Fiction, Pleasurable Tragedy, and the HOT Theory of Consciousness

Rocco J. Gennaro

[Final version in Philosophical Papers, 2000]

Much has been made over the past few decades of two related problems in aesthetics. First, the "feeling fiction problem," as I will call it, asks: is it rational to be moved by what happens to fictional characters? How can we care about what happens to people who we know are not real?1 Second, the so-called "paradox of tragedy" is embodied in the question: Why or how is it that we take pleasure in artworks (e.g. tragedies) which are clearly designed to cause in us such feelings as sadness and fear?2 Various solutions to these puzzles have been proposed, but my primary aim is neither to offer a novel solution nor to summarize and critique most of the alternatives.3 My focus instead will be on the issue of consciousness and, more specifically, to view these problems from the point of the view of the so-called "higher-order thought theory of consciousness" (the HOT theory). Although some work on these puzzles have raised important questions about the nature of consciousness and "aesthetic experience," no attempt has yet been made to examine them in light of a specific theory.

1. What is the HOT Theory?

In the absence of any plausible reductionist account of consciousness in nonmentalistic terms, the HOT theory says that the best explanation for what makes a mental state conscious is that it is accompanied by a thought (or awareness) that one is in that state.4 The sense of 'conscious state' I have in mind is the same as Nagel's sense, i.e. there is 'something it is like to be in that state' from a subjective or first-person point of view.5 Now, when the conscious mental state is a first-order world-directed state the higher-order thought (HOT) is not itself conscious; otherwise, circularity and an infinite regress would follow. When the HOT is itself conscious, there is a yet higher-order (or third-order) thought directed at the second-order state. In this case we have introspection which involves a conscious HOT directed at an inner mental state. When
one introspects, one's attention is directed 'back into' one's mind. Introspection is therefore a type of self-consciousness which involves having a conscious HOT directed at a first-order mental state.

For example, what makes my desire to finish this paper a conscious first-order desire is that there is a (nonconscious) HOT directed at the desire. In such a case, my conscious focus is directed at the paper and my computer terminal. When I am introspecting my desire, however, then I have a conscious HOT directed at the desire itself.\(^6\) We might also distinguish between momentary focused introspection which only involves a brief conscious HOT, and deliberate introspection which involves the use of reason and a more sustained inner-directed conscious thinking over time. In this paper, I will use the term 'introspection' as short for 'momentary focused introspection.' It is this type of introspection which is most relevant to the discussion here and, as we will see, particularly to the paradox of tragedy.

In any case, the fundamental question that should be answered by any theory of consciousness is: what makes a mental state a conscious mental state? The HOT theory says that what makes a mental state conscious is the presence of a suitable\(^7\) higher-order thought directed at it.\(^8\) The HOT theory can therefore explain what distinguishes a conscious mental state from a nonconscious state: the former has a HOT directed at it whereas the latter does not. In short, when we are in a nonconscious mental state we are not aware at all that we are in it, whereas when we are in a conscious mental state we are (nonconsciously, at least) aware that we are in it.

As I mentioned earlier, the HOT theorist must avoid definitional circularity and an infinite regress by explaining that the HOT is not itself conscious when one has a first-order conscious state. Otherwise, we would be answering our fundamental question by appealing to consciousness, which is circular. Moreover, we would have an infinite regress because for every conscious state there would have to be a higher-order conscious state and so on \textit{ad infinitum}.\(^9\)

2. The Feeling Fiction Problem
Let us now return to the feeling fiction problem. Once again: is it rational to be moved by what happens to fictional characters? How can we care about what happens to people who we know are not real? There seems to be something rather puzzling about the fact that we are often moved emotionally by non-existent people and events.

What is going on according to the HOT theory? It seems to me that we have first-order conscious mental states directed at something outside of us. We fear the shark in *Jaws*, we pity Anna Karenina, we grieve for Desdemona in *Othello*. There are various theories about what exactly our consciousness is directed at, but it is clear that while we are watching a play or movie, or reading a book, our conscious attention is typically focused outside of us. It is because of this outer-directed consciousness that we are moved by fiction.

