

*Making Sorrow Sweet: Emotion and Empathy
in the Experience of Fiction*

Alison Denham

And so in regard to the emotions . . . and all the appetites and pains and pleasures of the soul which we say accompany all our actions, the effect of poetic imitation is the same. For it waters and fosters these feelings when what we ought to do is to dry them up, and it establishes them as our rulers when they ought to be ruled, to the end that we may be better and happier men instead of worse and more miserable ones. Plato, *Republic* 606d1–7¹

Relations between experimental psychologists and analytic philosophers have often been marked by mutual distrust and misunderstanding. In recent decades, however, a new agenda has developed of collaborative theorizing. The resulting sub-disciplines of ‘experimental ethics’ and ‘experimental aesthetics’ have refined and progressed both the evaluative dimensions of naturalistic psychology and the naturalistic ones of evaluative philosophical theory. In principle, the claims of empirical and conceptual investigations should dovetail: psychological accounts of target phenomena should complement, rather than compete with, their philosophical counterparts. For theories of the origins, nature and value of affective engagement with fiction, however, this ideal has proven elusive. Psychological studies have focused principally on the causal mechanisms explaining our affective interactions with fictions, prescinding from questions concerning their rational justifiability. Analytic philosophers, by contrast, have often struggled to move beyond those questions, and to overcome longstanding doubts about fiction’s wider epistemic value. The result has been a theoretical impasse in which the power of fiction to ‘transport’ a reader is at once often lauded (by psychologists) as a privileged route to interpersonal understanding, and condemned (by philosophers) as an abdication of the authority of reason. This chapter surveys some of the central claims on both sides, tracing the source of the

debate, at least in part, to competing conceptions of rationality. I will focus primarily on emotional and empathic responses to fiction, but much of the argument extends to affective states more generally.

I begin with a brief review of some psychological findings about emotional transportation in fiction-reading and its relation to empathic responsiveness. I then turn to the so-called ‘paradox of fiction’ as addressed by philosophers, and the proposal that emotions targeting fictions are themselves fictional or ‘make believe’ gestures in a game of pretence. I argue that the allure of the putative paradox is dispelled by a more nuanced distinction between rational and irrational emotions; the norms of rationality governing our emotional responses to fictional characters are those appropriate to experience-based evaluations, rather than beliefs. As such, these responses – like our evaluations of expressive properties – are answerable to non-inferential norms of justification. The chapter concludes with some related observations on the prudential rationality and value of our emotional immersion in fictional portrayals of aversive experience.

Emotional Transportation: How Does Fiction Make Us Feel?

[Fiction] . . . is a form of consciousness that can be passed from one mind to another . . .

Keith Oatley²

Much recent research in the psychology of literature has focused on the mechanisms of affective and imaginative immersion in narratives, exploring both how it occurs and what are its wider effects. The central construct of this research agenda is the phenomenon of ‘emotional transportation’, whereby persons ‘become emotionally involved, immersed, or carried away imaginatively in a story’.³ The construct itself is ambiguous and ill-defined, and no consensus about its proper measurement exists within the literature. Nonetheless, Transportation Theory has delivered several striking, replicated findings. I will mention and comment on four of these here.

The first finding is that readers are more significantly transported by fictional narratives (e.g., fictional stories and novels) than by factual ones (e.g., journalistic accounts).⁴ Being ‘transported’, as the term is typically used, involves not only affective responsiveness but imaginative tracking of sense-based descriptions and attunement to both explicit and implicit evaluative attitudes expressed by a narrative. With respect to readers’ specifically emotional responses to fiction, the more transportation that occur[s] in reading a story, ‘the greater the *story-consistent* emotional

experience has been found to be'.⁵ It is perhaps unsurprising to find that fictional narratives are significantly more powerful transporters in this regard than journalistic articles (the usual control literature) when these are matched for general topic and valence. Even randomized studies which controlled for subjects' usual reading preferences delivered this result: fiction reliably elicits heightened transportation over non-fiction, even across subjects who normally prefer the latter. The possible reasons for the enhanced transportative effects of fiction are many, and will be familiar to literary theorists. For instance, fictional narratives have the liberty of describing events as from the subjective, first-personal points of view of its characters, in a way that factual narrative cannot (save perhaps through interview and third-party reports). In keeping with this 'licensed subjectivity', fictions also enjoy a freer hand at describing first-personal sensory, affective, motivational, and interpersonal experiences; details of such content are a natural way of constructing the internal perspective fictions often adopt. Additionally, because fictions do not allow the reader to affect the outcome of a narrative through his own actions, one common motivation for resisting transportation and remaining distant from characters is eliminated, namely, a felt obligation to intervene.⁶

A second finding concerns the psychological effects of such transportation: fiction reading more significantly than non-fiction enhances readers' wider empathic, affective and perceptual Theory of Mind skills.⁷ This result is robust across a wide variety of instruments and measurements, including The Mind in the Eyes test, narrative completion tasks and autonomic measures of excitation.⁸ Moreover, it holds for both episodic, occurrent empathy (tested during and after reading) and for sustained, dispositional empathy (tested up to a week later). Early studies probing this hypothesis were vitiated by failing to control for subjects' usual reading habits; it was thus impossible to tell whether more empathic subjects tended to read fiction, or reading fiction had made them more empathic. More recent randomized studies controlled for this, and also ensured that subjects were exposed to the *same* fictional and non-fictional material over the same period of time. The results were sustained: even following an incubation period of one week, the fiction-reading subjects scored significantly higher on empathic facility.⁹ In a sense, this is to be expected, given that empathic facility is very nearly built into the definition of 'transportation', higher transportation is likely to correlate with higher empathy (and regularly does) for fiction-reading.

