

**Iris Vidmar**

University of Rijeka, Faculty of Humanities and Social Sciences, Sveučilišna avenija 4, HR-51000 Rijeka  
ividmar@ffri.hr

**Thought Experiments, Hypotheses,  
and Cognitive Dimension of Literary Fiction**

**Abstract**

*Some authors defend literary cognitivism – the view that literary fiction is cognitively valuable – by drawing an analogy between cognitive values of thought experiments and literary fiction. In this paper my aim is to analyse the reasons for drawing this analogy and to see how far the analogy can be stretched. In the second part, I turn to the claim put forward by literary anti-cognitivists according to which literature can at best be the source of hypotheses, not of knowledge. I challenge this claim by showing that hypotheses can have valuable cognitive benefits on their own, thus hoping to restore cognitive benefits readers get from literature.*

**Key words**

cognitive values, hypotheses, literary cognitivism, realism, science fiction

**1. Introduction**

In this paper I want to explore the analogy, put forward by some literary cognitivists, between the cognitive values of literary fiction and thought experiments (TEs). Namely, in trying to explain the cognitive value of literature, some philosophers claim that literature functions in the same way as TEs in science and philosophy, and given that for the most part, the cognitive value (and benefits) of TEs is taken as justified, we should, on those same bases, accept that literature bears cognitive value. While I do accept that literature is cognitively valuable and that indeed there are many cognitive benefits we can gain from engaging with literature – in that sense, we will mostly speak of cognitive benefits and cognitive contribution of literature – I think we should further explore the analogy in order to see how far it takes us, but also to see how diverse the cognitive benefits available are. It is important to note that my aim in this paper is not to call into doubt any of the epistemic benefits attributed to TEs or to question different accounts that were put forward in order to explain them; for the most part, I rely on the account given by David Davies, who provides one of the most extensive analysis of this analogy.<sup>1</sup>

Literary cognitivists claim that literary fiction is cognitively valuable, which means there are some cognitive benefits we can get from engaging with it.

<sup>1</sup>

See Davies 2007a, 2007b, 2010a. Peter Swirski (2007) also offers and analyzes various accounts of TEs.

David Davies claims there are four main categories of cognitive values one can find in works of literary fiction:<sup>2</sup>

- (i) Knowledge of matters of facts (factual information about the world): at the most general level, works of literary fiction contain descriptions of different aspects of the world which readers can pick up and accommodate within their larger scheme. This can include rather detailed descriptions scattered throughout the novels which compose the background of the story (like Maria Edgeworth's *provincial novels* which depict different aspects of Irish society in the late 18<sup>th</sup> century, or Sir Walter Scott who "opened up the novel to the full panorama of revolution, dissent, rebellion and social change"<sup>3</sup>), or can be presented in just a few sentences describing particular city, event, holiday etc.
- (ii) Understanding of general principles (moral, metaphysical, psychological). The idea behind this claim is that works of literary fiction provide a suitable background against which different principles can be developed, taken from abstraction and presented in their complexities, given substance and developed through their different aspects. For example, many critics argue that Dickens' *Hard Times* can be read as calling into question the principle of utilitarianism.
- (iii) Source of categorical understanding: apart from enabling readers to deepen their understanding of general principles, literary works can also provide them with new conceptual framework, or categories, which can be applied to the real world. David Novitz claims that there are different skills, cognitive and practical, available to readers from literature.
- (iv) Affective knowledge: given the way literary works develop the story they present to the readers, literary cognitivists often claim that they are particularly suitable to show different aspects of emotional responses and complexities, which in turn helps readers understand what it feels like to be in particular circumstances. In engaging with the work, readers can see the situation from the "inside". Davies links this to moral growth of the readers claiming that this knowledge of what it would be like can bear upon "our ability to comprehend, and respond appropriately, to morally complex situations we encounter in the actual world".

