

Genuine Rational Fictional Emotions

Tamar Szabó Gendler
Karson Kovakovich¹

Forthcoming: *Contemporary Debates in Aesthetics* (Blackwell)

The Paradox

The “paradox of fictional emotions” involves a trio of claims that are jointly inconsistent but individually plausible. Resolution of the paradox thus requires that we deny at least one of these plausible claims.

The paradox has been formulated in various ways (some of which we discuss below), but for the purposes of this chapter, we will focus on the following three claims, which we will refer to respectively as the Response Condition (1), the Belief Condition (2) and the Coordination Condition (3).

Regarding certain fictional characters (and situations) F, it is simultaneously true that:

- (1) We have genuine and rational emotional responses towards F
- (2) We believe that F is purely fictional

At the same time, it is also true that:

- (3) In order for us to have genuine and rational emotional responses towards a character (or situation), we must not believe that the character (or situation) is purely fictional.

The inconsistency among the three claims is clear: the Response and Belief Conditions together tell us that we can have genuine rational emotions towards F while believing F to be purely fictional; the Coordination Condition denies that this conjunction is possible. So while any pair of the trio can be consistently maintained, endorsing all three at the same time results in a contradiction. But each of the three claims is also *prima facie* plausible. We do seem to have genuine and rational emotional responses towards purely fictional characters and situations (as when we shed authentic, appropriate tears at the report of Anna Karenina’s demise); at the same time, we seem to believe that those characters and situations are purely fictional (we do not expect to read a report of Anna’s suicide in the annals of the Leningrad Railroad Authority, nor do we expect to be able to intervene in any way regarding the events described in the novel). Still, there seems to be something irrational, inauthentic or even impossible about responding emotionally to things we believe to be purely fictional. (How can we rationally feel genuine fear for something we know to be merely imaginary, or authentically respond with anger to something we know could never have happened?)

Since they cannot be true simultaneously, it must be that one of the *prima facie* plausible claims is, in fact, false: but which one? Each has been forcefully denied by well-respected figures in the literature – and important insights about the nature of our

¹ For comments on an earlier draft, we are grateful to Gregory Currie, John Hawthorne and Zoltán Gendler Szabó.

emotional responses underlie each of these denials. At the same time, we will argue, those who attempt to resolve the paradox by denying either the Response or Belief Condition tend to rest their attempts on a distorted conception of the relations between emotions, beliefs and actions. Careful examination of these connections in light of recent empirical research on the nature of emotions offer grounds for thinking that it is the Coordination Condition that should be denied. Drawing on research by Antonio Damasio and insights of Paul Harris, we will suggest that it is *not* a condition on our having genuine and rational emotional responses towards a character (or situation) that we believe the character (or situation) to be non-fictional. Rather, we will suggest, our cognitive architecture is such that without the tendency to feel something relevantly akin to real emotions in the case of merely imagined situations, we would be unable to engage in practical reasoning. (Damasio 1997, 1999; Harris 2000.)

Before continuing, four caveats about what we will and won't be doing in this chapter. First, given limits of space, we will have little to say about the nature of emotion. In particular, we will set aside important questions about whether the collection of attitudes and feelings generally referred to as "emotions" forms a natural class, whether there are correctness-conditions for feeling emotions, and whether emotions are – strictly speaking – attitudes directed at particular entities. Instead, we will focus on two classic examples from the literature – pity and fear – each of which is at least a plausible example of an emotional response that, at least in certain cases, seems to be object-directed and correctness-evaluable. Second, the notion of rationality at play in our discussion will be an extremely thin one: in calling an emotion "rational" we are claiming only that it is *not irrational* – that it does not interfere with our capacity to function as agents who make effective use of means-ends reasoning, and that it does not directly involve us in inconsistent belief. In so doing, we neglect a number of important distinctions – between instrumental and intrinsic rationality, between theoretical and practical rationality, and between "act-rationality" and "rule-rationality". (The issue of rationality of emotions is discussed in more detail in Derek Matravers's companion piece.) Third, we will help ourselves to a loose notion of "fictional" – contrasting here with "non-fictional" – where characters such as Anna Karenina and places such as Oz are fictional in the sense that we do not consider them to be concrete denizens of the actual world, or of plausible continuations of the actual world. (We here ignore a number of important issues relating to authorial intent, truth in fiction, and – except for a brief discussion below – the ontology of fictional characters.) Finally, we will assume it to be common-ground among all parties that each of the three conditions in the paradox has at least *prima facie* plausibility, but is in-principle deniable. So we will have little to say about views that stipulate at the outset that we can feel genuine or rational emotions only towards entities that are actual, making the truth of the Coordination Condition a definitional matter, or views according to which we do not even *seem* to have emotional responses to fictional characters, rendering the Response Condition trivially false. Of course, it is not always an easy matter to separate terminological from substantive disputes, particularly when one is engaged in conceptual analysis – but we will do our best.

