ABSTRACT: This paper’s main thesis is that in virtue of being believable, a believable novel makes an indirect transcendental argument telling us something about the real world of human psychology, action, and society. Three related objections are addressed: a Stroud-type philosophical objection—as well as an empirical objection—questioning the force of this kind of transcendental argument, and the objection that a version of ‘the paradox of fiction’ applies to this account.

KEYWORDS: Currie, narrative, novels, paradox of fiction, Stroud, transcendental argument, truth in fiction

1. INTRODUCTION

This paper’s main thesis is that in virtue of being believable, a believable novel makes an indirect argument telling us something about the real world of human psychology, action, and society. This involves that believable novels are arguments, not in the sense that they are stories that explicitly offer arguments (perhaps didactically or polemically), but in the sense that, as wholes, they indirectly exhibit the distinctive structure of a kind of transcendental argument. As applied here, Stroud’s influential objection (1968) to transcendental arguments would be that from believability, the only conclusion that could be licensed concerns how we must think or conceive of the real world. Moreover, Currie holds that such notions are probably false: the empirical evidence “is all against this idea…that readers’ emotional responses track the real causal relations between things” (2011b). Finally, a version of the ‘the paradox of fiction’ pertains. Certainly, responding with a full range of emotions to a novel requires that it be believable. Yet since we know the novel is fiction, we do not believe it. So in what does its believability consist? This paper will address these three related objections.¹

I start with the idea that believability is ‘the master criterion of the novel’ (as one reviewer of an ancestor of this paper put it), or at least is a central criterion of assessment. It is always reasonable to ask about a novel—is it successful ‘make-believe’? No doubt the distinctive power and sweep of the novel is its unrivaled potentiality for intricate plot and associated character development. But for any believable plot/character development complex, we can ask—what principles or generalizations would have to be true about the real world (of human psychology, action, and society) in order for the fictional complex to be believable? Because this also always seems a reasonable question to ask, and because it can be an unanalyzed datum or given that a novel is indeed believable, the following transcendental argument scheme is generated:

¹ While this paper addresses these three possible objections, in two previous papers I consider other issues that arise in understanding some novels to be arguments (2011; 2012).
(1) This story (complex) is believable.
(2) This story is believable only if such and such principles operate in the real world.
(3) Therefore, such and such principles operate in the real world.

The believability premise, (1), is a proposition about the novel; it is not a self-referential claim made by the novel (although in degenerate cases such as parts of Henry Fielding’s Tom Jones the novel seems to be explicitly claiming about itself that it is believable). If (1) were an implicit or explicit claim made by the novel, the question of whether this claim itself is believable would arise, and so on into an unpleasant regress. The idea is that in virtue of being believable (not claiming to be believable), a novel makes an argument telling us something about the real world. (2) expresses the specific inference license or rule that allows a novel to be an argument, according to the present theory; it is not something that any novelist need intend or even be aware of. The idea is that the believability of a novel requires that certain principles or generalizations be true about the actual world. (3) is the conclusion. It indicates which principles operate in the real world, which is primarily of human nature given the subject matter of novels.

For illustration, consider Nussbaum (1990, pp. 139-140) on Henry James’ The Golden Bowl:

The claim that our loves and commitments are so related that infidelity and failure of response are more or less inevitable features even of the best examples of loving is a claim for which a philosophical text would have a hard time mounting direct argument. It is only when, as here, we study the loves and attentions of a finely responsive mind such as Maggie’s, through all the contingent complexities of a tangled human life, that . . .we have something like a persuasive argument that these features hold of human life in general.

As applied here, (3) is the generalized (and rosy) “claim that our loves and commitments are so related that infidelity and failure of response are more or less inevitable features even of the best examples of loving,” which is implicated by the believability of the plot/character development complex: “the loves and attentions of a finely responsive mind such as Maggie’s, through all the contingent complexities of a tangled human life.”