At the same time, of course, we do also have knowledge of the fictional nature of the people or objects in question. Although this knowledge is present simultaneous to the outer-directed conscious state, it is, however, clearly not conscious knowledge at the same time that we are consciously feeling the emotions in question. That is, while I am consciously fearing the shark in *Jaws* I am not also consciously focused on my knowledge of its fictional nature. Indeed, the HOT theory neatly explains the difference: The conscious fearing is an outer-directed first-order conscious state, whereas consciously focusing on my knowledge is a second-order conscious, or introspective, state. My consciousness cannot be focused outside of me and inwardly at the same time, although quick shifts back and forth no doubt frequently occur. However, the mark of great fiction is that it causes us to rarely, if ever, have such self-conscious moments. We become "caught up" in our outer-directed consciousness. Indeed, frequent interruptions of self-consciousness can spoil what could otherwise be a moving play. This is particularly true if one is consciously focused on the knowledge that "it's all fiction" or "those characters aren't real." If one consciously focused on that knowledge while being in the theatre, then even the greatest plays and movies would not move us at all. Try it sometime: go to see a
Shakespeare play and force yourself often to have the self-conscious thought that "this is all just fiction" or "those characters don't really exist." You will neither be moved nor will you enjoy the play.

We can now look critically at Colin Radford's influential paper on this topic. He argues that we are in an irrational state of mind when moved by fiction, and replies to six possible solutions to the feeling fiction problem. I wish to examine two of the six and am sympathetic to the first solution discarded by Radford. The basic idea is that when we read a book or watch a play, "we are 'caught up' and respond and we 'forget' or are no longer aware that we are only reading a book or watching a play. In particular, we forget that Anna Karenina, Madame Bovary, Mercutio and so on are not real persons." Radford replies that "this won't do...it turns adults into children." But does it? There is an ambiguity in his use of the term 'aware.' If Radford means to imply that we are always consciously aware that Anna Karenina, Madame Bovary, and so on are not real persons, then I think he is mistaken. As we saw earlier, while we are having conscious first-order mental states directed at the play, we are not also consciously aware of the knowledge in question. On the other hand, if Radford is simply saying that it is wrong to suppose that we are completely unaware (both consciously and nonconsciously) that we are watching fictional characters, then he is correct. But this is only to say that we have the nonconscious mental state in question, e.g. the belief or knowledge that Anna Karenina is a fictional character. This nonconscious mental state can be accommodated in the HOT theory in the way described earlier: it is a mental state that is not accompanied by a HOT. The difference between adults and children is that children sometimes really do mistakenly believe that fictional characters are real, and so they do not even have the nonconscious knowledge that the characters are fictional.

Now Radford asks "Do we shout to get on the stage when, watching Romeo and Juliet, we see that Tybalt is going to kill Mercutio? We do not...If we really did think that someone was
really being slain...we would try to do something..."14 But again, first, there is the ambiguity: If we really did consciously think that a real person was being slain, then we would try to do something. True enough. But, of course, we still have the nonconscious knowledge that no real person is being slain. Second, Radford overlooks the fact that such nonconscious knowledge can still be sufficient to prevent the behavior in question. Nonconscious mental states, like conscious ones, can cause or prevent behavior. Although this may not always be the case, a nonconscious state can cause or prevent behavior (e.g. not rushing to the stage to prevent Mercutio's death) while another conscious mental state (e.g. the visual perception of the action on stage) causes one to fear for Mercutio. So we have a nonconscious state preventing a behavior and a conscious state causing an emotion. Is this an irrational state of mind? I am really not sure, but it hardly seems as incoherent or inconsistent as Radford suggests.

