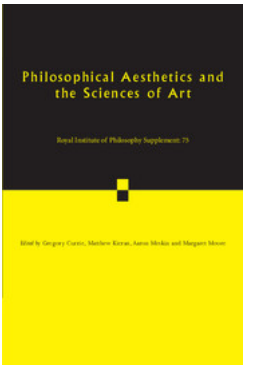

The Epistemology of Fiction and the Question of Invariant Norms

Jonathan Gilmore

Royal Institute of Philosophy Supplement / Volume 75 / October 2014, pp 105 - 126
DOI: 10.1017/S1358246114000204, Published online: 03 October 2014

The Epistemology of Fiction and the Question of Invariant Norms

JONATHAN GILMORE

I.

A primary dimension of our engagement with fictional works of art – paradigmatically literary, dramatic, and cinematic narratives – is figuring out what is true in such representations, what the facts are in the fictional world. These facts (or states of affairs) include not only those that ground any genuine understanding of a story – say, that it was his own father whom Oedipus killed – but also those that may be missed in even a largely competent reading, say, that Emma Bovary's desires and dissatisfactions are fed by reading romance novels.

How we uncover fictional truth parallels how we decide what is true in the real world. When forming beliefs, as well as revising, transitioning among, and relinquishing beliefs, we rely on standard sources of evidence such as testimony, perception, memory, the results of inductive and deductive inferences, and our affective responses. When all goes well, these sources provide the right sorts of reasons for our beliefs: reasons that justify or serve as warrants for what we believe. Analogous operations supply and justify what we imagine to be true in a given fiction.

Sometimes we rely on the testimony of a narrator who is largely transparent to the text. In these cases we treat what is said as true by stipulation: reading, 'it was a dark and stormy night…,' we don't typically need to look for other confirming evidence to be justified in believing that the description captures how things are in the fiction. In other cases where narrators or those in the text whom we rely on for information seem to be fully-realized individuals, we may discount what they convey to us according to the degree of reliability we attribute to them, as we do with Henry James' Maisie, who sees things with only partial comprehension through a child's eyes. Such discounting is of course what we do as good epistemic agents in response to testimony in real life. We also often infer what is true in a story, when it is not explicitly stated, from what is directly asserted to be true: I conclude that Charles Bovary (not a fully qualified physician but only an *officier de santé*) must have botched the surgery

doi:10.1017/S1358246114000204

he performed on the young groom's clubfoot because the boy's leg develops gangrene. Other things I imagine to be true are not described by a text, nor inferred from those descriptions, but are imported into a fiction from my beliefs about the real world – insofar, that is, as I see the fictional world as similar in relevant respects to our own. Hence, unless otherwise specified, we tend to assume in reading a realistic narrative that the human beings it represents are mortal and that the laws of physics hold.

Furthermore, just as I may be mistaken in my beliefs about the real world, so my imagining what is true in a fiction can be faulty – say, if I thought that the brawny peasant on a donkey that Don Quixote encounters really is the beautiful princess he hallucinates, or that Goneril's and Regan's professions of filial devotion are sincere, or that Ganymede in *As You Like It* really is a man within the story, and not Rosalind in disguise. In either case, the defect may be in the belief or imagining itself (its failure of correspondence) or in the epistemic means by which the belief or imagining is arrived at, such as when they arise out of practices that are not truth-apt, such as 'wishful thinking.'

Philosophers and psychologists have worked out a substantial picture of the kinds of normative constraints that are constitutive of epistemic rationality when applied to beliefs – what normative constraints govern a person's formation, maintenance, transitions among, and relinquishing of her beliefs. My question is whether such norms governing our beliefs about what is true in the real world apply *invariantly* to our imaginings of what is true in fictions.[1] Taking P to be a proposition expressing some fact, is it rational to imagine P is true in a given fictional world if and only if it would rational to believe P is true for the same kinds of reasons in real life? In short, is make-believing rational in the same sense as believing?

[1] Related questions concern the degree of commonalities in the way beliefs and imaginings are attributed to agents; whether imaginative states bear the same relations (inferential, causal, supervening, etc.) among themselves as belief states do; and what systematic relations exist between imaginative states and belief states. For discussion of these and other comparisons between beliefs and imaginings, see Tamar Gendler, 'On the Relation Between Pretense and Belief' in Matthew Kieran and Dominic McIver Lopes, eds. *Imagination, Philosophy, and the Arts* (Routledge, 2003), p.125–141; and Shaun Nichols, 'Introduction' in Shaun Nichols, ed. *The Architecture of the Imagination* (Oxford: Oxford University Press, 2006), 1–18.

Let me describe two opposed answers. On the one side are those who see imagining in response to a fiction as rational according to the same norms that govern whether a belief is rational. The norms of believing and imagining are invariant across the real and fictional divide. We can call this a commitment to invariance or *continuity*. On the other side, proponents of *discontinuity* see the standards of rationality for forming beliefs about the real world as in tension with, if not collectively inconsistent with, the standards that govern imaginings of what is true in a fictional world.[2]

On the side of continuity is the intuition that our epistemic behavior in relation to the contents of fictions is very much like that in relation to states of affairs in the real world – indeed, it isn't clear how authors could expect us to understand their fictions correctly, to import what needs to be imported, to infer what needs to be inferred, for the fiction to make sense, were they not able to rely on our rational processes for discovery of facts about the world being taken 'offline' and directed to the stories they create.