The Nussbaum quotation also illustrates what is not all that uncommon: a vague, undeveloped recognition of the (transcendental) structure of the argument of a novel. Here is another example: Rodden (2008, p. 155) says “in more didactic novels such as George Orwell’s 1984, we are often aware of a presence arranging and evaluating ideas and characters in building a convincing argument.” I am trying to shed some light on how characters can be ‘arranged’ into an argument, not, trivially, how (e.g.) the speeches of characters sometimes overtly state arguments.

These considerations mean that (1)-(3) constitute a schematic meta-level representation of the argument of a believable novel, which, at the object level, is only indirectly expressed by the novel.

2. BELIEVABILITY AND THE PARADOX OF FICTION

In what does believability consist? A novel’s believability seems to be determined mostly by what can be called the ‘internal’ and ‘external’ coherence of the event complex. I take Schultz (1979, p. 233) to be succinctly explicating internal coherence where he says: “the events must be motivated in terms of one another. . .either one event is a causal (or otherwise probable) consequence of another; or some events [sic] happening provides a character with a reason or
motive for making another event happen” (cf., e.g., Cebik, 1971, p. 16). A novel is not believable if in it things keep happening for no apparent reason or in a way that is inadequately connected with the other events in the novel. Certainly, this applies to some degree to James Joyce’s *Ulysses* and William Burroughs’ *Naked Lunch*, for example.

But even if the events of a novel are fully connected, the novel may still not be believable because those connections do not cohere well with our widely shared basic assumptions about how human psychology and society not only actually, but necessarily work. This is the main component of external coherence. The believability of a novel requires that its plot and characters be developed in ways that generally conform to our fundamental shared assumptions about human nature. It might be wondered whether there is circularity here. I am saying both that the believability of a novel requires this kind of external coherence and, with the rule of inference (2) above, that the believability of a novel implicates certain truths of human nature. However, it seems there is no pernicious circularity, mainly because both of these statements are meta-level generalities. Even though at the object level a given novel’s specific argument is only indirectly made by the novel itself, the reader or reviewer can summarize how the argument proceeds. And in this summary, there is no appearance of circularity. The summary starts with the unadorned premise that the novel—let Henry James’ *The Golden Bowl* again be the example—is believable. It seems that generally, believability is experienced by the reader as a simple, unanalyzed datum or measure of the novel, continuously updated as the reader progresses through the novel and imaginatively engages with it. And, like Aristotle said about judging the happiness of a person, you do not know for sure about believability until you reach the novel’s end. Of course, a few paragraphs back, there is already a conveniently short abbreviation of the remainder of this novel’s argument. Put another way, the experience of a novel’s believability is one thing, determining which specific truths of human nature are implicated may be quite another and may lie in the province of literary criticism.

A novel does not have to be realistic in order to be believable. The events of a novel can be far-fetched or remote, as in a science fiction, fantasy, or allegorical novel. Extremism of this sort seems to have little effect on believability so long as the events related are reasonably well-connected, and our fundamental shared assumptions about human nature, and about physical nature of course, are generally respected. Even with substantial alterations in fiction of physical or psychic reality, if the author’s development of these alterations is internally consistent and coherent and exhibits firm suspension of the author’s disbelief, and if the author successfully depicts the characters as believing what is going on as if it is normal, this can make the novel believable for the reader. (The author in effect says, ‘suppose for the sake of argument…’) There may be a kind of transference or transitivity of the suspension of disbelief here. For such a novel, trusting the characters and watching them for signs seems analogous to watching flight attendants for signs the flight is going well or badly—a kind of ‘reality check’, as it were. On the other hand, a novelist may push the envelope regarding physical nature (a possible example is H. P. Lovecraft’s novella *The Call of Cthulhu*) or psychic reality (a possible example is Max Beerbohm’s *Zukeila Dobson*), to the point where neither we, nor the characters, nor the author really understand what is going on. Here, believability breaks down, and consequently, no argument can get off the ground.