However, a third finding is important in this context. When *non-fictional* texts concerning the fortunes of real people and events elicit strong emotions, this tends to lead to lower empathy. So, for instance, higher

transportation levels often correlated with lower subsequent empathy in the case of non-fiction narratives of distress, e.g., journalistic accounts of suffering refugees. This result lends support to the idea that factual narratives can sometimes 'block' empathic responses by indirectly imposing obligations of practical action and eliciting heightened personal distress, well-known to be inversely correlated with other-directed empathic concern.¹⁰

A fourth finding is perhaps the most surprising: literary fiction is more efficacious than popular fiction in facilitating both transportation and empathy, exercising a more significant and more lasting influence on subjects' sensitivities to others' affective states.¹¹ Why should that be? After all, the category of 'middle-brow' fiction is more popular, and if readers seek out narratives in part to enjoy their transportative capacities, then 'high-culture' novels ought to be in greater demand. There are, however, several other variables in play that may explain its limited market appeal. For instance, high-culture literature is often informed by social, aesthetic, cultural and linguistic environments that differ significantly from those of the wider reading public. There are also many formal and structural features of literary fictions that are likely to facilitate transportation positively – properties remarked by both literary theorists (for instance, Mikhail Bakhtin and Roland Barthes) and psychologists. For instance, literary writing tends to demand more of readers, engaging them in a discourse that forces them to fill in gaps and search for 'meanings among a spectrum of possible meanings'.¹² While both popular and literary fiction enjoy the licensed subjectivity mentioned above, the latter more commonly relies on presupposition and indirect allusion (rather than explicit statement) and more often introduces multiple perceptual and evaluative perspectives. As Jerome Bruner has observed, the skills required to interpret information presented in these ways actually mimic those required for affective empathy and imaginative exercises in mindreading.¹³

These four results all directly or indirectly lend support to an empirical hypothesis at the heart of transportation theory, namely, that attentively following a well-composed fictional narrative instigates a process of *simulation*: the reader tracks the narrative experientially, mirroring its descriptions through first-personal perceptual imaginings, affective and motor responses and even evaluative beliefs.¹⁴ Simulation effects in each of these categories (perceptual, affective and motor) have been extensively explored, using behavioural measures, autonomic measures such as heartrate and skin conductance, and functional magnetic resonance imaging (fMRI).¹⁵ By every measure, vivid, high-imagery, semantically and structurally

complex literary descriptions most effectively elicit the predicted first-personal correlates. Even where subjects' self-reports did not reveal the predicted mirroring effects, these were often evidenced physiologically. When Keith Oatley refers to fiction as 'the simulation of selves in interaction', he does not overstate the case.

This brief outline of transportation theory indicates the approach to fiction reading now favoured by experimentalist psychologists. The operative framework recognizes affective responses to fiction (and especially fictional characters) as valenced, pre-rational, psycho-physical responses, both occurrent and dispositional, that are naturally elicited by fictional, and especially literary, descriptions. It is significant that these responses are closely associated with heightened empathy, not least because this suggests one reason that the practice of fictional discourse has evolved. Empathic attunement to our conspecifics is a nature-given, adaptive capacity, of first importance not only to successful social coordination but to many of the other-regarding prescriptions at the core of our human morality. What promotes empathy thus stands also to promote societal harmony. For this and other reasons, Frank Hakemulder describes fiction as our 'moral laboratory' and as an indispensable civilizing force by which we bind ourselves to others, forming a common moral community.¹⁶ The idea that responding emotionally to fiction might be irrational, let alone immoral, is wholly alien to this research framework. That thesis has been left to philosophers.

Emotions and Reasons

[W]e do not feel horror because we are threatened by a sphinx; we dream of a sphinx in order to explain the horror we feel.

Jorge Luis Borges, citing Coleridge¹⁷

Plato famously puzzled over the attractions of literary discourse (or 'poetry') in his *Republic* and other dialogues. His objections to literary devices are notoriously complicated and various, and do not target fiction alone.¹⁸ However, through them all runs the thread of his 'epistemology of affect' – his view that affective 'passions' are deleterious to our epistemic aims. Plato held that the passions are at best indifferent to the truth, and at worst systematically hostile to our epistemic interests. He combined this premise with two further observations: first, that literary uses of language (and of dramatic mimesis, specifically) tend to invite and arouse affective responses; secondly, that this arousal of emotion is alluring to us. Our love of literary and other artistic illusions is owed to the fact that they appeal to

‘the emotions . . . and all the appetites and pains and pleasures of the soul’.¹⁹ For Plato, this was no good thing; men at their best are truth-seekers, guided by their capacity to reason. To be ruled by one’s capricious, undiscerning passions is to be psychologically enslaved, and fictions are ruthless gaolers: they nourish the irrational parts of the soul, systematically undermining our capacity for reason and making us ‘worse and more miserable’ men.²⁰ Plato thus addresses both the question of whether emotional engagement with fiction is rational (it is not) and that of its value or disvalue (it frustrates the aims of reason, undermining our highest interest: knowledge of the Good).²¹

Almost no theorist today shares Plato’s wholesale skepticism about the epistemic perils of affectivity as such. On the contrary, in contemporary epistemology and moral epistemology in particular, affectivity (and especially the emotions) have moved to centre-stage in many naturalistic theories of the nature and origins of moral, aesthetic, and other evaluative knowledge.²² Perhaps Plato was right to observe that literary fictions ‘water and foster’ the emotions, but it does not follow that they are inherently hostile to our epistemic goals. Nonetheless, the worry has persisted that when affective states with intentional content – typically, the emotions – target fictive entities, they are irrational. This worry has been a focus of analytical aesthetics for more than half a century, taking shape as the so-called Paradox of Fiction. As formulated by Colin Radford, the paradox is produced by combining three, individually plausible premises:²³

- P1. We experience genuine emotions directed at fictional characters and situations.
- P2. To experience an emotion towards something, one must believe that thing exists.
- P3. We do not believe that fictional characters and situations exist.