In order to remain focused on substantive issues, we will devote most of our discussion to characterizing the descriptive and normative conditions governing our emotional responses to various sorts of real and imaginary situations, devoting the final

sections to applying this characterization to a number of proposed resolutions to the paradox. For ease of presentation, we will use the term “fictional emotion” to refer to emotional responses that we apparently have towards characters and events that we believe to be fictional, and “actual emotion” to refer to emotional responses that we apparently have towards characters and events that we believe to be actual. (So, “fictional emotion” as we use the expression does not refer to an emotional state of a fictional character (Anna’s anguish), but to an actual person’s apparent emotional response to such a character (your pity.)) We can then state our fundamental questions as follows: What is the significance of the manifest similarities between our fictional and actual emotional reactions, and what is the significance of their manifest differences? Is this configuration of similarity and difference indicative of something problematic in our emotional responses to fiction? And does this pattern of similarity and difference suggest that fictional and actual emotions are two species of the same genus?

In our discussion below, we will argue that the manifest similarities between fictional and actual emotions *are* significant, even in light of their differences, that this configuration of responses *is not* a pathological one, given facts about our cognitive architecture, and that it *is* reasonable to employ the expression “genuine, rational emotion” in describing both actual and fictional emotions. While those who deny the Response Condition are correct to note that our fictional and actual emotions differ in their subject-matter and motivational force, this difference is not sufficient to render fictional emotions either inauthentic or irrational. And while those who deny the Belief Condition are correct to note that the similarities between actual and fictional emotions indicate that we treat their apparent objects in comparable ways, this comparable treatment need not be a reflection of a belief in the actuality of fictional entities. Rather, we will contend, it is the Coordination Condition that articulates a false constraint on our emotional reactions: it is crucial to our ability to make rational decisions about various courses of action that we respond with genuine emotions to situations that we know to be non-actual.

The bulk of this chapter will be devoted to articulating and defending this claim. But before turning to this, we need to undertake two matters of housekeeping. In stating the paradox, we spoke of having genuine and rational emotional responses towards fictional characters. But there are important philosophical questions about what it even means to say that we feel fear or pity for something that does not exist; we discuss these issues briefly and inconclusively in the next section. Second, it is important to realize that there has been an extensive literature on the topic of fictional emotions (including several book-length treatments of it), and while this is not the place for a full literature review, we devote the subsequent section to showing how our own presentation of the puzzle accords with some of the most influential of these discussions. With this groundwork in place, we then turn to a presentation of our own view, followed by a closing section where we contrast our approach with a number of other extant approaches.

Fictional Characters

In our statement of the paradox, we spoke of “having genuine and rational emotional responses towards F” where F is a fictional character or event. But this raises certain puzzles. On most standard semantics, when a name occurs as the direct object of a verb or in a that-clause of an emotional attitude verb (“I pity Anna Karenina” “I fear that

Anna is going to kill herself”), we can quantify into the name position, and conclude that something exists that is referred to by the name ($\exists x (x = \text{Anna Karenina})$). (A similar puzzle arises with the prefix “It is true in the story that...” – from “It is true in the story that Anna Karenina killed herself” we can, it seems, conclude that Anna Karenina exists.) *Prima facie*, however, we seem committed to saying both that we pity Anna Karenina and that she does not exist. How is this puzzle to be resolved?

Four strategies suggest themselves, each with costs and benefits. The first is to reject the standard semantic picture that gives rise to the puzzle in the first place. Contrary to common assumption, we cannot, in general, quantify into name slots. So in saying that we pity Anna, we are not committed to her existence. Many find such revisionary semantics unpalatable. The second is to claim, following Meinong, that there *are* things, such as fictional characters, that do not *exist*. We might then distinguish a broad quantifier that can range over both existent and non-existent entities from a narrow quantifier that can range only over things that exist. While quantifying in with the broad quantifier is legitimate, quantifying in with the narrow quantifier is not. So in saying that we pity Anna, we are committed to *there being* Anna, but not to her existence. (Parsons 1980; Crittendon 1991.) Many find such a view ontologically suspect.

The third strategy accepts both the standard semantic picture and a non-Meinongian ontology, and thus accepts that in saying that we pity Anna, we are indeed committed to Anna’s existence. On one substrategy of this view, fictional characters exist but they do not inhabit our actual world; rather, they exist as flesh-and-blood entities in some other world spatiotemporally and causally isolated from our own. On another substrategy, fictional characters are abstract entities, such as abstract individuals (see van Inwagen 1977; Salmon 1998 among others), or abstract roles, kinds, or sets of properties (Wolterstorff 1980; Lamarque 1983). Each bears costs. The first substrategy requires acceptance of something akin to Lewis’s unpopular ontology (cf. Lewis 1978); moreover, it violates our sense that fictional objects are not the sorts of things that exist concretely at any world whatsoever. The second has the unpleasant result that we bear attitudes towards *abstracta* (such as pity) that seem appropriately borne only towards *concreta*.