Although developing an account of the cognitive dimension of literary works along these four lines is a commonplace in philosophy of literature, there are some authors who go even further and claim that literature has the capacity not only to give us knowledge and understanding, but also to make us more aware of or sensitive towards various aspects of the world we might have not noticed before.<sup>4</sup> It is also said sometimes that art generally, and literature in particular, can not only offer a perspective for readers' evaluation, but also broaden, challenge or argue for or against an established one.<sup>5</sup> I accept all of these claims; the lines between them might be a bit blurry, but certainly there is plausibility in the claim that different cognitive benefits and values are available to readers. The problem for the cognivist, however, is to show what it is that gives strength to these cognitive contributions, that is, what reasons we have for accepting them. Some philosophers have claimed that we can only accept claims like these if we have some kind of additional proof that they are true, that is, only after we have subjected them to further testing.<sup>6</sup> Given the strict epistemic conditions that need to be fulfilled in order for one to reach knowledge, these claims can at best be only hypotheses, at least until proved, confirmed or justified through some further testing. The problem then

becomes that of showing whether this testing process is internal or external to the process of reading and the answers to this are twofold. David Novitz argues that this testing is external to the process of reading, in that a reader is supposed to project the hypothesis into the world and accept it as valuable only if it proves correct. At this point, David Davies rightly emphasises the fact that we have to be careful and recognize different justificatory grounds for each of the category of claims literature can offer (i–iv above). In the case of (i) and (iv) reasons for accepting these as true can be either the readers' familiarity with them or some additional reason for finding the author reliable (from an internalist perspective) or the fact the author is in fact reliable in these matters. For (ii) and (iii) Davies claims that we should apply the same solution that grounds the cognitive value of thought experiments (TEs), the so-called moderate inflationist response:

"The suggestion then, is that the mental models through which readers comprehend fictional narratives also provide, through their mobilization of tacit or unarticulated knowledge of the world, a means of testing those claims to knowledge of the actual world that theorists have located in fictional narratives, and thereby validate the idea that fiction can be a genuine source of knowledge of the world."<sup>7</sup>

Given that according to this model we expand our knowledge by creating mental models that operate with the knowledge we already have, that is, cognitive resources we possess prior to reading, the testing process is internal, we do not need any extra evidence or argument to support the claims and we do in fact end up cognitively richer. Finally, Davies argues, if readers do in fact learn something from literary works, the justification for that should be grounded not in subjective feeling that the learning process has taken place, but in the right kind of unarticulated knowledge that has been mobilized. In that sense, "[l]earning from fictional narratives in this way requires 'reliability in the reader' rather than 'reliability in the author'"<sup>8</sup>

So this, in a nutshell, is the account of the cognitive value of literature put forward by those who defend it by invoking the analogy with TEs. What I want to do now is to expand it further. I will divide this in two parts: first, I want to see whether we can come up with some criteria that would specify which literary works in fact function as thought experiments. Second I want to say something more about the testing process, that is, about the value of hypotheses.

## 2. Literature as thought experiment

Davies is not alone in his claim that literature can function as extended TE. Noël Carroll, a well devoted defender of literary cognitivism, claims that

"... philosophy employs a gamut of techniques to produce knowledge and learning that are analogous to those found in literature. What I have in mind here specifically are thought ex-

<sup>2</sup> See Davies 2007, 2010. Other cognitivists accept more or less the same classification (see for example Novitz 1984, Gaut 2006).

<sup>3</sup> Carter and McRae 2001, 239.

<sup>4</sup> See Carroll 2002, Gaut 2006.

<sup>5</sup> See Kieran 1996, 2005.

<sup>6</sup> Novitz 1984, Davies 2007a, 2007b, 2010a.

<sup>7</sup> Davies 2007b, 44.