Much of this also applies to Radford's dismissal of the proposed second solution to the problem; namely, that we 'suspend our disbelief' in Mercutio's reality, even though we do not ever forget that he is only a character in a play.15 Radford urges, however, that the paradox still remains because we are "never unaware that we are watching a play..."16 Once again, does he mean consciously or nonconsciously unaware? If Radford is saying that we are never nonconsciously unaware that we are watching a play, then he is right. But that is only to say that we have, throughout the play, the nonconscious belief that we are only watching a play. However, if he is saying that we are never consciously unaware that we are watching a play, then, as we have seen, he is mistaken. Radford himself says so: "When we watch a play we do not direct our [conscious] thoughts to it's only being a play. We don't continually remind ourselves of this -- unless we are trying to reduce the effect of the work on us."17 So even Radford acknowledges the fact that our conscious focus is directed outside of us, and that self-consciously reminding ourselves of the fictional nature of the characters would only "reduce the effect of the work on us." As we saw earlier, this is absolutely correct. However, once we realize that there is
no contradiction in saying that we are nonconsciously aware that the characters are fictional but that we are also consciously caught up in the plight of the characters, Radford's replies lose much of their force.

Radford makes a similar mistake later in his paper when he contrasts our reaction to fiction to our reaction to real life. He again mentions the above solutions and says that we "are still aware that we are watching a play and that Mercutio is 'only' a character." The same reply is appropriate here: Radford is using the term 'aware' ambiguously between nonconsciously aware and consciously aware. If he means the former, then he is correct. But if he means the latter, then he is incorrect, as even he acknowledges.

The difference between our reaction to fiction and our reaction to real life tragedy has to do with the fact that in real life we (obviously) do not have the belief or knowledge at all that the persons are not real. Therefore, there is no such nonconscious mental state inhibiting us from acting to prevent the tragedy in question. We are consciously aware that the hostages in the bank are real people, and so we try to prevent their injury or death. If a hostage is killed, we grieve and feel for that person and her family. Moreover, since there is no nonconscious belief in the fictional nature of these real people, we cannot self-consciously focus on it either during or after the ordeal. This explains why real grief lasts so much longer than any grief that one might feel in a movie theater or at a play.

In any case, perhaps we can now see what Radford must mean by "irrational"; that is, having two simultaneous and contradictory mental states. Presumably the contradictory states in question are (1) the knowledge that some character is fictional, and (2) the belief that something bad is happening to a person. But if one of these mental states (the knowledge that the character is fictional) is not conscious at the time that one is consciously feeling fiction, then perhaps we should not say that the person who is pitying Anna Karenina is irrational. I do not have a general theory of rationality to offer, but it seems to me that several useful distinctions can be made.
Let us distinguish between the following senses of irrationality (though there may be others):

A. **Cognitive irrationality (CI)**: a person has at least two contradictory mental states.

B. **Conscious cognitive irrationality (CCI)**: a person has two conscious contradictory mental states.

C. **Behavioral irrationality (BI)**: a person is behaving in a way that contradicts at least one of that person's mental states.

I am inclined to think that, of the three, only CCI is sufficient for irrationality. At the least, it is the clearest case of irrationality. If a person is consciously aware that she is in two contradictory mental states, then we should say that she is being irrational. In these circumstances, we expect the person to "give up" one of her beliefs. However, as we have seen, a person moved by fiction is not irrational in the sense of CCI. Whatever contradiction there arguably might be between two mental states, such as believing both that a person is suffering and that the person does not exist, we are not consciously aware of both mental states at the time of being moved.

It is probably true, however, that one moved by fiction is irrational in the sense of CI because the two mental states are still present (though one is nonconscious). However, such a definition of irrationality seems much too strong. We would probably all be irrational according to CI. If this is what Radford means by "irrational" (as he seems to), then anyone with simultaneous and contradictory mental states is in an irrational state of mind. But surely this would apply to virtually every human being. Of the millions of (conscious and nonconscious) mental states that we have at any given time, each of us is likely to host a contradiction somewhere.