A final strategy holds that when we use fictional names, directly or in the context of attitude-ascriptions, we engage in make-believe, pretending that the names refer when they do not. On such a view, sentences such as “Anna is sad” do not really express propositions, and sentences such as “I pity Anna” do not really express propositional attitudes, though we pretend that they do. While it is not literally true that Anna Karenina exists, it is true in the fiction that she does, and we can fruitfully speak (in certain contexts) as if that fiction were true (cf. Evans 1982; Walton 1990; Brock 2002.) This has the result that we do not have genuine emotional attitudes towards fictional characters. We discuss this issue in more detail below.

It’s beyond the scope of this chapter to evaluate which of these strategies is ultimately most successful. We thus put the special semantic puzzles raised by names of fictional characters to one side.

Traditional Formulations of the Paradox

As we have noted, the problem that gives rise to the paradox is that we seem to respond to fictional scenarios in two different ways: emotionally, we respond to them as

if they might be actual; cognitively, we respond to them as if they could not be actual. And, in general, we expect our cognitive and emotional responses to run in synchrony. So there is a mismatch between two of our responses in a circumstance where we expect no such mismatch.

While a number of related issues concerning our emotional responses to fiction are addressed in the writings of Plato and Aristotle some 2500 years ago, specific attention to this particular puzzle finds its first hints in the work of Samuel Johnson in the mid-18th century and Samuel Taylor Coleridge some 50 years later. (Plato 1992; Aristotle 1984; Johnson 1765; Coleridge 1817). But it was only in the late 20th century – with the publication of Colin Radford’s “How can we be moved by Anna Karenina?” and Kendall Walton’s “Fearing Fictions” – that the paradox was explicitly formulated in a way that captured widespread professional attention (Radford 1975; Weston 1975; Walton 1978); the result has been something of a cottage industry in (analytic) aesthetics. (For a comprehensive bibliography, see Hjort and Laver 1997.)

Although they do not state the paradox in these stark terms, with a bit of excavation and reconstruction, we can see that both Radford and Walton are concerned with some variant of the issue we have identified. So, for example, in his seminal 1975 “How can we be moved by Anna Karenina?” Colin Radford makes the following observation: “It would seem then that I can only be moved by someone’s plight if I believe that something terrible has happened to him. If I do not believe that he has not and is not suffering or whatever, I cannot grieve or be moved to tears” (Radford 1975, 68). At the same time, he notes: “we are moved by the death of Mercutio and we weep while knowing that no one has really died” (Radford 1975, 71). It is reasonable to think that the first of these quotes essentially articulates the Coordination Condition, while the second points out a case where we simultaneously satisfy the Response and Belief Conditions.

In the same way, one can map Kendall Walton’s early discussion of the puzzle onto our formulation. In his 1978 “Fearing Fictions” Walton notes that: “It would seem that real people can, and frequently do, have psychological attitudes toward merely fictional entities, despite the impossibility of physical intervention” (Walton 1978: 5-6) [Response Condition]. At the same time, he notes, when confronted by a cinematic representation of approaching green slime, the filmgoer “knows perfectly well that the slime is not real and that he is in no danger” (Walton 1978: 6) [Belief Condition]. But, Walton suggests, “it is plausible that” psychological attitudes such as “pity, worry about, hate, and envy are such that one cannot have them without believing that their objects exist, just as one cannot fear something without believing that it threatens them” (Walton 1978: 21, fn. 15) [Coordination Condition]. (Comparable mappings can be offered for the other highly influential formulations: see, for example, Lamarque 1981: 291; Carroll 1990: 62; Currie 1990: 187.)

Our Resolution: Overview

Walton and Radford propose solutions to the paradox that involve the denial of the Response Condition: Walton contends that fictional emotions are not genuine; Radford holds that they are not rational. Others have suggested that it is the Belief Condition that is false: when we respond emotionally to fictional characters, we lost track

of our belief that they are fictional. (We discuss these views in more detail below.) Our own resolution to the paradox involves endorsing the Response and Belief Conditions while denying the Coordination Condition. In so doing, it will be helpful to have a particular case in mind. So consider the classic example of feeling pity for Anna Karenina. (Radford 1975.) What we contend is that, in keeping with the Response Condition, the pity that we feel for Anna is both genuine and rational, even though, in keeping with the Belief Condition, we believe Anna to be a purely fictional character. (We are, recall, bracketing ontological issues concerning the precise content of this apparent attitude.)