<sup>8</sup> Davies 2007, 163.

periments, examples and counterexamples that are often narrative and generally fictional in nature”.<sup>9</sup>

They aim at “mobilizing conceptual knowledge – a priori knowledge and/or the knowledge that underwrites our ability to apply concepts competently – in order to reach certain conclusions”.<sup>10</sup> The notion that plays crucial role in Carroll’s account is that of clarification. In describing different ways in which TEs (and obviously literature) are cognitively valuable, he says:

“They make connections – that were hitherto recessive or obscure – between what is already known and other parts of our cognitive stock. They illuminate the relevance of what is already known to the question at hand by refocusing that knowledge in a novel way. This counts as knowledge productions, because it clarifies linkages between parts of our cognitive map”.<sup>11</sup>

Apart from that, they can also raise counterexamples to well accepted theories, make argumentative points, motivate conceptual distinction and give counterexamples to widely accepted claims. All of these functions can, Carroll claims and I concur, be carried out by literary fiction.

One of the strongest formulations of the analogy comes from Daniel Dohrn, who claims that the process of interpretation is close to counterfactual thinking. Dohrn simply assumes that every literary work that comes close enough to real world scenarios to invoke the process of interpretation gets its cognitive value from engaging readers in counterfactual thinking. Narratives present possible scenarios which resemble real world scenarios in that they bring what is salient to view and thus help us understand the real world. His account also relies on the power of narrative to work with cognitive resources we have and mental models that we create, although his picture is an even wider one than the one Davies relies upon. Dohrn claims:

“The artistic activity of representing and conceptually evaluating fictional paradigm scenarios closely resembles the philosophical practice of counterfactual thought experiments (...) The author of a narrative may be regarded as making a TE: What would be the case if?”.<sup>12</sup>

One reason for claiming that all literature functions as TEs is the fact that we read literature as counterfactuals – as if the story happened, although we know that it did not, due to the fact that we, as readers, are invited to imagine that (something takes place) while knowing that it did not. Indeed, there is a sense in which this can be a strong reason for claiming that the cognitive value of literature derives from the same source as the cognitive value of TEs: when engaging with TEs, we are aware of the fact that the content is to be imagined and that it did not (and usually it cannot) happen. But I think this is wrong. For one thing, although some of the content described in literary works did not happen, some of it nevertheless did. Many aspects of many literary works offer us true accounts of real world state of affairs and in that sense it is just plain wrong to pretend that it did not happen. A more detailed analysis of the analogy between literature and TEs is found in Peter Swirski, who accepts the claim that literature gives us knowledge in the same way as TEs, i.e. that the ability of literary works to provide us with knowledge derives from their similarity to TEs. However, Swirski puts a restriction on the analogy and claims that only some works function in this way, while others do not:

“The basic premise behind *Of Literature and Knowledge* is that the capacity of literary fictions for generating nonfictional knowledge owes to their capacity for doing what philosophy and science do – generating thought experiments. Not that *all* knowledge in literature can be traced to thought experiments. Historical novels transmit knowledge of history in the same manner that historians transmit it.”<sup>13</sup>

I think this is a good place to start my analysis. Davies also accepts that only some literary works function in this way because only some provide “indirect support for a scientific or philosophical position by illustrating how certain things are *possible* given that position”.<sup>14</sup> In that sense some fiction functions as helpful, visualisable illustrations, not like TEs. My aim is to try and find criterion that would help us differentiate between those works that can function as TEs and those that cannot. In order to determine that, I think we should go back to the reasons for equating fiction and TE in the first place, that is, for drawing the analogy between the cognitive functioning of both of these. So the question is: in what sense is literary fiction similar (enough) to TEs for the analogy to work?

Dohrn and Swirski put forward the analogy in order to account for the problem of fictionality, which many see as the most pressing for literary cognitivism: how can we claim to be able to learn from literature given its fictive dimension? On the other hand, Carroll points out that the fictional dimension of TEs is never considered a problem for those who claim that TEs are cognitively valuable. According to Davies, the fictionality of literary fiction and TEs can be explained along two conditions: namely, the attitude of make-believe and the no-fidelity constraint. These are two footholds of *fictive utterance theory of fictionality*, according to which, for some narrative to be fictional, the attitude prescribed is not belief but pretence or make-belief, and the general aim of the author in composing the narrative is to tell a story, to fulfil some artistic aim, and the choice of propositions inserted into the narrative does not depend on whether or not they are true.<sup>15</sup> This is the basic similarity between literary fiction and TEs, and it enables literary cognitivists to ground the cognitive value of literary fiction in the similarity of functioning between literary fiction and TEs.