This ensemble of claims is inconsistent in the classical sense of entailing a contradiction: P2 and P3 entail the negation of P1. Endorsing jointly inconsistent claims is irrational, if anything is: any reader of whom all three are true (that is, who has genuine, existence-presupposing emotions directed at fictions whilst failing to believe their targets exist) is, the worry goes, guilty of ‘inconsistency and incoherence’. Radford, for one, embraced this unpalatable conclusion.²⁴ The three claims, he argued, accurately describe the practice and psychology of our emotional responses to fictional entities; we are adopting contradictory commitments when we fear Frankenstein’s ire, hope for Raskolnikov’s redemption, grieve for Anna Karenina’s shame and cheer Jane Eyre’s final words (‘Reader, I married him’). We are then in

much the same position as one who claims ‘Socrates was mortal and Socrates was not mortal’, or ‘It is raining here and now and it is not raining here and now.’

Radford’s diagnosis sits uncomfortably with the psychological framework of contemporary transportation theory, which regards fictional engagement as facilitating our capacity accurately to perceive others’ psychological, and especially their affective states. It also falls foul of everyday critical practice which likewise holds that affective engagement with a fictional work – and specifically *emotional* engagement with its characters – is as much a part of understanding a work as of enjoying it aesthetically. Within philosophical meta-ethics, for instance, recent decades have seen a rebirth of interest in the moral and psychological insights afforded by literature, and its epistemic value more generally.²⁵ Virtually all cognitive theories of literature regard our first-personal, emotional responses to fictional characters as not merely rationally permissible but indispensable to that value. It would be odd indeed if the knowledge fictions have to offer could be grasped only by those prepared to embrace a *reductio ad absurdum*.

The preferred path out of the paradox has been to reject one or more of the premises producing it. Premise 3 is a conceptual, and perhaps even an analytic truth, which makes it an unpromising target of criticism.²⁶ For this reason, challenges to Premises 1 and 2 have largely dominated the debate, producing what I shall call the ‘Pretence Theory’ and the ‘Imagination Theory’ respectively. I discuss these in turn.

Pretence Theory

One of the most influential analyses of the Paradox of Fiction is owed to Kendall Walton. It is of a piece with his wider treatment of engagement with the mimetic or representational arts, according to which this constitutes a mode of *pretence*. That wider theory characterizes representational artworks as ‘props’ which, by long-established conventions as well as certain natural qualities, come to prescribe specific imaginings, much as do children’s props in games of make-believe. (One of Walton’s early inspirations was Ernst Gombrich’s ‘Meditations on a Hobby Horse’.²⁷) Walton proposes an ontological distinction between the actual world and the various ‘fictional worlds’ which such games generate, and further demarcates the ‘game world’ of individual participants and that of the artwork itself, the content of which is circumscribed by its internal features. Artworks-as-props come in two varieties: sensory depictions (such as

programme music, paintings and sculptures) and verbal representations (verbal narratives as in stories and novels, and the linguistic constituents of film and theatre).

Applying this theory to the Paradox of Fiction, one might expect Walton's target to be Premise 2 (that genuine emotions presuppose belief in their objects). That is not, however, how he proceeds; indeed, Walton took Premise 2 to articulate a 'principle of commonsense' that 'ought not to be abandoned if there is any reasonable alternative . . .'²⁸ Instead, in his now-classic paper, 'Fearing Fictions', Walton proposes a distinction between ordinary or conventional emotions and those targeting fictional entities. The latter are 'quasi-emotions', manifesting much of the same first-personal phenomenology as ordinary emotions, but functionally identified as make-believe experiences. That quasi-emotions are distinct from conventional, garden-variety ones, and are recognized as such, Walton argues, is shown by the fact that they are motivationally inert, i.e., they do not move us to action: we do not attempt to stay Othello's hand before he slays Desdemona, nor do we flee the cinema as the zombies approach. On Walton's view, this is because we recognize that in the make-believe activity of reading and responding to fictions, the objects of quasi-emotions no more demand action than a child's mud pies demand eating. Walton's account thus acquits the fiction reader of Radford's charge of irrationality by stipulating a new psychological type, the quasi-emotion, which abjures the troublesome existential commitments and practical import he takes to feature in ordinary emotions.

Does Walton's Pretence Theory deliver an accurate phenomenology of our everyday experience of fictions? There is certainly something right in his suggestion that engaging with fictional representations involves a kind of make-believe and willing suspension of disbelief – a tacit agreement to cooperate with and track a work's imagined narrative. But it does not follow that the emotions we then bring to bear *themselves* become part of what is imagined – states of a distinctive psychological kind – mere make-believe responses. It certainly does not seem, first-personally, that when one grieves with Shakespeare's *Lear* or feels dismay at the injustices suffered by Thomas Hardy's *Tess*, one is pretending to experience these emotions. Moreover, the psychological evidence of subjects' behavioural dispositions, autonomic responses and neurological activations overwhelmingly indicates that emotions had in response to fictions are psychologically and physiologically manifested in the same way as everyday ones.