There are reasons to find each of these claims suspect. After all, in ordinary cases when we feel pity for a (living) person, we are at least in-principle motivated to take some sort of action regarding her – but to the extent that we realize that Anna is purely fictional, we seem to feel no such motivation. This suggests that our actual emotions are sufficiently different from our fictional emotions to render them different in kind: the former are genuine; the latter are not. If, instead, we hold that both actual and fictional emotions are instances of genuine emotion, then the latter seem somehow defective: if emotional responses are intimately tied to motivation, then an emotional response whose object is rightly believed to be fictional is surely irrational. The only apparent way to avoid this conclusion is to hold that when we experience fictional emotions, we temporarily fail to believe their objects are purely fictional (contra the Belief Condition). If this line of reasoning is correct, then the Coordination Condition expresses a genuine constraint on our emotional responses.

Below, we discuss each of these objections in some detail. Our goal in this portion of the paper is to make a positive case for our own position. Our view finds its source in recent empirical research (by Antonio Damasio and others) showing that when we make practical decisions about our own futures, our reasoning is action-guiding only in cases where we imaginatively engage with potential consequences to produce emotional responses that are then somatically encoded – that is, that result in particular sorts of bodily changes. (We describe this research in the next section.) This suggests that, far from being exceptional, emotional responses to non-actual situations are a fundamental feature of our cognitive repertoire. Moreover, because of the role they play in underpinning practical reasoning (allowing us to act on our preferences by somatically encoding our evaluations of potential outcomes), it is crucial that they resemble actual emotions as precisely as possible. Together, these features suggest that it is legitimate to consider such emotional responses to be both genuine and rational. (We defend this claim in more detail below.) Moreover, we will contend, despite their resemblance to actual emotions, fictional emotions do not rest on a confusion in belief about what is merely fictional. As we will argue below, we tend to initially interpret all cognitive and sensory input as indicative of the presence of the ordinary source of phenomena of its type. Such instantaneous interpretations are not sufficiently robust to be properly considered beliefs. Together, these considerations suggest both that we should embrace the Response and Belief Conditions, and that we should reject the Coordination Condition.

The Damasio Results

Individuals with damage to their ventromedial prefrontal cortex display a typical set of behaviors. Most strikingly, they manifest extraordinary difficulties in day-to-day activities: Though they are able to articulate reasons for pursuing various courses of action, they are unable to use those reasons as bases for behavior, acting instead in ways that seem erratic and unplanned, and that are often counterproductive and anti-social. In addition, they exhibit a number of patterns that are detectable in laboratory settings. So, for example, while they are easily able to identify whether a photograph is “disturbing,” and even to articulate why (“its front paw is caught in a trap” “the bodies are piled atop one another”), they typically lack autonomic reactions to such emotionally distressing images. This contrasts strikingly with normal patients, who consistently display such responses. (Harris 2000: 85, describing Damasio *et al* 1991.)

A similar pattern of deficits reveals itself in a testing paradigm developed by Damasio’s team. Experimental subjects are given four decks of cards and a pile of play money, with the goal of maximizing their profit by turning over cards one at a time from any one of the four decks. Cards from the four decks are pre-assigned values so that for each card from deck A or B that the subject turns over, she earns a sizable reward, while for each card from deck C or D, she earns a significantly smaller reward. But decks A and B are also associated with high penalties, whereas decks C and D are not, so that playing with decks A and B results in a net loss, whereas playing with decks C and D results in a net gain. (Bechara *et al*, 1994).

When normal subjects are presented with this task, they initially sample from all decks, but eventually settle on decks C and D (the overall advantageous decks with lower immediate rewards), and their performance continues to improve over time. By contrast, when the task is presented to subjects with damage to the prefrontal cortex, results are strikingly different: they soon settle on decks A and B (the overall disadvantageous decks with higher immediate rewards); moreover, they show no improvement in performance over time.

As normal subjects successively experience the consequences of choosing from the high-risk decks, they begin to exhibit skin conductance responses in anticipation of such selections. Soon afterwards, they begin expressing a “hunch” that these decks are more risky, and begin avoiding them in favor of the lower-risk decks; ultimately, many of them are able to articulate the basis for this avoidance. By contrast, subjects with damage to the prefrontal cortex exhibit no such skin conductance responses in anticipation of their high-risk deck selections, and no tendency to avoid such decks in their own selection process. This is so even when they are able to articulate conceptually the relative risks involved. (Summary based on Harris 2000: 86-7)

What Damasio, Bechara and others have concluded on the basis of this research is that autonomic responses play a central role in practical reasoning. Some sort of somatic realization of the potential consequences of a risky action seems crucial to prudent decision-making. Without it, the theoretical advantages of one or another course of action may be apparent, but these will not translate properly into action-guiding behavior.