In using Coleridge’s (1817, p. 314) phrase “suspension of disbelief” here, I do not mean to suggest that the believability of a novel involves believing that its event complex is true; rather, it involves believing that the event complex could have been true in a strong sense of ‘could’—stronger, for example, than that of mere logical possibility. As Aristotle famously said,
“the poet’s job is not to tell what has happened but the kind of things that can happen, i.e., the kind of events that are possible according to probability or necessity” (Poetics, Ch. 9). So while nonfictional narration (history, biography, etc.) aims at veracity, the novel aims at verisimilitude or depicting events and characters “according to probability or necessity,” which I would explain as determined principally by internal and external coherence.2

This approach suggests a solution to the much-discussed ‘paradox of fiction/of fictional emotions’. It certainly seems that the believability of a novel and our emotional response to the novel are interrelated: a novel’s being believable allows responding to it with a full range of emotions, or conversely, responding with a full range of emotions to a novel requires that it be believable. Yet since we know it is fiction, we do not believe it. So how can it be both steadfastly unbelieved and believable—known to be false and (e.g.) a tear-jerker? More formally, the paradox of fiction is that although all three of the following propositions seem plausible, they cannot all be true:

a: We have genuine emotional responses to certain fictional narratives.
b: We believe that those narratives are fictional.
c: (a) and (b) are incompatible (each implies the denial of the other).

Thus, solutions typically deny one or the other of these three propositions. What are generally regarded as implausible or distorted solutions, either deny (a), as in the case of Walton’s postulation of “quasi-emotions” (e.g., 1978), or they deny (b) (e.g., Suits, 2006).

The solution suggested by the above, like most solutions, denies (c), but I think it uniquely gives believability a prominent role. It is a possible-world solution. We believe that the plot/character development complex (event complex) of a novel is not real because we know that generally it is a merely possible (nonactual) world constructed by the novelist. However, for a believable novel, the possible world constructed by the novelist is strongly ‘accessible’ from the actual world, where the core idea of one world being accessible to another is that the one is possible given the facts of the other—in this case, notably, the basic facts of human nature. The basic facts of human nature are held common across the worlds. Thus, accessibility grounds believability, which in turn grounds emotional response. Although believability requires that perceived fundamental facts of human (and physical) nature be respected, a novel is a complex counterfactual. But it is commonplace that we have emotional responses, unquestionably genuine, to all manner of situations that are not presently actual—and so are counterfactual in at least this sense. Indeed, it is hard to see how there could be practical reasoning without such responses.

I don’t know about you, but I fear a stock market crash. This fear fully motivates me to take measures to minimize the financial damage to me should a crash occur. It may be that the particular kind of crash that I fear has not and will never in fact occur (though it could be significantly probable), and so, unbeknownst to me, the possibility is metaphysical and not merely epistemic (‘for all I know, we’re in for a crash’). Of course, the counterfactuals of a novel are generally metaphysical—the events and characters depicted have not and will never occur or exist (in the actual world). But this is by no means always the case. For example, consider some of the events of From the Earth to the Moon by Jules Verne, or consider any historical fiction.

2 The distinction between nonfictional and fictional narration with respect to believability may not be as sharp as suggested here. Olmos (2014; forthcoming) proposes a general account of credibility that covers both types of narration.
My key point is that it seems to make little if any difference to our emotional response whether the possibilities (counterfactuals) we consider are perceived to be metaphysical or epistemic, fictional or temporary, so long as they are believable. The critical link and parity among them is that they are all creatures of the imagination, wherein their believability is determined. However, the perceived status of the possibility may of course make a big difference in our behavioral response. Adapting a favorite example, we may be horrified by the events depicted in a horror film because they are believable; yet because we don’t believe them, we don’t flee the theater. In other words, we don’t flee the theater because we know the possible world of the horror film is metaphysical, in relevant ways. Failing to adequately take into account such differences in behavioral response perhaps (confusedly) leads to thinking that emotional responses to fiction are themselves qualitatively distinctive or are only “quasi-emotions.”