Neither do fiction-elicited emotions differ from everyday ones in other important respects. Consider the relation of emotions to agentic control: if

quasi-emotions were part of a willing game of make-believe, we ought to be able to opt in or out of them at will – to simply shut off our emotional responses by leaving the game (putting the book aside or exiting the cinema). This is not typically possible. As Noel Carroll observes, ‘if . . . [the fear produced by chiller films] were a pretend emotion, one would think that it could be engaged at will. I could elect to remain unmoved by *The Exorcist* . . .’²⁹ Secondly, no distinction between authentic and quasi-emotions can plausibly rest on the criterion Walton proposes, viz., motivational engagement. We are regularly subject to motivationally inert emotions in countless real-world contexts. Through memories, we respond but do not act on past, recollected events (the joy and relief of finishing one’s examinations, the grief and aloneness following a parent’s death). Likewise, through imaginings, we have non-motivating but genuine hopes, fears, etc. for our own and others’ future selves. Sometimes motivation is absent because time and distance make action impractical or too difficult, as when one sympathetically hears of a faraway disaster. On other occasions, however, we simply experience an emotion as ‘free-standing’, disengaged from our practical deliberations. It is far from obvious that action-intentions are an internal feature of emotions themselves, rather than a common response to them. If they are not, then the principal motivation for stipulating an independent category of quasi-emotions collapses. In that case, Pretence Theory offers little more than an *ad hoc* stipulation of a novel psychological kind – the only virtue of which is that it avoids the Paradox of Fiction. As Noel Carroll remarks, ‘Walton’s theory appears to throw out the phenomenology of the [affective] state for the sake of logic’.³⁰ If one’s sole motivation for positing an independent psychological kind is to avoid a theoretical paradox, then that seems a good reason to question the basis of the theory itself.

Emotion and Empathy in Fiction: Reasons and Causes

The failure of Pretence Theory is instructive as a caution against rejecting the authenticity of emotions experienced in response to fictional entities – the first premise of Radford’s troublesome trio. A better candidate is Premise 2, requiring an existential commitment to the objects of our beliefs. Transportation Theory, for its part, does not hesitate to abandon this premise. Indeed, psychologists typically conceive of emotions as experiences, not beliefs, and experiential contents as such need entail no existential commitments. This also seems to be true of many familiar instances of emotions. Consider the jealous husband’s wave of rage when he even *entertains* the idea of his (faithful) wife’s infidelity, or the intimacy enjoyed by the bereaved parent who converses with a photo of his deceased

child. These are both cases of genuine, episodic emotions: valenced, affective states with a distinctive phenomenology, identified and individuated by evaluative, intentional contents. In general, one need not believe that an object exists in order to be afraid of it, or saddened by it, or hopeful about it; one need only entertain the thought of it or *imagine* it. You can experience terror by entertaining the thought of an intruder in the night, or joy when fantasizing about professional acclaim, or sorrow when imagining the loss of a valued friend. In these cases, the causes of your emotions are counterfactual thoughts, and the objects of those counterfactuals – what they are about or represent – are possible, but non-actual states of affairs. This suggests that Premise 2 should be revised:

- Revised Premise 2: To experience an emotion towards something, one must represent that thing in thought.

Such is the strategy of the ‘Imagination Theory’ of emotional engagement with fictions.³¹

Imagination Theory

Imagination Theory re-describes what emotions are: they are responses to representations, requiring no existential commitment to their referents. Jenefer Robinson, for example, endorses a conception of emotions (and an ontology of their contents) which is substantially the same as that informing Transportation Theory, characterizing them as ‘ongoing interactions between an individual and the environment . . . where the environment includes not only the world of the physical sciences but the world as it appears to us in our thoughts and imaginings’.³² Applying this notion to the classic case of readers’ pity for Anna Karenina, she writes:

If I feel my interests and values to be at stake in my encounter with this object of imagination, then I can respond emotionally to it (‘her’). Just as I can get all worked up imagining my parents dying in a car crash, so I can get all worked up imagining someone called Anna Karenina going through all the wrenching experiences Tolstoy describes her as having. This is just a fact about how human emotions function. Furthermore . . . emotions do not require *beliefs* about anything, but only a perspective on things, in terms of our own wants, interests, and values. What Tolstoy succeeds in doing so masterfully is in getting us to find our own wants, interests, and values to be at stake in Anna’s story, so that we respond emotionally to her. Indeed, there are scenes in the novel which can induce almost the full panoply of

emotional responses: physiological changes, facial and vocal expression, and action tendencies.³³

Robinson's conception of emotional responses to fiction harmonizes well with the experimental findings discussed earlier. Moreover, those findings suggest that she is right to insist that (given how 'human emotions function') the inner world of our imaginings is efficacious in causing a wide range of vivid emotional responses. That is why we often resonate with fictional characters as we do, and why they can be proper targets of both emotion and empathy. In these respects, Imagination Theory is compatible with naturalistic accounts of emotions, focusing on the causes of, rather than reasons justifying emotions. (Robinson's own conception of them as 'bodily perturbations' happily identifies them as 'pre-rational'.) Moreover, it disposes neatly of the Paradox of Fiction.

Unsurprisingly, however, philosophers have questioned whether and how Imagination Theory can accommodate the apparent rationality of our feelings for fiction. By endorsing an 'inner world ontology' of imagined characters, events, etc., it seems to leave emotions answerable *only* to how things appear to the subject who has them. And if we are justified in having an emotional response to *anything* we imagine, can we still mark the difference between rational and irrational emotions at all?