Davies defines (scientific) TE as taking the

“... form of short narratives in which various experimental procedures are described. Competent reader understands that these procedures have not been, and usually could not (for some appropriate modality) be, enacted. She is invited, however, to imagine or make-believe that these procedures are enacted and to conclude that certain consequences would ensue, where this is taken to bear upon a more general question, which is the topic of the TE”.<sup>16</sup>

This is how TEs are used in science. Let us now see whether we can, relying on this account of TEs, explain the cognitive functioning of literary fiction.

### 2.1. Sense 1: similarity in the way the content is presented

One possible criterion that we can extrapolate from Davies’ account of TEs is that TEs (that is, literary fictions which can function as TEs) use the procedures that have not been (and could not be) enacted. I think we can briefly expand these procedures so as to include different kinds of scenarios that de-

<sup>9</sup>  
Carroll 2002, 7.

<sup>10</sup>  
Ibid.

<sup>11</sup>  
Ibid., 8.

<sup>12</sup>  
Dohrn 2009, 39–40.

<sup>13</sup>  
Swirski 2007, 4.

<sup>14</sup>  
Davies 2010a, 52.

<sup>15</sup>  
This is obviously not so in case of nonfiction, where the insertion of propositions depends on whether or not they are true.

<sup>16</sup>  
Davies 2010a, 51.

scribe some set of circumstances and the actions of different agents within it. It seems plausible to claim that in imagining the content presented in this way, the reader has to be aware that it is precisely because of this inability to actually carry out what is described, that the whole set of propositions is being put forward in the first place. To go back to the analogy, it seems that it excludes the following types of literary fiction from the analogy:

- i) works pertaining to the genre of realism, including historical novels which rely upon historical accuracy;
- ii) works that (set out to) present or portray a certain political regime, historical events, social circumstances etc, as truly as possible;
- iii) works that, although may be highly unrealistic from our perspective, nevertheless reflect certain realistic elements or beliefs of the target audience of the time the works were written, works like Dante's *Inferno*, Poe's *Eureka*, Ibsen's *Ghosts*.

It seems to me that in all of these examples, the attitude of “imagine that p although you know that not p” is not the guiding principle in the composition. The fictional dimension is strongly paired with the principle of verisimilitude, according to which events portrayed in the literary fiction are close enough to the real world. In that sense, these works provide us with the factual information and matters of fact.<sup>17</sup> However, that is not all, not even the most important cognitive aspect of these works, and I have already said that the analogy between literary fiction and TEs is not evoked to explain factual knowledge, nor is defending factual knowledge the most pressing aspect of literary cognitivism. We do not read literature in order to get the factual information, but to follow the story, the point of the story, to see what the story has to tell us about life, world, other people and our own circumstances. Problem for literary cognitivist is how cognitive benefits along these lines are possible, and suggestion put forward by Carroll, Dorhn, Swirski and Davies is to invoke the analogy. But my claim is that the scope of this analogy should be carefully delineated. The problem is, it is not plausible to approach these kinds of works as if they were TEs. For one thing, most of the authors pertaining to this genre insisted on the principle of fidelity at all levels of descriptions. In *The Cambridge Introduction to Nineteenth-Century American Novel* we find the following characterization of the writers pertaining to this literary period:

“these novelists share a general conception of fiction as a detailed and accurate representation of historically specific characters and settings – their manners, ways of dress, speech patterns, social habits, main concerns, and topics of conversation.”<sup>18</sup>

The same dedication to details is seen in the choice of topics realist writers wrote about:

“The realists were also critical of what they saw as the previous generation’s squeamishness when it came to depicting the more common, harsh, or even vulgar aspects of life, such as adultery, crime, alcoholism, radical violence, labor strife, and political corruption. (...) [R]ealists depicted the excesses of capitalism, the plight of the poor, and the narrow or strained circumstances of women, black Americans, and immigrants.”<sup>19</sup>