We use our emotional—or more generally, affective—responses to different possible courses of future actions or events (and their potential consequences) to help test them out and select among them where we have a choice, or to be prepared where what will happen is out of our control. The thought of such a possibility may bring fear, anger, disgust, anxiety, interest, arousal, joy, or whatever, but the bottom line seems to be that “emotions” have a “cognitive dimension” in that they “embody some of our most deeply rooted views about what has importance, views that could easily be lost from sight during sophisticated intellectual reasoning” (Nussbaum, 1990, p. 42; cf. Johnston, 2001). Such affective responses to fictional possibilities figure in the contribution that reading novels makes to enhancing practical reasoning skill, which is by “offering us the opportunity to practice thinking about difficult and interesting situations and complex personalities and providing us with examples of how to discriminate salient features of such situations and characters” (Depaul, 1988, p. 563; cf. also Clark, 1980, and Gendler & Kovakovich, 2006 for some similarities to the approach I take here).

3. THE STROUD-TYPE PHILOSOPHICAL OBJECTION

Transcendental arguments on the order, for example, of Davidson’s directed against skepticism about other minds (1991, pp. 159-160), reason that since certain aspects of our experience or inner world are undeniable, the external world must have certain features, on the grounds that its having these features is a necessary condition of our experience being the way it is. In my representation, the argument of a believable novel is of this type. Stroud (1968) famously objected to such transcendental arguments that they are too ‘ambitious’ (the terminology is Stern’s, 2007)—that the only condition and conclusion that could be licensed is that we must think or conceive of the external world as having certain features, not that it actually does. The objection as applied to the novels case is that it would be enough to allow our experience of believability if having this experience implicated only that we perceive the real world as operating in accordance with certain principles.

The first thing to note in response is that this ‘modest’ version of the transcendental argument of a believable novel is still an argument; there is still an argument whether we take “real world” in (2) and (3) of the schematic representation above to refer to the real world simpliciter or to how we must conceive of the real world.

Second, no doubt in certain cases I may find a novel believable, whereas you do not. But I think that there is no fundamental relativity of believability because there is such a thing as human nature, which we all share and to which we have significant introspective or ‘privileged’
access, or at least psychological attunement. The believable novel taps into and relies on these facts, bringing operant principles to the fore. If this general idea were not true, then it would be pretty inexplicable that there is widespread agreement about which novels are good novels.

Being believable is a central necessary condition for a novel to be a good novel. So in the case of the ambitious version of the argument of a novel that began this paper, the leap from the inner to outer worlds is limited and facilitated. The leap is from our psychological experience of believability of the novel to the real world of human psychology, action, and society—which is the primary subject matter of all novels. This subject matter is basically human nature, I take it. The inner and outer worlds of the ambitions argument are significantly the same; it is not as if the worlds are distinct as, for example, thought and a brain in a vat, as in Putnam’s memorable transcendental argument (1981, Ch. 1). And, as Nagel (1979, Ch. 12) forcefully argued, because after all we are human, we know what it is like to be human in a way we do not know what it is like to have a different nature, such as a bat’s (and perceive the world primarily through echolocation, be capable of flying, etc.).

Such philosophical considerations indicate that the principles identified in the argument of the novel resonate in believability largely because they are true of human nature; they indicate that some ambitious version of the argument of the novel is justified.