This research seems to show that our ability to engage in practical reasoning rests on the following sort of process: We imaginatively engage with the potential consequences of various courses of action, thereby activating our emotional response

mechanisms, and we encode the results of these simulations somatically; the presence of these “somatic markers” then helps to guide our future behavior. Call these emotions “simulated emotions.” It is clear that simulated emotions are a fundamental feature of our cognitive repertoire. It is also clear that there are striking resemblances between *simulated* emotions and *fictional* emotions, so much so that if we can establish that simulated emotions are both genuine and rational, then we will have done most of the work required for establishing the faultiness of the Coordination Condition.

Genuineness

There are at least two dimensions along which fictional and simulated emotions differ from actual emotions that seem relevant to whether the former are rightly considered genuine. Whereas the apparent objects of actual emotions are actual individuals and events, the apparent objects of simulated and fictional emotions are ostensibly non-actual. And whereas actual emotions feed directly into behavior in certain predictable ways (we move away from objects that we fear, and intervene on behalf of individuals that we pity), fictional and simulated emotions are not directly tied to action in this fashion.

Famously, Kendall Walton has contended that these differences mean that we do not feel genuine emotional responses to fictional scenarios. Instead, he contends, we experience phenomenally indistinguishable *quasi-emotions* that differ from genuine emotions along two dimensions: in contrast to genuine emotions, they do not require that we be existentially committed to their apparent objects; and they are not intimately connected to motivation and action. (Walton 1990; an alternative view is presented in Walton 1997.) So, for example, he writes concerning the first requirement:

Grief, as well as pity and admiration, would seem to require at the very least awareness of the existence of their objects. It is arguable that for this reason alone appreciators cannot be said actually to pity Willy [Loman] or grieve for Anna [Karenina] or admire Superman. (1990: 204)

And, concerning the second:

Fear is *motivating* in distinctive ways, whether or not its motivational force is attributed to cognitive elements in it... To deny this, to insist on considering...[a] nonmotivating state to be one of fear of [its purported object] would be to reconceive the notion of fear. Fear emasculated by subtracting its distinctive motivational force is not fear at all. (1990: 201-2)

Regarding the first quotation, there is the risk of degenerating into terminological debate. If one defines “genuine” (or “actual”) emotion so that such emotions “require...(awareness of) the existence of their objects,” then Walton is surely correct; the substantive question is whether such a restriction effects a natural cut in conceptual space. (We here ignore the difference between non-existence and non-being; see **Fictional Characters** above.) An argument can be made that there is a continuum of cases – from cases where the object of the emotion is an entity that exists only in the past

(where I pity someone who has died), to cases where the object of the emotion is a situation that may or may not occur in the future (where I fear a stock market crash), to cases where the object of the emotion is an entity that may or may not exist in the future (where I pity the oldest daughter of my great-grandson), to cases where the object of the emotion is a situation that is explicitly fictional (where I fear the flood that may drown the inhabitants of Alpha Centauri), to cases where the object of the emotion is an entity that is explicitly fictional (where I pity Anna Karenina). We have no inclination to withhold attributions of genuineness from cases involving past or future or merely-possible persons or events (“She doesn’t *fear* a stock-market crash she just *quasi-fears* it”...?). It thus seems that we do not require that the target of a genuine emotion exist in the here and now. What, then, would incline us to withhold such attributions in the case of explicitly fictional persons and events? The source seems to be either worries about empty names (“she doesn’t pity Anna Karenina – there is no Anna Karenina”) or related worries about misattribution (“what he fears isn’t the bear in the closet – there is no bear there – what he fears is the sound of the wind...”). We are setting aside worries of the first sort. And it is interesting to note that worries of the second sort dissipate somewhat when we are explicitly aware of the fictionality of our character. When someone (apparently) feels pity for Anna Karenina, knowing full well that Anna is a fictional character, we are not inclined to think that he has misidentified the target of his emotion. In short, unless we stipulate at the outset that attitudes such as fear and pity can take as their targets only certain sorts of entities, then, assuming a standard picture of what exists and what does not, and setting aside legitimate worries about empty names, it seems arbitrary to insist that they are genuine only when their objects exist.