Because of these features, I think it is plainly inappropriate to ask readers to approach these literary fictions as TEs. Obviously, in what they describe, they do not tackle “experimental procedures” that have not been and cannot be enacted. In fact, due to its nature, realism is further described as “generally empiricist in its orientation, privileging concrete examples of experience over totalizing systems of thought.”<sup>20</sup>

What remains, what literary works are excluded from the sense 1? Obviously, works that in some sense deviate from the verisimilitude principle, and that demand the readers adopt “imagine that *p* even though you know that *no-p*” attitude. This is so due to the nature of what they describe; it can be seen in the breaching of the laws of physics, in portraying patterns of human behaviour and relations that are radically different from our world, in creating alternative worlds and communities, etc. The relevant background against which readers address these works has to include awareness of this breach, as well as acknowledgment of how big it is. I think that the best examples are novels pertaining to science fiction, novels that rely on state-of-the-art scientific developments and exaggerate them even further, fantasy genre, horror stories, utopias and dystopias. Examples that come to mind are *Brave New World*, *1984*, *The Strange Case of Dr. Jekyll and Mr. Hyde*, *The Island of Dr. Moreau*, *Herland*, *The Female Man*, etc. Precondition of making sense of these works is the awareness of the fact that such scenarios have not been enacted, that is, at the time<sup>21</sup> these fictional stories were written, these scenarios were fictional. However, as I will show later, that doesn’t mean they are not cognitively rich.

The genre of science fiction is

“... arguably, and in several respects, the most challenging form of literature yet devised. (...) [I]t can be shown that reading (or even viewing) any form of science fiction does involve one extra *intellectual* step over and above those necessary for reading other forms of fiction”.<sup>22</sup>

What Shippey here primarily has in mind is the fact that science fiction presents to a reader what he calls *nova data*, that is, new things given:

“The basic building-block of science fiction is accordingly the *novum* – a discrete piece of information recognizable as not-true, but also as not-unlikely-true, not-flatly (in the current state of knowledge)-impossible”.<sup>23</sup>

I think that it is here that we come close to what Davies describes as “procedures that are not enacted” and it is because of this breach from the familiar

17

Notice that this is so even in case of works like Ibsen’s *Ghost* which present wrong facts regarding syphilis. Even though the facts are wrong, the reason for that is not Ibsen’s intention to lie or deceive, but it is the outcome of the wrong (scientific) account of syphilis Ibsen was relying upon, which was however the account considered true at the time the work was written. So although we are given false information, we still learn new things, for example what people in those days believed about syphilis.

18

Crane 2007, 156.

19

Ibid., 157.

20

Ibid., 157–158.

21

It is important to emphasize ‘at the time these stories were written’, a good explanation of that is Jules Verne’s novel *20 000 Leagues under the Sea*. At the time the novel was written,

the idea of living under the sea was unbelievable. Today, we do not make fuss about submarines. I think this explains why certain literary works that were once seen as depicting unrealistic scenarios and procedures stopped being treated in that way once we managed to enact these procedures in the real world. To read Jules Verne’s novels as TEs today would be strange, but at the same time reader has to be aware that at the time they were written they were presented as TE. Notice however that this is in no way distinctive of science fiction: in order to understand almost any literary work, a reader has to be aware of certain background (common knowledge regarding the social rules, political regimes, religious beliefs etc.) shared by the author and his target audience. This also explains why those literary works that are in any sense “false” are not considered as of less value.

22

Shippey 2005, 11.