The discontinuity view, however, relies on the intuition that imagining exhibits a freedom that seems to distinguish it from many other representational states of the mind. It seems, for instance, that I can successfully imagine *at will* that almost any facts hold, but this cannot be said, both conceptually and practically, of remembering, desiring, perceiving, or believing, which seem more greatly constrained by the

[2] Whether or not a given theory of fiction-directed imagining commits to or denies invariance tends to be only implicit. Some of the more salient expressions of continuity can be found in: Ruth M. J. Byrne, *The Rational Imagination: How People Create Alternatives to Reality* (Cambridge, MA: MIT Press, 2005); Shaun Nichols and Stephen Stich, 'A Cognitive Theory of Pretense', *Cognition* **74** (2000), 115–147; and Gregory Currie and Ian Ravenscroft, *Recreative Minds: Imagination in Philosophy and Psychology* (Oxford: Oxford University Press, 2003). Discontinuity is a tenet of Romantic theories of the imagination, as in Coleridge's *Biographia Literaria* (1817), and existentialist and phenomenological treatments of fiction, such as, respectively, Jean-Paul Sartre's *What is Literature?* and Maurice Merleau-Ponty's 'Indirect Language and the Voices of Silence.' Among contemporary theorists, Kendall Walton endorses what appears to be a qualified thesis of discontinuity in stressing the absence of any "simple set of principles" governing the generation of fictional truths (*Mimesis as Make-believe: on the Foundations of the Representational Arts* [Cambridge, MA: Harvard University Press, 1990], 185). My aim in this paper, however, is not to offer critical exegesis of the views of theorists of fictions but to expose a significant conflict between two positions in which they cannot avoid taking a side.

circumstances I find myself in, and the other memories, desires, perceptions and beliefs that I already have and will not relinquish.

No doubt, much recent work in cognitive psychology, neuroscience, and philosophy of the imagination supports the idea that there are several forms of *descriptive* continuity to be observed across the stances of believing and imagining.[3] The psychological mechanisms that process believing that *P* appear to operate in ways parallel to, and employ much of the same cognitive architecture, as those that process pretending or imagining that *P*. However, my question is about the *norms* associated with those propositional attitudes: whether the criteria governing their epistemic justification hold invariantly across our beliefs about the real world and our imaginings of what is true in fictional works of art.[4]

In what follows I introduce and assess some considerations in favor of these two theses of continuity and discontinuity, both of which have a prima facie plausibility. Ultimately, I defend a version of the discontinuity thesis: for readers and audiences of fictions, there are epistemic reasons to attribute facts to a fictional world that would not count as epistemic reasons to identify analogous facts in the real world. More generally, the norms in light of which our imaginings can be epistemically warranted are not, as a whole, consistent with those in light of which our beliefs are epistemically warranted.

A few caveats are in order:

(1) There is no doubt that the concept of rationality when applied to either beliefs or imaginings requires careful qualification. There is no consensus over what theoretical rationality consists of, hence no easy way to ask whether the norms of belief are altogether invariant over believing and imagining.[5]

[3] Representative studies are: Byrne, *The Rational Imagination: How People Create Alternatives to Reality* (Cambridge, MA: MIT Press, 2005); Paul L. Harris, *The Work of the Imagination* (Oxford: Blackwell Publishing, 2000); Shaun Nichols and Stephen Stich, 'A Cognitive Theory of Pretense', *Cognition* **74** (2000), 115–147; and Timothy Schroeder and Carl Matheson, 'Imagination and Emotion' in *The Architecture of the Imagination*, Shaun Nichols, ed. (Oxford: Oxford University Press, 2006), 19–40.

[4] For the debate between theories of continuity and discontinuity over the norms governing our emotional responses to fictions and the actual world, see Jonathan Gilmore, 'Aptness of Emotions for Fictions and Imaginings', *Pacific Philosophical Quarterly* **92**.4 (2011), 468–489.

[5] In this discussion I treat theoretical and epistemic rationality as largely identical capacities. In other philosophical contexts, however, the two may

In what follows I appeal to the putative invariance of only some of the most familiar and relatively uncontroversial norms identified with theoretical rationality, not to exotic norms that only a perfectly rational person, say an android decision theorist possessed of unlimited working memory, might be guided by. I will also not enter into the debate as to whether such normativity applies in the first instance to beliefs and only derivatively to believers, or to the converse (as in some theories of virtue epistemology). I assume that anything I say about the rational grounds for a belief can be translated into an attribute of someone's epistemic disposition to rely only on such rational grounds.

(2) Nothing I say here is meant to address the metaphysics of fictional worlds, or more mundanely, what makes something true, part of a story, make-believe, and such, in a fictional world. My only relevant commitment is to the idea that there is criterion of representational correctness in what we imagine when we submit our imaginative activity to the objective constraints of a work of fiction, allowing that most fictions underdetermine what we may imagine of them consistent with correct comprehension.

(3) Finally, I am not addressing norms that govern the correctness or aptness of beliefs or imaginings, *all things considered*, but only those that govern their intentional, or more specifically, *representational* correctness and the putatively justifying means by which such correctness is achieved. There may be practical, prudential, aesthetic, moral and other norms in virtue of which one has a reason to believe or imagine something, or, more specifically, to put oneself in a position in which one will *come to* believe or imagine it. Sometimes an instrumental reason (e.g., it would be too distressing) might trump an epistemic reason to believe something or to imagine it to be true. But even if practical, moral, aesthetic, and other kinds of reasons can trump epistemic reasons, they do not silence epistemic reasons, in the sense of making them wholly inapplicable. My concern is only with the

---

be distinguished, particularly in how the former but not the latter requires that one be sensitive to certain kinds of instrumental reasons pertaining to the achievement of one's cognitive goals. See Thomas Kelly, 'Epistemic Rationality as Instrumental Rationality: a Critique', *Philosophy and Phenomenological Research* **66**.3 (2003), 612–640.

> invariance or otherwise of epistemic norms governing how our beliefs and imaginings present their contents – as, respectively, true of the world, or true of a fiction. Do the same standards of epistemic rationality apply to beliefs and fiction-guided imaginings? Not: Do the same standards of rationality, *in toto*, apply to the two domains?[6]

II.

A continuity proponent might suggest that our abiding by the same rational standards for belief formation and transition is what creators of works of fiction rely on to let us know what is true in a fiction.[7] If the astronaut crew crash-lands on a primitive planet ruled by apes but then comes upon the charred fragments of the Statue of Liberty, audiences can be expected to infer that the strange planet is actually Earth (!) in a post-apocalyptic future. If we are directed to imagine that a fictional world is much like our own, we are entitled to assume that a character in the fiction who is in London in the morning and New York later the same day has traveled there by air.[8] If our epistemic norms were not continuous, such identifications of the facts in a fiction would not be so predictable.