Perhaps Radford and others only mean to put forward something like BI. For example, I
am behaving (e.g. weeping) in a way that contradicts my knowledge that the characters in the play do not exist. Perhaps there is something irrational in this situation, but we must be careful here. First, we are not talking about a cognitive kind of irrationality as in CI or CCI. Second, although we have seen that nonconscious mental states can cause or prevent one's behavior, it is also clearly possible for one's conscious mental state at a time (conscious pitying) to be the principle cause of one's behavior (weeping) at a particular time. We might say, in those circumstances, that the conscious mental state causes a behavior which overrides the typical causal role of the nonconscious mental state. This is not always so, but it is the case whenever we are consciously moved by fiction.21

3. The Paradox of Tragedy

Let us shift to the paradox of tragedy which, again, is basically this: Why or how is it that we take pleasure in artworks (e.g. tragedies) which are clearly designed to cause in us such feelings as sadness and fear? How can we enjoy seeing miserable and suffering characters in a play?

Although both Aristotle and Hume22 have been historically influential, I believe that Susan Feagin's more recent approach is closest to the truth.23 Moreover, we can see how it lends itself to analysis by a HOT theorist. Feagin speaks of "two kinds of responses to art: a... direct response is a response to the qualities and content of the work. A meta-response is a response to the direct response."24 We derive the pleasure of a tragedy in the meta-response "arising from our awareness of, and in response to, the fact that we do have unpleasant direct responses to unpleasant events as they occur in the performing and literary arts. We find ourselves to be the kind of people who respond negatively to villany, treachery, and injustice. This discovery, or reminder, is something which, quite justly, yields satisfaction. In a way it shows what we care for, and in showing us we care for the welfare of human beings and that we deplore the immoral forces that defeat them, it reminds us of our common humanity."25 She later explains that the
pleasurable feeling brought on by tragedy is the same one that makes possible moral action in real life.\textsuperscript{26}

Thus, if we adopt Feagin's approach we see that two kinds of mental states are involved when we enjoy a tragedy: (1) the unpleasant conscious first-order mental state directed at the tragedy (the direct response); and (2) a pleasant conscious higher-order awareness directed at the first-order mental state (the meta-response).

Two crucial points need to be made here: First, the pleasure of tragedy only appears in (2); namely, in the self-conscious or introspective state. One has conscious thoughts of one's own empathy, displeasure, or pain and that is what results in the feeling of pleasure for the reason that Feagin gives; namely, that only morally sensitive people would feel such pain and compassion for others. As I mentioned earlier, however, we can and do frequently move back and forth between the conscious and self-conscious state. Second, in contrast to the feeling fiction problem, the pleasure of tragedy involves a second-order conscious state directed at the feeling itself. The problematic mental state in the feeling fiction problem was instead a second-order conscious state directed at the knowledge that one is only seeing or reading fiction.

So the emotion felt in the feeling fiction problem is best understood as involving only first-order conscious states whereas the paradox of tragedy is handled best once we recognize that a conscious HOT is present and is directed at one's own mental state of sadness or empathy. When we are moved by fiction we are not in introspective mental states, but when we derive pleasure from a fictional tragedy we are in those conscious HOTs. The force of the meta-response is particularly vivid after the play or movie is over. We are then more free to reflect on our feelings because there is no longer the perceptual input causing a direct response.

In any case, we can see how Feagin's theory fits in with the HOT theory, and, in turn, how the HOT theory can shed light on her view. We must therefore disagree with Packer when he says that "[a]n apparent difficulty with Feagin's argument is that distinguishing these two levels
of direct response and meta-response, either in aesthetic experiences or in other situations, is problematic."27 Packer does, however, ultimately "proceed on the assumption that distinguishing such levels of response is indeed plausible,"28 but my point here is only that there is no need to treat the distinction as problematic at all.