23

Ibid., 13–14.

that science fiction makes readers take an extra step in approaching the story and following it. Another important aspect of science fiction is its engagement with the “far away” in time and space; in this sense, Fred Botting talks about science fiction’s “unbounded explorations of change, outsiders, escape: its ‘freedom of imagery’ is freedom from realistic conventions”<sup>24</sup> and many other science fiction theoreticians talk about defamiliarisation of the familiar, “radical disjunction from the real”, and cognitive estrangement typical of science fiction. Another essential feature of science fiction is that it deals with science, but usually with science gone wrong,<sup>25</sup> or with presenting political, social and/or religious systems and communities that operate on significantly different principles from the ones we know. In explaining the workings of science fiction, David Seed invokes Arthur C. Clarke’s claim according to which science fiction can challenge conservative mind-sets through narratively embodied thought experiments.<sup>26</sup> This is an important point for our research here: the invocation of thought experiments to explain the nature of science fiction.

My intuition here is that, if we compare these two very broadly conceived genres of literature, we have to recognize different reading protocols (to use Seed’s term) readers have to employ in order to follow the story, and in the case of science fiction, this will include treating the work as presenting something non-existent, that is, existent only in the mind. On the other hand, this is not necessarily so for realism. In a similar vein, I think, in engaging with TEs, scientists and philosophers rely on the willingness to accept the description as given, even in cases where what is described is highly unlikely. We all know that “the violinist case” cannot happen, yet that doesn’t prevent us from engaging seriously with it. This can lead us to explore another reason for drawing the analogy: the motivation behind the construction of TEs.

## ***2.2. Sense 2: motivation for particular kind of description***

I think there is a valuable lesson to be drawn from our discussion so far: if we are to explain the analogy, we should consider the aim of constructing TEs, why is it that they are used in the first place? J. R. Brown lists several reasons why TEs are postulated, and these include: to fulfil specific function within a theory, to help illustrate and clarify abstract states of affairs, accelerate the process of understanding, as examples in conceptual analysis, to provide evidence for or against a theory.<sup>27</sup> I think this is the line we should pursue. Namely, a lesson from the previous chapter was that so many literary works due to their structure cannot function as thought experiments, primarily due to the fact that they do not ask readers to take that extra step. However, if we focus on the artistic aims that govern the construction of the story, we might get ahead. Davies himself mentions this criterion and it seems to me we should take it more seriously.<sup>28</sup> A common feature of novels like this is that they describe scenarios which bear close resemblance to our world, yet differ radically from it, in that they have never been actually enacted. These novels exaggerate what they describe, yet everything they describe is plausible, easy to imagine, not contrary to natural, psychological or social laws. In order to trigger a certain attitude or response, or just an intuition regarding something, authors who write this sort of novel describe scenarios that should present to readers a common yet disturbed picture of reality, so as to highlight the consequences that can be caused. This is very close to how philosophical TEs are used. I believe this model rightly captures another important aspect of works such as *Brave New World* and *1984*: these authors wanted to make

political statements and the right attitude (literary appreciation) should, as the end result, recognize that. To treat *Brave New World* as representation (or report) about some society that is real would be to terribly miss the point. On the other hand, it is only if we accept the “imagine that” that we can truly trigger our cognitive resources. I think this point is obvious if we analyse works like *The Possibility of an Island*, which consists of parts that are to be read as ‘real life episodes’ and parts that are based on thought-experimental scenarios. A competent reader, in order to fully appreciate the novel, has to have the ability to shift these perspectives as he goes along reading.

That, however, still does not exclude all the possible senses in which analogy might work. What we want to explain here is how it is that readers can deepen their understanding of the principles – moral, metaphysical and psychological – through engaging with literature. If we focus on this aspect, it appears that what matters is not only what is represented in the work and with what intention, but also what the reader brings in. Critics and defenders of cognitivism agree that in order to follow the story and understand what is going on, a reader already needs to have at least some elementary concepts and principles. The claim put forward by literary cognitivists is that in engaging with the work, these principles become more encompassing, and the reader becomes more sensitive towards different aspects of experience. That is what the analogy should explain.