[6] There is a narrow sense in which pragmatic factors may plausibly be counted as providing epistemic reasons relevant to acquiring a belief, as when the degree of importance associated with being correct in some claim affects what one counts as a sufficient level of evidence to believe it. Whether or not such pragmatic reasons count in epistemic justification need not be addressed here as my question is only whether the kinds of reasons, whatever they may be, that justify beliefs apply invariantly to the justification of imaginings.

[7] Defenders of the descriptive continuity of believing and imagining or pretending stress what Nichols calls 'inferential orderliness': that individuals working out what is true in a given pretense often make inferences that mirror those that they would employ if the pretense were in fact real. See Shaun Nichols, 'Introduction', in Shaun Nichols, ed. *The Architecture of the Imagination* (Oxford: Oxford University Press, 2006), 1–18.

[8] Currie and Ravenscroft write: 'It is this capacity of imaginings to mirror the inferential patterns of belief that makes fictional storytelling possible. …If imaginings were not inferentially commensurate with beliefs, we could not draw on our beliefs to fill out what the story tells us.' Gregory Currie and Ian Ravenscroft, *Recreative Minds: Imagination in Philosophy and Psychology* (Oxford: Oxford University Press, 2003), 13–14. Such preservation of inference in imagining is also demonstrated in studies of child

Of course, one norm that a commitment to invariance should not require the observance of is that a person imagine something to be true only when she has good grounds to believe it to be true. But once we index the *grounds for imagining* some fictional or make-believe fact to the *fictional* world that the imagining is about, we can say that the same constraints that govern the epistemic behavior of any agent with respect to his beliefs govern his epistemic behavior with respect to his fiction-directed imagining. His beliefs present certain facts as holding in the real world and his imaginings present certain facts as holding in a fictional one.

It certainly seems part of the phenomenology of our engagement with fictions that we perform many of the same epistemic operations in imagining what is true in a fiction as we do in coming to believe what is true outside of it. We infer via deduction and induction from what is explicitly described as being the case to other facts of the fiction that are not so described. Through our affective and emotional reactions we imaginatively assign values to things represented in fictions just as we impute values to things in real life. We try to monitor the consistency among our imaginings in response to a fiction just as we monitor such consistency among our beliefs, sometimes giving up what we initially held to be true as a story unfolds. We think it is no more theoretically reasonable to base one's imaginings in response to a fiction on how we desire events to transpire than we do in connection with our beliefs. That I want Anna Karenina to survive, is, I recognize as a reader, no epistemic reason to justify imagining that in Tolstoy's narrative she somehow continues to live. There are, of course, stories that seek to satisfy such desires, as in 'fan fictions' that continue and sometimes revise a narrative without the sanction of the original creator. And there are interesting cases such as when the pseudonymous Alonso Fernández de Avellaneda determined that Don Quixote was in fact more pious than Cervantes posited, and wrote a narrative featuring the character after Cervantes composed the first part of his novel, but before he had finished the second. However these imaginings are not epistemically justified by reasons internal to the original fictions, although they may be justified on aesthetic, moral, or other terms. Indeed, writers sometimes try to redeem characters from others' novels, say, because they see those characters or actual people like them as

---

psychology; see, e.g., Alan Leslie, 'Pretending and Believing: Issues in the Theory of ToMM', *Cognition* **50** (1987), 211–38.

deserving of different attributes or experiences than those with which they were originally endowed.[9]

It should also be noted that audiences for fictions regularly discuss fictional characters and events *as if* they were real, debating the fine points of what a protagonist's motivations are on the basis of her behavior. This is evidence that we cite when we are called upon to justify what we imagine to be true in those fictions – as when we debate what really happened at the end of the film *Inception* (was it still a dream?). That is, we appeal to reasons in an inferentially norm-governed way that is continuous with how we appropriately justify what we believe to be true in real life.

Finally, in discovering truths about the actual world we rely on various forms of counterfactual imagining akin to our imagining what is true in a fiction. These include thought experiments, predicting the future, simulating another person's point of view, apportioning legal responsibility for some event, and appealing to scientific models featuring, e.g., frictionless planes. The epistemic value of these imaginings is sometimes controversial, whether as a source of useful moral intuitions or as a guide to metaphysical possibility. However, a proponent of continuity might contend that the very possibility of counterfactual imagining serving as a source of knowledge about the real world depends on our abiding by the same rational constraints in imagining as we do in forming beliefs. Of course, some might embrace that point, not as an intuitive support for continuity, but as a *reductio* showing that some of the products of such imagining, e.g., intuitions drawn from fictions involving trolley problems or dopplegängers exiting from teletransporters, are not a good source of knowledge about morals and metaphysics in this world where the circumstances the fictions prescribe us to imagine don't typically arise.[10]

Still, one might object to the continuity view that there is an essential dimension of the process by which we discover what to imagine as true in a fiction that has no obvious analog in the process by which we form beliefs about the real world. That is where we take an *external* approach to the fiction as an ordinary artifact in our world and

[9] As in Jean Rhys' *Wild Sargasso Sea* (1966), an alternative imagining of the life and mind of the 'madwoman in the attic' of Charlotte Bronte's *Jane Eyre* before her arranged marriage to Rochester and relocation from the Caribbean to England.