It is useful in this context to distinguish two ways in which our mental states can fail to be rational. A first way, as we have seen, is to transgress against the principle of non-contradiction, which prohibits the simultaneous endorsement of contradictory assertions or beliefs ('I believe that P and I believe that not P'). It is this requirement which Premises 1, 2 and 3 jointly contravene. Non-contradiction, however, is only the most minimal requirement: rationality norms typically demand more than the avoidance of inconsistencies, and apply to attitudes other than belief. Specifically, rationality requires that many of our intentional states are backed by reasons. As Thomas Scanlon puts it, they are 'judgement-sensitive attitudes':

[There] are attitudes that an ideally rational person would come to have whenever that person judged there to be sufficient reasons for them and that would . . . 'extinguish' when that person judged them not to be supported by reasons of the appropriate kind. Hunger is obviously not a judgment-sensitive attitude; but belief is, and so are fear, anger, admiration, respect, and other evaluative attitudes such as the view that fame is worth seeking . . . [J]udgment-sensitive attitudes constitute the class of things for which reasons . . . can sensibly be asked for or offered.³⁴

Emotions are typically subject to a norm of rationality in this wider sense; we expect ourselves and others to be 'reasons responsive' in respect of them. For instance, we very often expect a person to offer reasons when we ask them why they are angry, delighted, sad and the rest, and to take into account reasons why they should or should not have these attitudes.

At the same time, emotions are also 'reasons resistant' in familiar and intelligible ways. Everyday emotions sometimes are indifferent to belief, without failing to be intelligible. Even after your neighbour reveals that his snarling Pit Bull is toothless, you may still find the dog's demeanour threatening; that is just how he looks to you, and that is the response he evokes. Literary fiction, too, is replete with examples of unjustified emotions, often to their characters' great detriment. In *Tess of the D'Urbervilles*, Angel Clare may believe that Tess was 'more sinn'd against' than sinning, but his emotions of indignation and disgust prevail. Likewise, Tess herself appreciates that she is morally blameless for her fall, and yet she is overwhelmed with shame. Such cases of 'emotional perseverance' against our best reasons are not, by Scanlon's lights, rational, but they are far from incoherent. Is our emotional engagement with fiction defective in a similar way? Even if it involves no internal contradiction, does it remain irrational in this wider sense? Consider the agoraphobic's terror of the open space of his garden, or the habitual gambler's hope for his 'big win', convinced (against all probability) that this lies just around the corner. These are responses to imagined representations. They are surely genuine emotions, but they also seem to be paradigm cases of *irrational* emotions. And so they are. Can Imagination Theory distinguish between justified emotions and cases such as these? If not, then it threatens to place the emotionally immersed and empathic fiction reader on the same rational footing as phobics and fools.

The motivating worry here is that if emotions are not governed by an epistemic norm of correspondence-to-the-facts, then they are ungoverned altogether. Like many skeptical worries, it is misplaced. To see why, consider an analogy: the epistemic norms governing judgements of the expressive properties of artworks, where the expressed content is an emotion, e.g., that the music is cheerful, the poem is whimsical, the painting is gloomy, and so on. Different accounts of artistic expression tie such judgements to actual, experienced emotions in different ways. Specifically, some accounts take attributions of expressed emotions to be justified by (true) *beliefs* about the work in question, while others take them to be justified by *experiences* of the work – experiences typically including occurrent emotional responses. Accounts of the first kind include certain

so-called ‘appearance’ theories of expression. Appearance theories deny that genuine emotions are required either for the expression of emotion or for justified judgements of expressive properties. Rather, works of art express a given emotion if and only if they appear like some natural manifestation of it – for instance, by looking or sounding or moving in ways that resemble a person expressing that emotion. Just as a basset hound can have a melancholy look without either possessing or causing a melancholic state of mind, so a piece of music can sound sorrowful without this reflecting anyone’s actual sorrow – either the composer’s or listener’s. A work’s expressive properties are thus constituted by its appearance only – an appearance that floats free of any actual, affective experience. Judgements of the work’s appearance are accordingly both inferred from and justified by *beliefs* about the work’s properties, e.g., the formal and sensible properties it shares with the ‘look’ of natural manifestations of the attributed emotion. Appearance theories of expression are, for obvious reasons, congenial to Pretence Theories of emotional responses to fiction: if a reader responds to a fictional work with, say, sorrow, he is responding to a ‘mere appearance’, requiring no existential commitment to its target.

By contrast, an ‘arousal’ theory of artistic expression holds that what it is for a work to express this or that emotion is just for it to arouse, or be disposed to arouse, that emotion in an audience, whether actual or ideal. Arousal theories count attributions of expressive properties as justified just if the constitutive arousal occurs (or if the work is disposed to elicit them in suitable conditions). Arousal theories thus take expressive attributions to be more like perceptions than beliefs: they are *experience-based*, where the relevant experiences are emotions themselves. While most versions of arousal theories are vulnerable to well-known objections, they respect the intuition that expressiveness in a work of art ‘must be perceivable, not just inferable, in order to deserve that appellation’.³⁵ As Robinson comments, ‘If you smile a Duchenne smile that expresses your happiness, I am able to *see* your happiness *in* your smile; I don’t just make inferences from your behavior to your state of mind’.³⁶ Robinson herself endorses a hybrid ‘romantic’ theory of expression which holds that what it is for a work to express some emotion is for it to ‘articulate and individuate’ an emotion experienced by some *persona* (which could, but need not be the artist). On this view, the arousal of emotion in the audience is not constitutive of its expressive properties. However, such arousal – along with other experiential states such as empathic attunement and perceptual imaginings – nonetheless provides reliable evidence of those properties and justifies our judgements of them. As Robinson

puts it, 'expressive qualities . . . can be grasped through the emotions they arouse', and 'a good criterion for what should count as expressive qualities is that they evoke corresponding emotions in audiences'.³⁷