### 2.3. *Sense 3: the analogy with TEs at the thematic level*

One possible objection to the analysis of sense 1 and 2 can come if we dissect particular literary works not into wider categories of its genre or in the amount of realistic details, but if we concentrate on different aspects, that is, the point, that the story presents, the concepts it develops, questions it raises, etc. Namely, someone might claim that where the story takes place (whether in 19th Chicago or in the distant future in space) bears no significance for what the story is actually about in terms of humanly important issues, and these are what matters. For example, although Dreiser’s *American Tragedy* can abound in details regarding the economic development of American society, first factories, etc, what we really need in order to follow the story is an understanding of psychological mechanisms, or principles that guide the actions of the main characters (for example, in order to explain Roberta’s trust in Clyde and her failure to see his lies and deceptions, we should invoke the notions of naivety, innocence, good upbringing and the way these principles

24

Botting 2005, 112.

25

See different accounts of this in Carroll 1990, and Seed 2005.

26

Seed, p. 5.

27

See Brown 2007. Carroll also provides a similar list of reasons for constructing TE, see his 2002.

28

Here is Davies: “... the narratives in TE’s differ from the narratives in standard works

of fiction not in failing to meet the requirements for fictionality, but in being presented to receivers with a different motivation. (...) For example, writers of utopias or dystopias such as *1984* or *Brave New World* plausibly intend that, as a result of the receiver’s making-believe the content of the narrative, she will come to believe that this is how certain societies would turn out, and will therefore amend her views about the merits of alternative political or socio-economic systems. Perhaps then, we should simply allow that some works of fiction are properly viewed as much more fully elaborated TE’s”. (Davies 2007b, 33)

collide in highly moral young girl who believes in love and doesn't suspect her lover to be unfair or immoral). While the love story between Clyde and Roberta is set against industrial development, what the author is really interested in the story is the question of agency, individual choices and free will, the reasons behind immoral behaviour and manipulation, the possibility of repentance, etc.

And this is, it might be claimed, where the analogy between the cognitive functioning of TEs and literature is evident: it is at this level of the story that the narration works upon our cognitive resources, enabling us to recognize the workings of general principles. If we focus on this aspect (so called thematic aspect) of literary fiction, then the differences between realism and science fiction become less important, because the setting of the story is not important. At this level, science fiction, as well as realistic literature, tackle the same "humanly important issues" which readers find relevant. For all its exaggeration, sublimity, gothic atmosphere and surreal setting, destruction of the world and technological innovations, science fiction deals with deeply important human issues and "asks fundamental questions about the world and the nature of selfhood"<sup>29</sup> that bear importantly on how we conceive of ourselves, other people and the world around us. To give but a few examples, one theme that is repeatedly being developed through the writings of H. G. Wells is that of genetic engineering, as well as the impact of Darwinism and evolution on people, which raises important questions about the identity and agency of human beings, as well as the boundary between species and the role of science and its impact on humans. Stephen R. L. Clark analyses the way science fiction treats religion and religious issues, principles of organized religion and faith, showing that what is ultimately brought into question is the possibility of a religion that is true. Particularly important in this respect are works pertaining to utopias and dystopias, which not only explore different political principles and social arrangements, but also question the nature of human beings, their willingness and abilities to connect with each other and form meaningful relations, not to mention their identity and self-awareness which can only find expression within society and wider social context. Philip E. Wagner shows the dependence of utopia on the particular political regime, that is, the historical context out of which it emerges:

"... through its presentation of this alternative community, the Utopian narrative has the effect of both highlighting in a negative light many of the problems of the reigning social order, and perhaps even more significantly, of showing that what is taken as natural and eternally fixed by the members of that society is in fact the product of historical development and thus open to change".<sup>30</sup>

Considerations along these lines show that, despite a high degree of breach from "reality" and "here and now", science fiction is deeply rooted with what is of the deepest human concern and can therefore certainly contribute in substantial ways to our cognitive endeavour of understanding the world.

On the other hand, it can also be claimed that the fact that realist literature is primarily concerned with "here and now" and in that sense informative of social, political, economic, etc. reality of what it describes is not an obstacle to its tackling universal themes which go beyond immediate reality presented in a work. To go back to Dreiser, no one can deny that his analysis of the class distinctions, poverty and richness and the influence these factors had on social interaction and personal relations among people is of no relevance for modern readers.