[10] That we form intuitions from thought experiments according to the same principles by which we discover truth in fiction is defended in Jonathan Ichikawa and Benjamin Jarvis, 'Thought-Experiment Intuitions and Truth in Fiction', *Philosophical Studies* **142** (2009): 221–246.

appeal to its style, tradition, function, genre, author, technique, and so on, in forming beliefs about its content – what is true from the *internal* perspective.[11] In adopting such an external stance on a work of fiction we refer to properties it has as a vehicle of representation but not (directly) to its represented content. By contrast, in adopting an internal stance, we refer to that content as if it were real or were a story being recounted by a real narrator. Othello's speech is rough from the internal stance ('rude am I in my speech') but eloquent from the external, in the poetic language Shakespeare uses. Mark Antony's funeral oration is eloquent from both the internal perspective – although he claims not to be an orator – *and* the external, in the phrasing Shakespeare employs. External features of a narrative do not lie within the scope of the operator 'it is fictional that' or 'it is part of the content that,' however, they can cause us to form propositional attitudes – beliefs, imaginings, emotions, desires – toward what is. For example, P.G. Wodehouse's novels often feature an earl or lord raising an older child in the absence of her mother, who has died before the period of the story begins. We assume that the characters don't dwell much over that loss (even though it would be natural to import that assumption from real life) because we know that would be foreign to Wodehouse's comic aims. Similarly, we are usually correct in inferring that the party who appears guilty of the murder in the first few pages of a traditional mystery story is not genuinely the villain, for paradigm mystery novels don't give up the game that early. Finally, a viewer of the film *Clueless* about a group of American high-school kids can make reliable assumptions about the significance of various turns in the plot if he's familiar with its acknowledged model, Austen's *Emma*.

However, those appeals to what is true *of*, but not *in*, a work of fiction do not count against continuity. In principle, if we had access to such an external source of understanding our world – say, through reliable beliefs about Providence, Fate, or Karma – we would use it to infer what is true in our world. No such more-than-human-knowledge is available, but in principle it would serve as a

[11] For discussions of internal and external stances on a fiction, see Peter Lamarque, *Fictional Points of View* (Ithaca, NY: Cornell University Press, 1996), chapters 2 and 8; and, Gregory Currie, 'Two Ways of Looking at a Narrative.' In *Narratives and Narrators: a Philosophy of Stories* (Oxford and New York: Oxford University Press, 2010), 49–64. An analogous distinction is noted by Kendall Walton, 'How Remote are Fictional Worlds from the Real World?' *Journal of Aesthetics and Art Criticism* **37** (1978–9), 21.

source of beliefs about what is true in our world as much as external features of a story serve as a source of what to imagine as true about its world. This scenario, of course, is sometimes explored thematically within works of art, as in *The Truman Show* – about a character whose life is orchestrated for the sake of a television series – and *The Comforters*, a short novel by Muriel Spark in which Caroline discovers that she is only a character in a fiction (she continually hears typing on a keyboard) and resolves to frustrate her author's plans.

III.

Let me now turn to the discontinuity view, which also seems prima facie plausible. This is the view, recall, that the rational norms that govern the formation of imaginings with respect to what is true in a fiction can be inconsistent with the rational norms that govern the formation of our beliefs. Sometimes, as in Shelley's Romantic primer, *Defense of Poetry* (1821), this is construed as the denial that reason has any role in the activities of the imagination.[12] In other formulations, imagination is reason-governed but perhaps – this is the question – not subject to the same norms of reasoning as believing. I want to first address, and suggest we reject, the most familiar point appealed to in favor of the thesis of discontinuity – one that pertains to the unconstrained contents of fictions. In its place, I introduce a defense of discontinuity that I think better survives philosophical scrutiny.

The most familiar point made in favor of discontinuity is that it is a highly salient feature of our engagement with fictions that they call for us to imagine things as true that are not, and sometimes could not be, true in our world. Fictional worlds can present fantasies as reality, featuring radical departures from standard laws of physics and states of affairs that are internally inconsistent. And while our real world, like all possible worlds, exhibits logical closure – any genuine proposition is either true of our world or false of our world – fictional worlds are typically incomplete: some propositions, such as that Emma Bovary has blue eyes, are neither true nor false in the world of Flaubert's novel, there being nothing in the narrative that gives us a reason to accept or deny that claim. If fictions call on us to imagine such fantastical states of affairs, so different,

[12] In *English Essays: From Sir Philip Sidney to Macaulay. With Introductions and Notes,* edited by Charles W. Eliot (New York: P.F. Collier and Son, 1909).

physically and metaphysically, from the actual world, they must rely on our acceptance of epistemic norms that govern imagining that are distinct from those that govern believing.

But that observation based on the contents of fictions does not succeed as a challenge to continuity. For a proponent of continuity can plausibly propose that what is embedded in those fictions is a kind of metaphysical or physical principle in light of which it would be rational to infer or make-believe the truth of those other fantastical parts of the fiction. If we accept that it is true in the state of affairs of Kafka's *Metamorphosis* that human beings *can* wake up as insects, or as other creatures, then we don't in any straightforward way depart from ordinary rational judgment if we imagine on the basis of the narrator's description that Gregor Samsa *has* indeed woken up as a bug. This is just as when Alice concludes from matters being so queer in general in Wonderland – she's been shrunk and is swimming in a pool of her own tears – that it isn't that odd that the mouse she encounters is able to speak French. The question is whether we would be rational in coming to imagine such fantastical states of affairs without there being such a principle of generation internal to the fiction that serves to license such an imagining.[13]

I suggest that if we were to encounter a fictional world in which such fantastic things occur but where there is no implicit (fictionally true) physical or metaphysical principle that licenses such departures from ordinary reality, we would be just as warranted in assuming that we are reading a story recounted by a *deluded narrator* – one who only imagines all that she or he describes to be true – as we would be in assuming that, e.g., French-speaking mice really do exist in the fictional world. But in most cases there is very little pressure or apparent motivation to assume the presence of a narrator so out of touch with that fictional world. Instead, we assume that the facts of the fictional world really are as they are described because we can readily assume that it is a fact in the fiction that such bizarre, non-naturalistic events can occur – the nature of that world permits it. This experience, of course, should be contrasted with cases in which in a story presents reasons that motivate us to wonder about the narrator's reliability, as in Nabokov's *Pale Fire*, where we try to piece together what is true about the events leading to the death of the poet John Shade

[13] On principles of generation, see Kendall Walton, *Mimesis as Make-believe: on the Foundations of the Representational Arts* (Cambridge, MA: Harvard University Press, 1990), 138–40.

through an obviously delusional commentary on the source and meanings of his accompanying poem.