Another advantage of bringing this problem together with the HOT theory is that it can nicely handle one of the rather perplexing aspects of the paradox. As Neill puts it: "[a]ny plausible account of the paradox of tragedy must involve a recognition that the pleasure and the "pain" that the tragedy gives rise to are in some way internally related; that the tragic pleasure lies at least partly in the pain."29 The HOT theory recognizes this crucial feature by making it clear how the conscious HOT, which is pleasureable, is directed at the first-order unpleasant mental state. The tragic pleasure is, at the very least, intimately related to the pain state. On the HOT theory, the lower-order pain is the very content of the higher-order thought. The HOT is of the form "I am currently aware of my feeling pain (or empathy, or compassion) for character C." This is precisely the kind of internal relation that can address Neill's challenge. The conscious HOT is directed at the first-order mental state of pain.

This point is crucial because, as Neill also notes, we can then remain faithful to the very statement of the paradox by both Aristotle and Hume. Aristotle rightly saw that "the tragic pleasure is that of fear and pity."30 It is therefore a pleasure of that pain. The higher-order pleasure is directed at the lower-order pain. Hume also explained that the "unaccountable pleasure" is pleasure "which the spectators of a well-written tragedy receive from sorrow, terror, anxiety and other passions that are themselves disagreeable and uneasy."31 As the HOT theory states, the HOT arises from32 and is directed at the first-order mental state.
NOTES


3. But see the references in notes #1 and #2 for most of the basic theories.


6. It should be noted that I am mainly concerned here with an analysis of the locution 'x (mental state) is conscious.' This ought to be distinguished from 'x (an organism) is conscious' as well as the two-place predicate 'x is conscious of y.' For some discussion of this terminological matter, see CSC pp. 3-5.

7. One might reasonably ask: "Just what makes a higher-order thought 'suitable'?" A full answer to this question would lead to a lengthy digression which I cannot pursue here. One condition, for example, would be that the HOT must be a 'momentary' and 'occurrent' state as opposed to a 'dispositional' state of some kind. See CSC, chapters three and four, for my attempt at answering the above question.

8. The terminology here can be a bit confusing. Sometimes thought is contrasted with perception whereas sometimes the term 'thought' is also used as a generic term covering virtually all kind of mental states. For our purposes, we can think of the higher-order state as some kind of higher-order awareness. For some discussion of the alleged differences between the 'perceptual' and 'thought' models, see CSC pp. 95-101. See also CSC pp. 36-43 for at least one attempt to address these terminological difficulties in the context of the HOT theory.

10. For example, our minds might be directed at possible real people or we may be reminded of similar actual people or events. In either case, one might argue that fictional entities do exist in some sense. See Peter Van Inwagen, "Creatures of Fiction," *American Philosophical Quarterly* 14 (1977), pp. 299-308.

11. See the reference in note #1. Radford's paper is reprinted in *Arguing About Art*, eds. Alex Neill and Aaron Ridley (McGraw Hill, 1995), pp. 165-175. All page references come from this anthology.


13. Ibid.


16. Ibid.
17. Ibid.

18. For example, Radford pp. 173-5.


21. It seems to me that BI (behavioral irrationality) is closely related to the traditional problem of *akratic* action, or "weakness of will," but this is a topic that I will not pursue in this paper. However, for one attempt to link weakness of will to the paradox of tragedy, see Christopher Williams, "Is Tragedy Paradoxical?," *British Journal of Aesthetics* 38 (1998), pp. 47-62.

22. See the references in note #2.

23. See the reference in note #2. Feagin’s paper is reprinted in *Arguing About Art*, eds. Alex Neill and Aaron Ridley (McGraw Hill, 1995), pp. 204-217. All page references are from this anthology.


25. Feagin, p. 209.

27. Packer, p. 213.

28. Ibid.


30. Aristotle, p. 52, emphasis added.

31. Hume, p. 216, emphasis added.

32. I use the general expression "arises from" not only to reflect Hume's wording but also because whether or not the HOT is caused by the first-order state in some strict sense is actually a somewhat controversial claim partly due to other well-known problems with the nature of causation. See *CSC* pp. 73-75 for some discussion on this topic.