Note that both appearance and arousal accounts of expressive judgements may be governed by a norm of correctness – an epistemic norm determining whether a given attribution is or is not rationally justified. They differ in respect of the proposed justifying grounds (beliefs versus experiences) and the justificatory procedure (rational inference versus emotional response). But experience-based judgements of expressive properties, just as much as belief-based ones, are accountable to other, non-expressive and even non-evaluative features of the work.³⁸ Suppose that you judge Mozart's *Requiem* to be expressing awe and wonder because it elicits those emotions in you, while I experience it as expressing dismay and terror. Our judgements are answerable to indefinitely many other salient features of the work – for instance, its melodic, harmonic, rhythmic and structural properties. True enough, our experience-based judgements are not inferred or reasoned from beliefs about those properties of the work; they are based on and caused by our listening experiences and the affectivity with which those are imbued. They do not derive from 'good reasons', in Scanlon's sense, nor could any such reasons entail them, simply because experiences are entailed by nothing: they are caused. (Only hearing Mozart's music can bring them about.) Nonetheless, they are not arbitrary nor ungoverned by rational norms. They are answerable to the internal properties of the work, and justified (or not) by the considerations that can be adduced for and against them. Not all rationally relevant considerations license inferences to conclusions; sometimes they simply nudge us to experience things in a different way, or from a different point of view.

Likewise, our emotional responses to fictions, and especially our empathic responses to fictional characters, arise from natural, psychological trajectories. Those trajectories, as psychologists describe them, are causal, not inferential ones. They are, however, governed by a norm requiring them to be sensitive to other salient features internal to their referents. If I find Anna Karenina's tragic circumstances *funny* or (like the dreadful Countess Lydia) merely *distasteful*, or (like Vronsky, over time) rather *boring*, my emotional responses have gone badly awry. Emotions, qua experiences, remain accountable to indefinitely many natural properties which both cause and justify them; some responses do, and others do not, count as getting it right, and the wrong emotions are rationally corrigible. In these ways, the logic of emotional engagement with fiction is not at all like the phobic's irrational fears or the gambler's irrational

hopes. Consideration of the other, salient properties of the phobic's and the gambler's referents (the familiarity and security of the garden, the vanishingly small chances of a big win) fail altogether to support their emotions. They are not simply embracing an alternative ontology along the lines of Robinson's 'inner world'. Strictly speaking, their problem is that they have no alternative ontologies, and conflate the referents of their emotions with the mind-independent, outer world. The phobic and the gambler are making a specifically epistemic error: they fail to recognize that their emotions have imagined contents, and that these imaginings are not veridical. Not so the reader of fiction, who knows exactly where to look for both the explanation and justification of his experience, viz., the words on the page, and the imaginings they inspire.

Neither is the fiction reader guilty of irrational emotional perseverance, such as Angel Clare's moral indignation and Tess's shame. Angel's and Tess's attitudes fail to be sensitive to a rationally salient feature of the situation eliciting them, viz., that Tess was more a victim than an agent of her fall. Their attitudes are still largely intelligible to us, of course, in a way that the phobic's and the gambler's attitudes are not. This is because we can appreciate that other, salient features of their actual situation, while not sufficing wholly to justify their attitudes, suffice to render them rationally intelligible (human psychology being what it is). We may even empathize with Angel's and Tess's epistemic predicaments, and with the imperfection of their reasons, in part because we recognize such imperfections in ourselves. That being so, reading *Tess of the D'Urbervilles* can be an exercise in suffering twice over: once for its characters, and once again for ourselves – just as transportation theory predicts.

Transportation and the Paradox of Tragedy

[When I read Celine] I do not learn that love does not exist, that all human beings are hateful and hating (even if – and I am sure this is not the case – those propositions should be true). What I learn is to see the world as it looks to someone who is sure that hypothesis is correct.

Hilary Putnam³⁹

At the start of this chapter I presented the central claims of transportation theory and surveyed some of the evidence supporting it: evidence that readers of literary fiction engage first-personally with affective states appropriate to fictional characters, and respond empathetically to those characters,

simulating their specific moods, emotions and evaluative attitudes. The resulting profile of the interactions of fiction, emotion, and empathy supports the pre-reflective experience of many readers, but poses challenges to its rational justification. I have attempted to outline how those challenges can be addressed, and how empathic and other emotional responses to fiction remain answerable to epistemic norms.

Epistemic norms are not, however, the whole of rationality. In conclusion, I should mention too the challenge that fiction has seemed to some to pose to *prudential* rationality. Very often – as in the classic example of Anna Karenina and other narratives of misfortune – engaging emotionally with a fictional text compels us to simulate experiences that we would find highly aversive in everyday life. Why, then, do we not only submit to but enthusiastically seek out occasions for such experiences in fiction? We often actively appreciate aversive events presented in fictions, when we do not appreciate, and indeed strive to avoid, such events in life. This is one way of formulating the so-called ‘Paradox of Tragedy’ – a psychological question which, while not strictly a paradox in the logical sense of the term, is perhaps more genuinely puzzling than the putative Paradox of Fiction. If Pretence Theory were true, then it would offer a solution to both paradoxes, for if fictions only delivered ‘quasi-affects’ – make-believe emotions, moods and motivations – this might explain why we do not avoid them. Pretence Theory is mistaken, however, so that solution is not available.