I suggest that a better argument for discontinuity can be found not in the sundry contents of fictions but in some of the myriad ways they succeed in eliciting our imagining of what is true. My proposal is that certain kinds of experiences generated by a fiction do serve as grounds for the imagination of certain associated facts holding in the fiction when those experiences are reliable indicators of those facts. More formally, when an experience E in reading a fiction is a reliable indicator that *P* is the case in the fiction, E is a (pro tanto) reason for imagining that *P*. At a general level, this structure of justification holds as well in relation to beliefs: an experience E can serve as a reason for a belief that *P* if E is a reliable indicator that *P*.[14] However, at a lower level of description, a fiction can provide an experience that justifies imagining something being true in the fiction while analogous experiences in the actual world may not justify an analogous belief.

It is true that in *Oliver Twist* Fagin is filthy and physically grotesque, as we learn from the attention paid by the narrator to his greasy clothes and matted hair. Yet, we imagine him as morally corrupt as well via the text's exploitation of our well-studied irrational tendency to conflate such feelings of mere physical disgust with justified moral opprobrium. No doubt, other facts internal to the fiction also explain and serve as reasons for this imagining just as they would in an analogous case of belief – such as that he exploits children. But those facts do not exhaust the pro tanto reasons warranting that moral judgment.[15] For in engaging with such a work, we implicitly accept a norm under which such physical disgust is a reliable indicator of such moral facts. No such reliable relation, hence no norm sanctioning an epistemic reliance on it, holds in the actual world. A feeling of disgust prompted by someone's filth in the real world would not offer a reason for judging him immoral. This suggests that we can have some imaginings on account of – warranted by – being caused to have other imaginings where an analogous justificatory relation

[14] I appeal here to a "reliabilist" notion of epistemic justification that does not preclude other grounds of justification. For a defense of an epistemic reliabilism as an exclusive account of justification, see Alvin Goldman, 'What Is Justified Belief?' In Alvin Goldman, *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, MA: MIT Press, 1992).

[15] See Simone Schnall, Jonathan Haidt, Gerald L. Clore, and Alexander H. Jordan, 'Disgust as Embodied Moral Judgment', *Personality and Social Psychology Bulletin* (2008), 1096–1109.

between one set of facts and another would not hold outside of our engagement with the fiction.

There are other kinds of cases in which works of art exploit automatic or subdoxastic tendencies to cause us to imagine as true what we would not have reason in analogous contexts to believe in real life. From these arise two claims. First, such phenomena serve as evidence of a descriptive discontinuity in what can serve as the bases of imaginings and epistemically-rational beliefs. Second, more controversially, these cases reflect normative discontinuity as well. They show how the causes of what we imagine to be true in a fiction can be epistemic reasons for those imaginings even if they would not be such for analogous beliefs about the actual world.

For example, in some genres of art we are induced through physical descriptions of characters – their beauty or ugliness, stereotypical racial or ethnic features of appearance, deportment, size, and so on – to conclude (correctly in relation to the story) that they have certain virtues or vices of character and certain kinds of capacities. Ugliness is often employed to provoke a judgment of nefariousness even though, of course, that would not be a proper inference between such a perception and belief. Many studies of human beauty – through what is sometimes termed the 'Halo Effect' – show that it can elicit not only an attribution of moral goodness but also intelligence: a psychological explanation of the historical idea of psyche and body mirroring each-other in 'beauty of soul.'[16] We are often solicited to construe the literal qualities of the media of some types of visual works of art as literal or figurative properties of whatever content the works depict, evoking a judgment about a represented person or state of affairs that is not grounded in the properties of that person or situation considered independently of the medium of representation. A film may cause us to think of the lives it depicts as happy through presenting them in warm tones and soft focus or a state of affairs as menacing through the use of cold blues and greys.[17] The names of characters, such as Roger Chillingworth,

[16] One representative study is Karen Dion, Ellen Berscheid, and Elaine Walster, 'What is Beautiful is Good', *Journal of Personality and Social Psychology* **3** (1972): 285–290; See also, Richard Nisbett and Timothy D. Wilson, 'The halo effect: Evidence for unconscious alteration of judgments', *Journal of Personality and Social Psychology* **35** (1987), 250–256. On the idea of 'beauty of soul' see Robert E. Norton, *The Beautiful Soul: Aesthetic Morality in the Eighteenth Century* (Ithaca, NY: Cornell University Press, 1995).

[17] Compare the metaphorical transfer exhibited in recent experiments that address the processing of tactile information: in one, volunteers asked

the cerebral husband of Hester Prynne in *The Scarlet Letter* and Gradgrind in *Hard Times,* induce us to attribute qualities to those characters that the mere possession of a name, in real life, would not indicate.

Furthermore, we often exhibit highly irrational forms of in-group/out-group bias in favoring even arbitrarily individuated communities in which we are primed to recognize our membership, but that bias is easily exploited in having us favor and judge as merited or objectively valuable the ends of the characters in a fiction with whom we are made intimate – say through having us simulate their perspective – even if independent of the fiction we wouldn't believe that those are good ends to have.[18] We value, for example, the elegant thief's finely calibrated heist even if means a loss to others with whom we don't identify. Indeed, the devices employed to prime our identification with a character can lead us to appraise the facts in the story as that fictional individual does even if a description of such facts outside of a fiction would be unlikely to garner that evaluation.[19] We worry with Tony Soprano as he frets over threats to his mob dominion and we feel a thrill implicitly approving of the ends of the hired killer in the film version of *Day of the Jackal* as he ingeniously pursues his mission to assassinate the fictional Charles de Gaulle. It is controversial to claim that all affective responses entail concomitant judgments about their objects; but in these cases our emotions do seem to evince certain context-conditioned judgments about the evaluative dimensions of their objects: e.g., "loss of control over his criminal organization would be a bad thing," or "succeeding in the assassination is the right goal to pursue."