4. THE CURRIE-TYPE EMPIRICAL OBJECTION

In recent years, Currie has made something of a cottage industry for himself questioning such claims on empirical grounds—questioning, as he likes to put it, ‘whether we learn about the mind from literature’. It is no doubt common to think that we do so-learn; for example, consider Lehrer’s 2007 book Proust Was a Neuroscientist. Currie’s writing on this topic includes pieces in the popular press (2011a; 2011b; 2013). Perhaps his most strident, though scholarly, articulation of his view is this (2012, p. 30):

And could [Samuel] Johnson have been rationally confident that Shakespeare has shown how human nature acts in real exigencies, when he, Johnson, carried out no surveys, no carefully structured experiments, to find out whether it really was so? Johnson was delightfully confident in his opinions of many things, and rated himself a great observer of his fellow creatures, but the last 50 years of psychological investigation has shown how often we are wrong about our own motives and actions, and those of others, and how little penetrating intellect and common sense can help us overcome our ignorance. When Leavis says, rather grudgingly, that Hard Times does not give “a misleading representation of human nature” (Leavis 1948, p. 233) it is tempting - to ask how he could possibly know something that not even the greatest psychologist would think of claiming: what human nature is.

Of course my answer to Currie’s last point is that the believability of Hard Times has something to do with it. Currie’s view constitutes a challenge to my claim that some ambitious version of the transcendental argument of a believable novel is justified, which would require that our conceptions of human nature are generally true. Again, I claim that the believable novel taps into and relies on these conceptions, bringing operant principles to the fore.

---

3 A recent influential article on introspection (Schwitzgebel, 2008) poses little threat to my points here concerning human nature and its operant principles, because the focus of the article is on the untrustworthiness of introspection of immediate conscious experience.

Differences among readers in the perceived believability of a novel may be largely attributable to relatively extraneous factors, such as the setting of the novel. For example, if I could get past the fantastic details of Tolkein’s trilogy, I think I could better appreciate these novels as implicating truths of human nature.
Let us for the moment try to step back from the possible detail of “surveys” and “carefully structured experiments” and look at the big picture. By virtually any biological measure such as population numbers and adaptability to different environments, *Homo sapiens* are an extremely and uniquely successful social species. (Indeed, we are so successful that in some ways we are victims of our own success: overpopulation, pollution, etc.) Is it not obvious that this success would not be possible if we were largely “wrong about our own motives and actions, and those of others” or in general about our conceptions of human nature, and if “penetrating intellect and common sense” were of little use in augmenting self- and social knowledge? We know ourselves and others and the operant psychological/sociological principles or generalizations well enough that our actions and interactions are mostly predictable, often drearily so. Our fundamental, shared conceptions of human nature allow us to function and flourish, and this is evidence of their (at least approximate) truth, in much the same way that the spectacular success of the physical sciences in their predictive power and technological applications (“they work”) is evidence of their (at least approximate) truth.

This seems to be so even if something like epiphenomenalism is true, whereby our conscious and self-conscious life is not causally efficacious in the physical world. As far as any competition for world domination by a “social” creature goes, ants are perhaps our only real rival. But we are sharply unlike ants. We have a mental life, and it is a rich mental life. It is hard to see how we could function and flourish if our mental life were so out of sync with reality as in Currie’s bleak picture, even if mental processes only attend physical processes—where the real action is. It seems that such a mental life would consist largely of bewilderment and confusion.

But epiphenomenalism is a radical view. Suppose rather that conceptualization and thinking come to the fore and are causally efficacious primarily in such things as problem-solving, including in response to when we act or interact in unexpected fashion, and that otherwise we mostly unthinkingly function with reliable ‘animal’ expectations of our behaviors (behaviors that are predictable by us but not predicted). This seems to be more like what is actually the case. Yet of course it is then all the more implausible that we could function and flourish and our mental life be so out of sync with reality as in Currie’s view.