I have argued that there is nothing epistemically irrational about our emotional responses to fiction. Is there nonetheless something *prudentially* irrational about our willing, affective engagement with fictional depictions of distress? It is natural and prudentially rational to seek out experiences which promote, rather than frustrate or defeat our personal ends, and few readers have as ends the sufferings of an Oedipus, or an Anna, or a Tess. Even if we elucidate prudential rationality so that one’s personal ends have a wide scope that includes the well-being or flourishing of other persons, or (even more widely) our species, it is difficult to see how these would be served by engaging with fictional tragedies.

Here again, descriptive psychology may offer part of the answer. Recall that the sort of emotional transportation promoted by fiction, and by literary fiction especially, tended to facilitate both affective Theory of Mind and empathic attunement. Psychologists often mention the instrumental value of these capacities. Oatley, for example, suggests that ‘fiction is the simulation of social skills’. Just as people improve their aviation skills in flight simulators, ‘those who read fiction might improve their social ones ...’⁴⁰ The findings cited

earlier lend some support to this claim, and they arguably identify a beneficial psychological effect of fiction reading. However, they do not speak to the Paradox of Tragedy, for that is a puzzle about motives, not consequences: the puzzle is that we find something *intrinsically* valuable about experiencing human suffering through fiction. We do not typically engage with tragic works of literature as part of an instrumental programme of skill enhancement, and when a novel (or film or drama) is commended as ‘moving’, this is free-standing praise, even if the way in which one is moved is to grief or fear. Why do we value the experience of being moved by fiction? A careful answer to that question would have as many parts and dimensions as there are types of fictions, and would require a much more elaborated psychology of affectivity reaching beyond the emotions to moods, evaluations and the wider panoply of human attachments and aversions. I cannot pursue that here. In closing, however, I offer an observation that I take to be an important part of the answer.

Marcel Proust wrote that within an hour a novelist can present ‘all kinds of happiness and misfortune, which would take years of our ordinary life to know’.⁴¹ The novelist achieves this not by introducing new beliefs, nor even by providing evidence for existing beliefs, but by provoking first-personal experiences of something that is typically opaque to us: others’ inner lives. It is a feature of our natural constitutions that, while we are designed to pursue intimacy with our fellows, each of us occupies a distinct, physical location and site of consciousness – a discrete and isolated locus of experience with its own path through space and time, its own, often silent, hopes and fears, and its own impending death. In responding to a well-composed literary work, however, the reader is carried beyond the outer world shared by all, to another site of subjective experience, complete with its own vivid and moving phenomenology of perceptions and sensations, animated by its own affective life. ‘Transportation’ may be a metaphor, but it is an apt one. What happens through engagement with literature has something in common with the *folie-à-deux* of human love, turning others’ sorrows into sweet episodes of intimacy and understanding. By delivering different sites – not of beliefs, but of experience – literary fictions deliver us to one another.

Notes

1. Plato, *Republic*, trans. G. M. A. Grube, rev. with Introduction by C. D. C. Reeve (Indianapolis/Cambridge: Hackett, 1992), 606d1–7.
2. Keith Oatley, 'Fiction: Simulation of Social Worlds', *Trends in Cognitive Sciences*, 20:8 (August 2016), pp. 618–628.
3. P. Matthijs Bal and Martijn Veltkamp, 'How does Fiction Reading Influence Empathy? An Experimental Investigation on the Role of Emotional Transportation', *PloS One*, 8:1 (2013): e55341.
4. Keith Oatley, 'On Truth and Fiction', in Michael Burke and Emily T. Troscianko (eds), *Cognitive Literary Science: Dialogues between Literature and Cognition* (New York: Oxford University Press, 2016), p. 261.
5. *Ibid.* The quality or intensity ('greatness') of emotional experience is often measured by self-report questionnaires of a commonsense sort, asking for Lykert Scale responses to questions such as 'How moved were you by N's experiences?'
6. See Bal and Veltkamp, 'How does Fiction Reading Influence Empathy?', e55341: 'Because fiction does not follow the reader into real life, the reader can allow oneself to freely experience strong emotions, without immediate transfer of these emotions to real life. Moreover, we can allow ourselves to sympathize strongly with a character of a fictional story, because we do not have obligations towards the characters of a fictional story, while sad reports in a newspaper may cause feelings of obligation towards the victims to help them.'
7. Keith Oatley, 'Fiction', pp. 618–628.
8. The 'Reading the Mind in the Eyes' test was developed and published by Simon Baron-Cohen et al. as a novel measure of Theory of Mind proficiency. It presents images of human eyes which subjects are tasked to pair with one of thirty-six target mental states such as 'playful', 'frightened', 'depressed', 'delighted'. It has been shown successfully to discriminate adults with Aspergers Syndrome or High-Functioning Autism from controls, and has also been used to evaluate differing levels of mind-reading proficiency in neurotypical (non-pathological) subjects. See Simon Baron-Cohen et al., *Journal of Child Psychology and Psychiatry*, 38 (1997): (813–822).
9. Raymond A. Mar, Keith Oatley, Maja Djikic, and Justin Mullin, 'Emotion and Narrative Fiction: Interactive Influences Before, During, and After Reading', *Cognition & Emotion*, 25:5 (2011): (818–833), pp. 828–829.
10. Alison E. Denham, 'Empathy and Moral Motivation', in Heidi L. Maibom (ed.), *The Routledge Handbook of Philosophy of Empathy* (Abingdon: Routledge, 2017), p. 234.
11. Maja Djikic, Keith Oatley, Sara Zoeterman, and Jordan B. Peterson, 'On Being Moved by Art: How Reading Fiction Transforms the Self', *Creativity Research Journal*, 21:1 (2009): (24–29), p. 26.