Our tendency to see actual events as having a narrative-like structure that goes beyond mere causal and explanatory connectedness can be relied on by authors to supply the kind of closure and unity

---

to assess the quality of candidates for an alleged job tended to rate those applicants whose resumes were attached to heavier clipboards as being, themselves, more serious (i.e., 'weighty'). Joshua M. Ackerman, Christopher C. Nocera, and John A. Bargh, 'Incidental Haptic Sensations Influence Social Judgments and Decisions', *Science* **328**.5986 (2010): 1712–1715.

[18] See the suggestion that mere (arbitrary) categorization of individuals serving as research subjects generated in-group bias in S.Otten, and G. B. Moskowitz, 'Evidence for Implicit Evaluative In-group Bias: Affect-biased Spontaneous Trait Inference in a Minimal Group Paradigm', *Journal of Experimental Social Psychology* **36** (2000), 77–89.

[19] See Morton Ann Gernsbacher, et al, 'Do readers mentally represent characters' emotional states?' *Cognition & Emotion* **6**.2, (1992): 89–111.

among fictional events that traditional plots require. It could only be figuratively true of a person's life, or of a romance, that it had an organically structured and internally related beginning, climax, and denouement, but it can be literally true in a fictional world that such is the case. As both prosecutors and con artists know, embedding facts and explanations within an aesthetically satisfying narrative is more convincing than merely stating the facts and explanations outright. A successful narrative can gloss over major explanatory and causal gaps in what we are to imagine as true, without thereby having any less claim on us to evoke that imagining. The narrator of Proust's novel tells us, from a first-person perspective, of his life and emergence as a writer. But certain sequences, particularly those in the sections recounting the relationship between Swann and Odette, could not have been witnessed by the young Marcel, yet are recounted – and we as readers go along imagining those facts – as if he were there. The seamlessness of the narrative gives us reasons to imagine certain states of affairs as obtaining in the fiction even though other facts in the fiction would make those states of affairs impossible.

In his remarks about the nature of moral demands, Nietzsche portrayed our psychological need to attribute a meaning to suffering as resulting in unjustified beliefs about its redemptive significance: that it is a test of character, a divine punishment, a curse.[20] But works of fiction regularly rely on that tendency to endow objects and events with a significance that is then treated as objectively and independently possessed by them. In, for example, *It's a Wonderful Life,* the character played by Jimmy Stewart undergoes various travails that, satisfyingly, come to appear to have existed *for the sake of* his eventual enlightenment.

We also readily accede to a biased understanding – the fundamental attribution error or correspondence bias – of people's motivations, in seeing their actions as explained by stable character traits and deep psychological dispositions or motivations, rather than much more powerfully explanatory contextual or situational factors.[21] This

[20] One of the themes of his 'On the Genealogy of Morals', in *'On the Genealogy of Morality' and Other Writings,* edited by Keith Ansell-Pearson, translated by Carol Diethe (Cambridge: Cambridge University Press, 2006).

[21] On the explanatory limits to relying on a notion of character see John M. Doris, *Lack of Character: Personality and Moral Behavior*, (New York: Cambridge University Press, 2002); and Timothy D. Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious* (Cambridge: Belknap Press, 2002).

tendency distorts our beliefs about why individuals act as they do, but it is often relied on by traditional fictional narratives, (perhaps it is essential to certain genres) in systematically eliciting from us insight into why characters in the fiction behave as they do.

I've appealed to the ways in general that fictions draw on our automatic and subdoxastic tendencies, including biases and heuristics, to elicit our imaginings of what is true in those fictions. A more specific account of such devices might show how particular kinds of biases may be indexed to the successful functioning of, respectively, particular categories or genres of fictions. For example, there is the bias of the 'hot hand' in which we unjustifiably tend to believe that gamblers or ball players can enjoy streaks, that they can be 'on a roll,' or 'in the zone' where these aren't merely short runs in a random process.[22] This may be the result of a confirmation bias, but whatever the psychological explanation, it seems tailor made for every film about an underdog team trying to make it to the championships.

There are studies of what has been called the 'Rhyme-as-Reason Effect' in which statements that rhyme are taken to be more truthful or insightful than those that don't even when the meaning is the same.[23] This may be an instance of a more general phenomenon in which a statement's truth is unwittingly evaluated on aesthetic terms.[24] In any case, that seems a cognitive bias made for pop, rock and hip-hop songs where in the midst of absorption we exhibit cognitive, affective, and behavioral cues that suggest we imaginatively endorse, say, a singer's genre-typical promise that love is eternal and unconditional or that violence and mayhem are the only answer to society's ills, even though we would not endorse such claims if we subjected them to scrutiny as candidates for belief. One might worry whether, in being largely constituted by an emotional response, such imaginative endorsement of the lyrics of such songs exhibits any cognitive content. However, if the operative emotions

[22] Amos Tversky and Daniel Kahneman, 'Belief in the law of small numbers', *Psychological Bulletin* **76**.2 (1971), 105–110.

[23] Compare Nietzsche's remark: '[E]ven the wisest of us occasionally becomes a fool for rhythm, if only insofar as he *feels* a thought to be *truer* when it has a metric form and presents itself with a divine hop, skip, and jump.' *The Gay Science,* edited by Bernard Williams, translated by Josefine Nauckhoff (Cambridge: Cambridge University Press, 2001), 85–6.

[24] See Matthew S. McGlone and Jessica Tofighbakhsh, 'The Keats Heuristic: Rhyme as Reason in Aphorism Interpretation', *Poetics* **26**.4 (1999), 235–244.

impute descriptive or evaluative facts to their objects, as I suggest they do, they should count as instantiating judgments.