None of this is like a suspicious evolutionary argument about the origins of some specific creature feature. One may easily get tangled up in alternative possible explanations of particular adaptations. For example, at one point paleontologists thought that the regression relationship between the dorsal fin area and the body volume of the pelycosaur showed that this ‘sail’ fin was a temperature-regulating mechanism. Later, this explanation was more or less replaced by the behavioral explanation that the fin was used for sexual display (Gould, 2007, p. 253). Of course it could have had both functions, or neither. Our conceptions of human nature, as a whole, lie at an altogether different level. There is no alternative possible explanation of their existence and entrenchment other than that they have evolved in answer to millions of years of human needs. So what are the kinds of psychological “surveys” and “carefully structured experiments” Currie uses to make his case that our conceptions of human nature are largely wrong, that “our insight into the mind generally is very limited”? One is the “imagined professor” experiment, which indicates that to do better at a game of Trivial Pursuit, for example, imagining a professor helps, whereas imagining a soccer hooligan hurts. This is supposedly surprising, and illustrates that “our minds are prone to capture by unconscious imitation.” More significantly, this principle is said to be borne out in the strong empirical evidence of a causal relationship between “media violence and imitative aggression,” about which there is supposedly a huge “disconnect between research results and public opinion” (2010, p. 201).
Another allegation is that folk psychology, like the novel, believes in character and character explanations, and that makes us prone to error, as when we “infer good character from attractive appearance.” Experiments suggest that small changes in circumstances can make a big difference in our behavior, as where “people who have just found a dime in a phone booth” are a lot more likely to assist someone outside in need of help than those who had no such luck. (I know the example is dated!) Of seminarians on their way to “give a short talk on the parable of the Good Samaritan,” one group was told there was “no hurry,” and the rest that they were “slightly late.” On the way, “a confederate faked a collapse.” Compared to those in the no-hurry group, the others were a great deal more likely to ignore and even step over the collapsed person (2010, pp. 202-203).

Another allegation is that our minds are prone to illegitimately link the literal and the metaphorical, as in the case of “our ready use of a warmth-coldness scale for persons…from developmentally important experiences of physical closeness to caregivers.” If you briefly hold a hot cup of coffee, you are then more likely to behave generously and classify a person with whom you are interacting as “warm” (2010, p. 204).

It seems that each of these specific allegations is to some degree disputable, but I will not do that here. Similarly, I will not respond to Currie’s ad hominem against novelists and other creative people; for example, he cites “a mid-1990s study of creative groups which found that only one of fifty writers (Maupassant) was free of psychopathology” (2011b). It should be enough to point out that compared to the reasons for believing that our conceptions of human nature, on the whole, allow us to function and flourish, the kind of evidence of detail that Currie presents seems to be a case of not being able to see the forest for the trees. Indeed, it is hard to see how any amount of such evidence of detail would be equal to the task Currie assigns it.

To be sure, at a higher level, Currie says “we have little grounds on which to trust our folk-psychological theories—any more than we these days trust folk physics, which has been shown to be substantially at odds with scientifically informed theories of the interaction of bodies” (2010, pp. 201-202). Yet does this just confuse the general vagueness of folk psychology and folk physics with falsity, or is it trying to say what anyone should admit, that as you go from folk to scientific theories, the truths identified tend to become less approximate (where this trend is less clear or more plagued with historical exceptions in the “social” sciences)? Should we stay off the pyramids because the ancient Egyptians used folk physics? At perhaps a less exacting level than the pyramid builders, we are always or almost always interacting with bodies in ways that could reasonably be said to require our use of folk physics, e.g., cooking dinner, driving a car, or playing baseball. Current theoretical physics should undermine our trust here not one wit, or if it did, one wants to say ‘that way insanity doth lie’.

5. CONCLUSION

Finally, Currie says that “we have been strangely complacent in assuming that we do learn [from fiction], without any better evidence than our own feelings of having learned something” (2011a, p. 49). This paper has tried to show that, on the contrary, such feelings may be firmly grounded in the believability of the fiction, and all that is entailed by that, so the complacency is not strange. It is warranted.
ACKNOWLEDGEMENTS
This paper has benefited from correspondence with Trudy Govier, and from comments by Jason Dickenson, Teresa Plumer, and the audience at the ISSA reading. Some of the material in this paper derives in earlier form from my 2011 and 2012.

REFERENCES