12. Jerome Bruner, *Actual Minds, Possible Worlds* (Cambridge, MA: Harvard University Press, 2016), p. 25.
13. Bruner, *Actual Minds, Possible Worlds*, *passim*.
14. David Comer Kidd and Emanuele Castano, 'Reading Literary Fiction Improves Theory of Mind', *Science* 18 (October 2013), pp. 377–380.
15. Diana I. Tamir, et al., 'Reading Fiction and Reading Minds: the Role of Simulation in the Default Network', *Social Cognitive and Affective Neuroscience* 11:2 (2015): 215–224.
16. Frank Hakemulder, *The Moral Laboratory: Experiments Examining the Effects of Reading Literature on Social Perception and Moral Self-Concept* (Utrecht: John Benjamins Publishing, 2000).
17. Jorge Luis Borges, *Dreamtigers*, trans. Antonio Frasconi (Austin, TX: University of Texas Press, 1964), p. 48.
18. In Book II of the *Republic*, Plato notes that the content of Homer's narratives stands to corrupt the spectator's soul, and that 'even if they were true' they ought not to be told.
19. Plato, *Republic*, 606d1–607.
20. It would be wrong, however, to suppose that one can draw a hard line between a literary work's descriptive and expressive properties; in literature, as in other representational artforms, expression often occurs by way of representation. For instance, Giorgione's painting, *La Vecchia* (*Old Woman*, c. 1508–10) expresses both the fatigue and wistfulness marking the final stage of a long human life by depicting a face that itself manifests those emotions. Likewise, in literature, it is how agents, scenes and events are described that they individually, and the work overall, express what they do. In general, one thing X expresses another A, if it conventionally or naturally denotes A, and is a realization or manifestation of A. Thus, for instance, your tears express your sadness: they naturally denote or indicate your sadness, and do so by manifesting it.
21. One might argue that both answers cannot be correct; after all, if fictions deliver pleasure, then it surely *is* rational to engage with them.
22. Neo-Humean theorists, of course, have long held that our evaluative beliefs derive from the sentiments, while denying that such beliefs ever constituted knowledge. Hume himself may be interpreted, for instance, as holding that the properties targeted by our evaluative beliefs are mere projections of sentiment, and thus have no mind-independent existence; in that case, their truth is always relative to our particular, sentimental constitutions. Some contemporary expressivist theorists have more radically denied that our evaluative commitments are proper beliefs at all, arguing that evaluative statements are disguised avowals or expressions of affect.
23. Colin Radford, 'How Can We be Moved by the Fate of Anna Karenina?', *Proceedings of the Aristotelian Society*, Supplemental Vol. 49 (1975): pp. 67–93; Colin Radford, 'Tears and Fiction', *Philosophy* 52 (1977), pp. 208–13.
24. See Radford, 'Tears and Fiction'.

25. For examples of this literary turn in metaethics, see Martha Nussbaum, *Love's Knowledge* (Oxford: Oxford University Press, 1992), A. E. Denham, *Metaphor and Moral Experience* (Oxford: Oxford University Press, 2000), and Richard Moran, *The Philosophical Imagination* (Oxford: Oxford University Press, 2017).
26. That said, some have been undeterred, proposing either that we suspend disbelief when absorbed in a fictional 'world', or that our emotions are directed at various real-world correlates.
27. E. H. Gombrich, *Meditations on a Hobby Horse and Other Essays on the Theory of Art* (London: Phaidon Press, 1963).
28. Kendall Walton, 'Fearing Fictions', *Journal of Philosophy* (75.1), 5–27, pp. 6–7.
29. Noel Carroll, *The Philosophy of Horror, or Paradoxes of the Heart* (New York: Routledge, 1990), p. 74.
30. *Ibid.*
31. Imagination Theory, sometimes referred to as 'Thought Theory', is credited in its current form to Philip Lamarque. See Philip Lamarque, 'How Can We Fear and Pity Fictions?', *British Journal of Aesthetics*, 21:4 (1981): pp. 291–304.
32. Jenefer Robinson, *Deeper than Reason* (Oxford: Oxford University Press, 2005), p. 184.
33. *Ibid.* p. 185.
34. Thomas M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Belknap, 2000), pp. 20–22.
35. Jerrold Levinson, *The Pleasures of Aesthetics* (Cornell, NY: Cornell University Press, 1996), p. 105.
36. Robinson, *Deeper than Reason*, p. 264.
37. *Ibid.* pp. 291–292.
38. While I will not argue for it here, my own view is that attributions of aesthetic properties generally, including expressive ones, can never be inferred from nor conclusively justified by beliefs about its non-aesthetic ones.
39. Hilary Putnam, 'Literature, Science and Reflection' in *Meaning and the Moral Sciences* (London: Routledge, 1979), p. 488.
40. Keith Oatley, 'Fiction', p. 619.
41. Marcel Proust, *Swann's Way (Remembrance of Things Past, Vol. 1)*, trans. C. K. Scott Moncrieff and Terence Kilmartin (1913, repr. London: Vintage, 1989), p. 84.