In many of these cases we implicitly ascribe the facts to a fictional world that would rationalize our irrational responses. But this rationalizing, if adopted in relation to our beliefs, would be only of a spurious sort. For in the cases I've described we do not discover genuine evidence of what we imagine to be true, but, rather, are caused to impute that evidence to a fiction and treat it as existing there independent of our imagining. In this respect, our tendency to interpret what we perceive in a way that preserves our rationality – treating, e.g., our physically-caused "moral" disgust for a character as a correct recognition of vices that would genuinely warrant moral disgust – exemplifies the widely studied phenomenon of cognitive dissonance: people are systematically motivated to reduce the dissonance among their cognitions, even when doing so isn't rationally warranted by their sources. Employing spurious rationalization aimed at reducing cognitive dissonance is not a truth-conducive manner of forming beliefs. However, when provoked by a work of fiction, such rationalizing can be a reliable truth-in-fiction-conducive cognitive process.

Artists may exploit not only our tendencies to think in these ways to make certain things true in their fictions but also to create unexpected or ironic discoveries, such as that a character has qualities that run counter to what our automatic responses would impute to him.

It runs counter, for example, to our implicit assumptions about evil to find that Milton's Satan is attractive and charismatic – he is described in topoi more fitting of a heroic figure like Achilles or Aeneas – and has none of the appearance of a foul fiend. Yet, Satan is diminished as an object of fascination as Milton's epic comes to enlighten us of his real nature. Compare the obverse phenomenon in Alexander Nehamas' account of how the initial visual repulsiveness of John Merrick, the title character in *The Elephant Man* (1980), is diminished as we come to empathize with him. His appearance is enhanced as we come to better recognize his dignity.[25]

The epistemic errors and departures from rationality exemplified in the activation of these tendencies are importantly, for my purposes, *systematic*. If we were not systematically biased or irrational in certain ways in forming our beliefs, creators of works of art could not *predictably* exploit such tendencies and rely on them to direct our imagining

[25] Alexander Nehamas, *Only a Promise of Happiness: the Place of Beauty in a World of Art* (Princeton, NJ: Princeton University Press, 2007), 59.

of what is true in a work. Our imagining from such tendencies would not be truth-in-fiction tracking. An author wouldn't know that – or wouldn't write in such a way that reflects an implicit awareness that – she could elicit the desired response, the desired emotion or imagining, unless it was likely that her readers would have such biases. Furthermore, while many of the cases I describe illustrate how a work can prime us to attribute certain truths to the fiction, such priming is not just an *a*rational causing to which the application of epistemic norms would be irrelevant, i.e., a category mistake. Rather, like the exploitation of various irrational tendencies, such priming employed by fictions can be systematically directed at particular ends. Specifically, unlike the cases in psychology experiments in which a subject may be differentially primed to adopt one of a plurality of different perspectives on some essentially ambiguous state of affairs, the priming performed by a fiction is usually systemically directed at the discovery of what is true in it, what sorts of things it is *correct* to imagine.[26]

IV.

One way for a defender of continuity to respond to the examples arrayed above would be to say that while they illustrate how the epistemic norms that govern our beliefs don't always govern our imaginings in accord with a fiction, this only shows that such imaginings are often epistemically unjustified or irrational. Authors exploit some of our irrational tendencies to cause us to imagine certain things to be true in a fiction that we would not, on the basis of like causes in real life, be justified in believing. There is no discontinuity in epistemic norms if the illustrations above confirm only that we are by and large epistemically rational in what we believe but epistemically irrational in some of what we imagine.

The problem with this way of describing such cases is that we need to preserve a distinction between instances in which our responses to a fiction are epistemically rational and others in which they are, indeed, irrational. In some cases, that is, fictions are designed to exploit our subdoxastic tendencies to reliably cause us to recognize what is true in a story. What we imagine is the output of a reliable process by

[26] See John A. Bargh and Tanya L. Chartrand, 'Studying the mind in the middle: A Practical Guide to Priming and Automaticity Research' in Harry T. Reis, ed., *Handbook of research methods in social and personality psychology* (New York: Cambridge University Press, 2000), 253–285.

which we discover what is true in the fiction. In other cases, those tendencies lead us to misunderstand a story, to attribute facts to it that do not hold, as when because of one's own irrational racial prejudices one fails to recognize that it is true in a story that a character has certain virtues. If a defender of continuity holds that all (otherwise unsupported) imaginings we form due to the activation of such tendencies are epistemically irrational, we lose the ability to identify those *distinctively* irrational responses to a fiction that do not result from a systematic means of discovering what is true in it.

Another approach for a defender of continuity might be to say that the examples above do not show that we are epistemically irrational in what we imagine in responding to a fictions; rather, they demonstrate how, we are epistemically rational in inferring what is true in a fiction from recognizing how the fiction is designed to affect us. One might suggest, for example, that in the illustrations above, we rely on external factors of the work to imagine what is fictionally true in the same epistemically warranted way in which we form ordinary beliefs: I come to infer that something is true in a fiction from my recognition that the author or artist has designed it in such a way as to make my discovery of that truth possible. But not all such elicitations to imagine function this way. For there are two kinds of cases here: one that poses no threat to continuity is represented by the case, where characters are named so as to give us reason to believe something about their qualities. There, our imagining that the character has that quality follows from an ordinary rational process of relying on the stipulation of the author or testimony of the reliable narrator. The nouveau-riche Veneerings, in Dicken's story *Our Mutual Friend,* really do live a life of superficial gloss, Daffy Duck really is a daffy duck, and Thwackum, the tutor in Fielding's *Tom Jones*, does have a penchant for the cane. In the second kind of case, however, the names of characters would not give us reason to believe something about their qualities, but, instead, are designed to cause us to attribute those qualities to those characters, through, e.g., activating stereotypes or implicit associations. The fiction presents a character or state of affairs as having certain features and our response to those features causes us to correctly imagine the presence of other features too.