Regarding the second quotation from Walton, it is far from clear that fictional and simulated emotions differ from actual emotions to the degree that Walton seems to be suggesting. Note first that in order for the reasoning process described above to operate effectively, our simulated emotions and our actual emotions must line up as closely as possible: otherwise, the process of considering alternative outcomes would not give us proper information about how we would respond once one of those outcomes became actualized. So simulated emotions and actual emotions should be, in a well-functioning person, as similar as possible. This non-accidental similarity provides grounds for considering simulated emotions to be genuine, and insofar as fictional emotions exploit similar mechanisms, it provides parallel grounds in that case. (It is in this regard that part of the truth about fictional emotions is captured by the position known as “Factualism,” according which the objects of our emotional responses to fiction are actual people in situations literally or metaphorically akin to those described in the fiction. For various versions of this position, see Paskins 1977, Johnson 1765, Weston 1975, McCormick 1988.)

Moreover, if Damasio is right, both simulated and fictional emotions produce bodily changes akin to those produced by actual emotions. That the latter feed directly into action whereas the former feed only indirectly into action can be traced, we propose, to a difference in processing, not in motivation (modulo differences in vivacity of stimulus). An alarm clock set five minutes fast can motivate us to rise, even if we are fully cognizant of the misinformation it gives; we are reluctant to pull the trigger in “Polish roulette,” even when we are certain that the gun contains no bullets. (See also our discussion of optical illusions in the next paragraph.) So while Walton is surely correct to

note that fictional emotions do not feed into behavior in the ways that actual emotions do, it does not follow that they do not have similar motivational structure.

But this line of thought brings out a certain tension in our view. We have been emphasizing the similarities among fictional, simulated and actual emotions. One elegant explanation of these similarities would be that we momentarily lose track of the non-actuality of the simulated and fictional stimuli when we respond emotionally to them. If so, then perhaps it is the Belief Condition that is at fault (insofar as the Response Condition holds. (This is the explanation to which advocates of the position known as the “Momentary Confusion” or “Suspension-of-Disbelief” view.)

As before, there is a risk of turning a substantive dispute into a terminological one: if “belief” is used thinly enough, then the Belief Condition may indeed be inapplicable as stated. The substantive question is whether there are other reactions to non-standard stimuli where we are not inclined to say that the similarity between ordinary and divergent cases rests on a false belief. We think there are many. In cases of optical illusions, for example, we may perceive a bent stick as being straight – or perceive the two lines in the Muller-Lyer illusion as being of different lengths – without *believing* that things are as they seem. If we stand near the edge of a high glassed-in platform, we may recoil slightly, without *believing* that we are at risk of falling. One explanation for this and other such cases is that we respond to nearly all cognitive and sensory input as being indicative of the presence of its ordinary stimulus-source. (So, for example, objects exhibiting the retinal-stimulation pattern of the water-embedded stick are, in ordinary cases, objects that are bent; situations exhibiting the retinal-stimulation pattern of the glassed-in high platform are, in ordinary cases, dangerous.) But precisely because these initial response patterns are so evidence-resistant, there is good reason to think that they are subdoxastic. If a response-pattern cannot be changed in reaction to the presentation of reasoned evidence (we cannot “talk ourselves out of” optical illusions), it seems misleading to categorize it as belief-involving. So the similarities between simulated and actual responses need not be seen as impugning the Belief Condition.

In sum, there are similarities and dissimilarities between fictional and actual emotions. Whether the differences are sufficient to warrant referring to them with distinct expressions is to some extent a terminological dispute. The substantive issue concerns what forms these similarities and dissimilarities actually take. It is our contention that the similarities are more striking than the differences.

Rationality

In a series of some dozen articles over nearly a quarter of a century, Colin Radford has contended that fictional emotions are irrational. As with Walton, part of our dispute with Radford may be terminological. Radford writes:

What is necessary for the occurrence of these [emotional] responses is missing when the objects which elicit them are (believed to be) fictional. There is then literally, nothing to be concerned about, no one—indeed nothing—to pity.... However natural, almost universal, such responses are, we can and do come to see that they are irrational, do we not? (Radford 1989: 96)

These remarks are somewhat perplexing: if “what is necessary for the occurrence of these [emotional] responses is missing” then it is hard to see how such responses could be “almost universal.” If, with Walton, Radford is suggesting that the very concept of emotional response precludes our responding emotionally to the non-actual, then our reply is as above. But this does not seem to be his primary worry. Rather, the source of discomfort for Radford seems to be the thought that fictional emotions somehow involve us in competing, even contradictory, commitments. He writes: “we are frightened for ourselves of characters we know to be fictional and are irrational, incoherent, and inconsistent in being thus frightened” (Radford 1995: 75).

By employing the terms “incoherent” and “inconsistent,” Radford seems to be suggesting that in feeling fictional emotions, we reveal ourselves to be holding contradictory beliefs. The thought behind such an analysis might be the following: when we respond with genuine emotion to a character, or describe someone as responding with genuine emotion to a character, we reveal that we believe that character to be actual; but when we believe a character to be fictional, we believe that character not to be actual; so when we respond with genuine emotion to a character that we believe to be fictional, we believe that character to be actual and not to be actual. But this reasoning, as we have argued at the end of the last sub-section, is flawed: it is a mistake to think that feeling genuine emotions requires a belief, temporary or otherwise, in the actuality of its purported target.

