending just to the image of the tiger. This is why we feel no need to run for the exits.

5- For (3') to hold, I need an example that one can have an emotional response just by attending to the objects of that emotion without also attending to any belief about the ontological status of those objects. In any other essay, I would give fiction as the primary example; however, given this is the object of my discussion, I shall avoid accusations of circularity by turning to B. J. Rosebury's account of the hypothetical. Rosebury suggests that the case 'in which one propounds to oneself an attractive or repellent hypothesis ... do[es] not

presuppose beliefs in the actuality of [its] content'[20]. I do not agree with Rosebury that the hypothetical is a useful model for fiction (I think both fictions and hypotheticals are two examples of the same phenomenom), but I do not have to. The hypothetical is simply, for the purposes of this argument, an example outside of fiction where one can have an emotional response without attending to any sort of belief in the existence of the situation that is the object of that emotional response.

When reading a novel as a novel then, and engaging with the characters as characters, one does not attend to the ontological status of the situations the novel describes or of the characters. (1), (2) and (3') are not inconsistent and as (2) does not commit us to *attend* to the belief we hold there need be no inconsistency between my belief that a character is not real and any reality implicit in my emotional responses to them. My experience can therefore be rational. To feel emotions about characters whilst holding the belief that they do not exist is only incoherent (and gives rise to the paradox) on the level of beliefs-that-one-holds, *not* the level of beliefs-to-which-one-is-currently-attending. A corresponding belief in the reality of the object(s) of an emotional response is a necessary condition only for a rational account of that emotional response.

6- A few final points. Firstly, not attending to beliefs is not to be confused with suspending them (although Coleridge and those who advocate his view were

---

[20] B. J. Rosebury, 'Fiction, Emotion and "Belief": A Reply to Eva Schaper', *The British Journal of Aesthetics* 19 (1979): 128.

perhaps attempting to describe an intuition similar to my own). Ideas of suspension of disbelief, as Eva Schaper shows in *Fiction and the Suspension of Disbelief*[21], become incoherent very quickly. As I have said, one can hold beliefs (or disbeliefs) but not attend to them; indeed, in terms of viewing things in categories at all, one must hold beliefs to which one does not attend to shape the way in which one does attend to features that are variable with respect to each category.

Secondly, it is not simply the case that there is a category for each set of features we might attend to in any given moment. Rather, categories deliniate certain, often named, sets of features which we have been taught (or learned) to recognise as sets to which we *could* fully attend. Importantly, our attention is fluid. We could fully attend to just those features which were variable with respect to a particular category, and thus percieve the gestalt of that category in our experience, or we could attend (with respect to the same category) only partially to those features which are variable with respect to it, and also to some features which are standard with respect to it. This does not mean that we are now attending just to the features which are variable with respect to some new category, although prolonged attention just to those features might result in such a category being created. In our experience of reading, we attend to the different features of the experience as we are differently prompted; there is a constant shift of attention. Alex Niell describes this process in almost exactly the way I would in a paragraph towards the end of his essay *Fiction and the Emotions*: 'In

---

[21] Eva Schaper, 'Fiction and the Suspension of Disbelief', *The British Journal of Aesthetics* 18 (1978): 31-44.

responding to a work of fiction, we tend to adopt a variety of attitudes or stances towards the work; the focus of our concern shifts between the various aspects of that work'[22]. He explicitly discusses attention, but only in connection with pity, and to explain how we can pity a character but also want her to suffer. He does not, however, seem to see that examining the whole process as an attentional one can explain why the paradox occurs.

This fluidity of attention also has the potential to help explain the problem of conscious engagement with one's emotions; why, for example, telling myself something doesn't exist can reduce my fear of it. If I am scared by fully attending just to the zombies in a zombie movie, also attending to the fact they don't exist mitigates my emotion because it changes the set of features which I am attending to. My attention just to the zombies is reduced, and so my fear of the zombies is reduced. Lamarque makes a similar point in *How Can We Fear and Pity Fictions?* when he says that 'The propensity of a thought to be frightening is likely to increase in relation to the level of reflection or imaginative involvement that is directed to it'[23], but he sees this in terms of *adding* related thoughts to a thought or thought cluster, whereas I see this in terms of shifting one's attention away from just those features which are variable with respect to a particular category which one might find frightening when attended to fully.

An interesting side point is that it can be difficult to successfully refocus my attention onto my belief in the zombies' ontological status; if the zombies are

---

[22] Alex Niell, 'Fiction and the Emotions', *American Philosophical Quarterly* 30 (1993): 11.
[23] Lamarque, *Fear and Pity Fictions?*, 330.

particularly compelling, they will keep dragging my attention back to just them. This is why it can be the case that it is very hard to stop yourself being frightened of things that you **know** don't exist. I suspect that this is more the case with children than adults because adults have learned to control their attention better. Children can repeat 'zombies don't exist' to themselves whilst watching a film; I find that when watching something I think might be about to scare me on a laptop, I will push the laptop further away from me, or lower the sound, or deselect full screen. All of these activities make it harder for my attention to remain just on the characters and situations that might scare me. If I push the laptop away, the screen occupies a smaller portion of my visual field, and so it is harder for me to attend to it alone. If I lower the sound, sounds from other sources will become more apparent, and harder to ignore, and so my attention will be less focussed. Spreading my attention buffers me against difficult emotions, like fear.

Some emotions are particular to attending to both variable features and the features which would be standard with respect to the category with respect to which the former features were variable; the emotions associated with irony and satire would be good examples.

7- In conclusion, I think addressing the problem in terms of Walton's categories, rephrased in attentional terms, allows us to describe a coherent experience of fiction, and helps explain why the paradox arises.