Finally, a defender of continuity might say: if it's true in the fiction that Fagin is morally corrupt, the beautiful person is intelligent and honest, the mobster's ends are merited, and so on, then that justifies imagining such things as true. *Whatever* the means might happen to be that such fictional truths are conveyed to us, they are fictional truths and therefore we are justified in imagining them as such, just

as we are justified in believing whatever is true. But that does not employ an adequate concept of justification. A belief that *P* is not justified on the basis of '*P*' being true if having the belief does not come about in the right way. It must not be, e.g., an accident, a knock on the head, or a deviant causal chain that explains why one believes that *P* if that belief is to be justified. Speaking of a fiction from the internal perspective, one does not have direct access to any non-stipulated facts; the only internal evidence one has for imagining what is true within a fiction is what else one imagines to be true in the fiction. The important point is that rational norms governing beliefs do not speak directly to their contents in isolation, but rather to the reasons in favor of the formation or retirement of those beliefs, and to their relations – such as their consistency-while they are held.

V.

Although in making the case for discontinuity I've referred to the various tendencies exploited by fictions as irrational dispositions and the like, it is a mistake to assume that they are always defects, or flaws, or evidence of improper functioning in our reasoning. There may have been evolutionary trade-offs that produced these forms of cognition and behavior so as to allow other beneficial forms. They may, say, reflect evolutionary history not being able to pass over a fitness valley required to attain a more optimal state. And they may reflect asymmetries in the cost of making an error in judgment and the benefit of getting it right ('better safe than sorry' is a low-cost/high-benefit policy when deciding whether a snake in the wild is dangerous).

My interest is in how the suboptimal aspect of these tendencies means we *try* to correct for them when we can – when it would be irrational not to – in theoretical reasoning with our beliefs. But we do not recognize an epistemic norm calling for us to engage in such correction in response to fictions, when, that is, such ways of thinking are exploited by the fiction in order to reveal what in the fiction is true.

That our epistemic norms for believing and imagining are discontinuous stems, I think, from how the kinds of reasons we countenance as justifications for our cognitive representations depend on the *functions* of the practices in which those representations are formed.

Thus, as we have seen, beliefs are typically directed at accurately representing things as they are.[27] Accordingly, the only reasons that

[27] Exceptions may be found in the sort of motivated believing and reasoning involved in thinking of oneself as a better athlete than one is in order

count in favor of a belief in its representational dimension are evidential reasons, those that speak to its truth and to the reliable means by which that truth is obtained.

Analogously, in some cases one's imaginative activity has an epistemic or practical role analogous to that of belief and perception, where the function is to aid in discovering some truth about the actual world, as when we need to plan for the future. That purpose is better realized if one's imaginings are based on reasons that speak to the objective qualities of, and relations among, what they represent.

Many fiction-directed imaginings, however, are generated in activities with ends – such as pleasure, entertainment, and absorption – in virtue of which they can be epistemically warranted on grounds that would not count as justifications for analogous beliefs. One may decide that a character in a film is trustworthy because she has a, so-to-speak, honest face (notice the familiarity of that expression). Even if that judgment is not justified by an inference from any facts imagined to hold in the fiction, it may be still be justified if it is part of the design of the work that it induces audiences to see that character as having that virtue.

It should be clear that in speaking of the rational norms governing fiction-directed imagining my aim has not been to address the canonical philosophical question of whether fictional works of literature are a good source of insight or understanding vis-à-vis the real world. However, I do think that my defense of discontinuity poses certain problems for those who attempt to treat our responses to fictional scenarios as a source of evidence for how we respond to like situations in real life.

I will describe only one such problem here, which concerns the widespread practice in the psychological study of cognition to use fictional narratives and films to ascertain the nature of such things as memory, the emotions, inferences, and perceptions.[28] The problem is that if individuals in those studies have internalized different

---

to perform better than one would in light of a wholly accurate appraisal, or in the various paradoxes of rationality in which one is motivated to adopt an attitude of belief toward what one does not believe. However, it is plausibly a conceptual constraint on the identification of a given attitude toward some content as constituting a belief that it is governed in some sense by a norm of truth. Cf. Nishi Shah and David Velleman, 'Doxastic Deliberation', *Philosophical Review*, **114**:4, 2005, 497–534

[28] See, as an example, James Gross and Robert W. Levenson, 'Emotion Elicitation Using Films', *Cognition and Emotion* **9** (1995), 87–108. For a favorable discussion of this method see Keith Oatley, et al., 'The Psychology of Fiction: Present and Future,' in I. Jaén and J. Simon, eds.,

criteria of warrant or justification – different norms – for their imagining what is true in the fiction and their believing what is true outside of it, their responses to the scenarios may differ according to whether they construe the events as real, or as merely the content of a fictional artistic representation.

A subject, for example, may allow herself to be affected by formal and stylistic aspects of a description – and not put the brakes on her automatic and usually distorting responses – when she approaches it as a fictional work of art but may try limit those effects on her responses insofar as she believes the description is intended to be a representation of the real world. It's been shown that people often do this: unwittingly adjust their truth-governed mental representations to take account of the effects of distorting forces if they've been primed to be aware of those effects. Individuals, for example, who were asked about their current level of happiness gave less negative responses when they were primed to be aware of the day's bad weather (and its presumed effect, without any explicit connection being made, on that judgment).[29]

Of course, if we do generally respond to fictions in a way that gives free rein to our biases and automatic subdoxastic tendencies, this does not mean that, in all cases, we ought to. I denied earlier that reasons that speak to the moral, aesthetic, or instrumental aspects of an imagining can serve as warrants for the *representational correctness* of that imagining. But, of course, non-epistemic reasons may be built into a theory of the proper ends as a whole that we should have in engaging with fictions. Perhaps we should not allow ourselves to have our irrational tendencies exploited by a fiction when, for example, it is designed to trigger our highly fallible in-group biases and implicit associations in shaping our judgments about a character with stereotypical racial or ethnic characteristics. Likewise, we may have reasons of self-respect not to succumb to the overly sentimental or sure-fire causes of imagining relied on by kitsch. However, it is the insidious power of art that our better judgment doesn't always constrain what we may be elicited to imagine.

Columbia University
jsg8@columbia.edu

Cognitive Literary Studies (Austin, TX: University of Texas Press, 2012), 235–249.

[29] Norbert Schwarz and Gerald L. Clore, 'Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states', *Journal of Personality and Social Psychology* **45**.3 (1983), 513.