Alternatively, perhaps Radford’s worry is that in responding to fictional characters with genuine emotion, we violate a norm that we tacitly hold: that we should respond emotionally only to things that we believe to be actual. The irrationality of our behavior would thus consist in repeatedly violating a principle that we reflectively endorse (in the same way that it would be irrational for someone regularly to eat meat if she had a principled commitment to vegetarianism.) But it is far from clear that we *do* (or *should*), on reflection, subscribe to such a principle. If, as we have suggested above, simulated emotions play a central role in allowing us to make and act on decisions about our future well-being, then far from impeding our capacity to act as agents who make effective use of means-ends reasoning, they contribute directly to it. If so, it is hard to see why we would want to endorse a principle telling us that we should respond emotionally only to things that we believe to be actual (even if, as a practical matter, this ideal proves unattainable).

But a residual issue remains about whether we would want to endorse a weaker principle according to which we should not respond emotionally to things we explicitly believe to be fictional. After all, one might contend, fictional emotions do not seem to play the same direct role in our capacity for practical reasoning that simulated emotions do, so we should aim not to feel them. But it is not clear the even this weaker line of argument can be successfully maintained. One line of thought, stemming from Aristotle and emphasized more recently by thinkers such as Susan Feagin and Martha Nussbaum, stresses the instrumental role of fictional emotions in the cultivation of moral and intellectual character. (Aristotle 1984; Feagin 1983; Nussbaum 1986, 1990.) By engaging emotionally with fictional characters and situations, we broaden our range of simulated encounters, gaining insights about others’ experiences that are processed much as if they had been our own. Without such a capacity, actual experience would be our only source of such emotional encounters, severely limiting the range of our reactive possibilities. So

fictional emotions may contribute to our capacity for rational action through the role they play in educating our sensibilities. If so, then there is little reason to think that we should endorse a categorical principle according to which we would, ideally, fail to feel such emotions.

In saying this, we are not denying that there are plenty of circumstances where such emotional responses are very much out of place. Cases where emotional investments in fictionality exceed corresponding investments in actuality are exemplary instances of irrationality (as both psychoanalytic theory and common sense remind us). And cases where prop-based pretense produces responses that, given ordinary practice in such games of make-believe, are unprompted are likewise paradigmatically irrational (as, for example, when a child feels genuine fear that her stuffed animal will catch a cold). But the existence of such cases does not impugn the possibility of rational genuine emotional responses to certain sorts of fictional scenarios. (Though they do reveal that our assessments of rationality and irrationality are, here as elsewhere, governed by conventional norms of appropriateness.)

Conclusion

As we have formulated it, there are three basic ways to resolve the Paradox of Fictional Emotions, by rejecting the Response Condition, the Belief Condition, or the Coordination Condition.

Those who reject the Response Condition deny that we have genuine and rational emotional responses to fictional characters or situations. But grounds for rejecting the position vary widely. Those who endorse the *Quasi-Emotion Theory* and *Irrationalism* hold that we do have emotional responses to fictional characters and situations, but that these responses are either *non-genuine* (Quasi-Emotion Theory) or *irrational* (Irrationalism). Those who endorse *Factualism* or *Non-Intentionalism* hold that the, contrary to appearances, we do *not* have emotional responses to fictional characters and situations; rather, our fictional emotions are actually directed at real-world analogues of their apparent targets (Factualism), or are not directed at anything, instead being diffuse, objectless moods (Non-Intentionalism). We have discussed the first and second of these (*Quasi-Emotion Theory* and *Irrationalism*) in some detail above. Regarding both the third (*Factualism*) and the fourth (*Non-Intentionalism*), we note that while they provide plausible explanations for certain cases, they do not seem to have the generality required for a full solution. In too many cases, there is no plausible real-world analogue to serve as the requisite surrogate. And in too many cases, the emotional response is far too focused to be classified as merely a mood.

Those who reject the Belief Condition hold that when we respond emotionally to fictional scenarios, we do not believe them to *be* fictional. So the similarity between our emotional responses to actual and non-actual scenarios can be traced to a certain sort of (albeit temporary) confusion. Although this view brings out the importance of recognizing similarities among actual, simulated and fictional emotions, as we noted above, it overstates its case in claiming that these similarities arise from a false belief about the fictionality of the scenarios in question.

Those who reject the Coordination Condition allow that we can respond with genuine, rational emotions to targets that we believe to be fictional. Above, we have offered reasons for holding such a position.

References

- Aristotle. (1984). The Rhetoric and the Poetics of Aristotle. (R. Roberts & I. Bywater, Trans. & E. P. J. Corbett, Intro.). New York: McGraw Hill.
- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. Cognition, 50, 7-15.
- Brock, S. (2002). Fictionalism about fictional characters. Noûs, 36, 1-21.
- Carroll, N. (1990). Philosophy of horror. New York: Routledge.
- Coleridge, S. T. (1817). Biographia literaria. London: Rest Fenner.
- Crittendon, C. (1991). Unreality. Ithaca: Cornell University Press.
- Currie, G. (1990). The nature of fiction. Cambridge: Cambridge University Press.
- Damasio, A. R., Tranel, D., & Damasio, H. (1991). Somatic markers and the guidance of behavior: Theory and preliminary testing. In H. S. Levin, H. M. Eisenberg & A. L. Benton (Eds). Frontal Lobe Function and Dysfunction (pp. 217-229). New York: Oxford University Press.
- Damasio, A. R. (1997). Descartes' error. New York: Grosset/Putnam.
- Damasio, A. R. (1999). The feeling of what happens. New York: Harcourt.
- Evans, G. (1982). The varieties of reference. Oxford: Oxford University Press.
- Feagin, S. (1983). The pleasures of tragedy. American philosophical quarterly, 20, 95-104.
- Harris, P. L. (2000). The work of the imagination. Oxford: Blackwell.
- Hjort, M. & Laver, S. (Ed.). (1997). Emotion and the Arts. New York: Oxford University Press.
- Johnson, S. (1765). Mr. Johnson's preface to his edition of Shakespear's plays. In S. Johnson (Ed). The Plays of William Shakespeare (pp. A3r-D4r). London: J. and R. Tonson and others.
- Lamarque, P. (1981). How can we fear and pity fictions? British journal of aesthetics, 21, 291-304.
- Lewis, D. (1978). Truth in fiction. American philosophical quarterly, 15, 37-46.

- McCormick, P. J. (1988). Fictions, philosophies, and the problems of poetics. Ithaca: Cornell University Press.
- Nussbaum, M. (1986). The fragility of goodness. New York: Cambridge University Press.
- Nussbaum, M. (1990). Love's knowledge. New York: Oxford University Press, 1990
- Parsons, T. (1980). Nonexistent objects. New Haven: Yale University Press.
- Paskins, B. (1977). On being moved by Anna Karenina and *Anna Karenina*. Philosophy, 52, 344-347.
- Plato. (1992). Republic. (G. M. A. Grube, Trans. & C. D. C. Reeve, Rev.). Indianapolis: Hackett.
- Radford, C. (1975). How can we be moved by the fate of Anna Karenina? Proceedings of the Aristotelian Society, supplemental volume, 49, 67-80.
- Radford, C. (1989). Replies to three critics. Philosophy, 64, 93-97.
- Radford, C. (1995). Fiction, pity, fear, and jealousy. Journal of aesthetics and art criticism, 53, 71-75.
- Salmon, N. (1998). Nonexistence. Noûs, 32, 277-319.
- van Inwagen, P. (1977). Creatures of fiction. American philosophical quarterly, 14, 299-308.
- Walton, K. L. (1978). Fearing fictions. Journal of philosophy, 75, 5-27.
- Walton, K. L. (1990). Mimesis as make-believe. Cambridge: Harvard University Press.
- Walton, K. L. (1997). Spelunking, simulation, and slime: on being moved by fiction. In M. Hjort & S. Laver (Eds). Emotion and the Arts (pp. 37-49). New York: Oxford University Press.
- Weston, P. (1975). How can we be moved by the fate of Anna Karenina? Proceedings of the Aristotelian Society, supplemental volume, 49, 81-93.
- Wolterstorff, N. (1980). Works and worlds of art. Oxford: Clarendon Press.

Suggested Further Reading

Matravers, D. (1998). Art and emotion. Oxford: Clarendon Press.

Thomasson, A. L. (1999). Fiction and Metaphysics. Cambridge: Cambridge University Press.

Yanal, R. J. (1999). Paradoxes of emotion and fiction. University Park, PA: Pennsylvania State University Press.

TAMAR SZABÓ GENDLER
KARSON KOVAKOVICH

Notes on Contributors

Tamar Szabó Gendler is an Associate Professor of Philosophy at Cornell University. Her work focuses primarily on issues in epistemology, philosophical psychology, metaphysics and aesthetics. She is currently working on a cluster of problems surrounding the relation between imagination and belief.

Karson Kovakovich is a doctoral student in the Philosophy Department at Rutgers University. His research interests are in epistemology, metaphysics, and philosophy of mind.