To go back to my question, does the analogy with TEs help us explain the cognitive values of literary fiction? If Thomson's violinist case is supposed to show that we should make a distinction between the concept of 'right to life' and the concept 'right to what is needed to sustain life',<sup>31</sup> and if it does so by mobilizing unarticulated cognitive resources grounded in our experience of the world, can the same account be applied to literary fiction? Can we claim that the very same process happens when, upon reading *The Scarlet Letter*, we are lead to consider in more details the moral implications of crime and punishment and to better understand different aspects of the psychology of punishment?

Several things need to be said here. First of all, most of those who defend literary cognitivism will try to ground it at the thematic level, that is, at the level where we are no longer concerned with propositions describing the fictional world and actions happening in it, but are concerned with principles regarding moral, metaphysical and psychological aspects of our experience. We said before that literature can help us cognitively in this respect by enabling us to see different shades of concepts we employ (such as the development of the concept of free will and acting from one's deliberation that is one of the central aspects of Dreiser's work<sup>32</sup>), or by its ability to give substance to bare principles (like calling into question the principles that determine male and female roles that some critics find central to Nathaniel Hawthorne's short stories<sup>33</sup>). It is also plausible to claim that some literary works, through the way they develop their themes, not only make different principles more available to the readers, but also call into question some established norms of thought and behaviour or present a new perspective for consideration. What needs to be recognized here is the plurality of values and different accounts are being put forward to explain how that is possible. We saw that Carroll claims that what actually happens in cases like this is the process of clarification, where a reader deepens his understanding of the (moral) concepts he has and learns how to apply them in new ways and in new situations. John Gibson claims that one of the effects of engaging with literature is its capacity to make us more aware of what we are supposed to do, and Berys Gaut develops a theory according to which we learn from literature by its capacity to employ and guide our imagination. To go back to the analogy with TEs, certainly it is at this level that it finds its strongest grounding. However, I think there is one more sense in which analogy might be grounded.

#### 2.4. *Sense 4: internal and external perspectives*

The notion of internal and external perspective as I will use it here comes from Peter Goldie, from his analysis of the way readers engage with the non-actual as well as actual events.<sup>34</sup> Although his focus is on engagement with narrative and the emotional responses it elicits in the audience – including being emotionally carried away as well as being emotionally resistant to go along with what is to be imagined – I think he makes a point that we might use

29  
Seed 2005, 3.

30  
Wegner 2005, 80.

31  
See Brown 2007.

32  
See Crane 2007 for interpretation along this line.

33  
See Person 2007.

34  
Goldie 2003.

in our analysis of the analogy between TEs and literary fictions. As Goldie sees it, narratives invite us to take these two different perspectives:

“The internal perspective is involved where the narrative represents or otherwise indicates the perspective, including the thoughts, feelings, and emotions, of one or more of the people who are internal to the narrative. Then, external to the narrative, there is the narrator’s own external perspective”.<sup>35</sup>

I think a similar approach can be taken by the reader of a literary work: we can read it from external perspective, (although obviously not in the authorial sense Goldie postulates), but as, let us say, observers of what is going on in the fictional world. On the internal perspective, a reader can also place herself within the fictional world, in place of one of the characters. This is the thought behind the idea that we can identify with literary characters and see what it feels like to be in their shoes. This is then certainly a powerful tool for gaining affective knowledge.

At one point, Davies also tackles the idea that we can shift perspectives towards TEs:

“Thompson’s TE does differ from standard narratives in works of literary fiction in one respect. The reader is invited to imagine certain things occurring to her, rather than to a fictional character or group of fictional characters. But this is merely a rhetorical device on Thompson’s part, and is in no way essential to the TE. We could replace references to ‘you’ with references to ‘a woman’, etc., without changing the way in which narrative functions.”<sup>36</sup>

So the question then becomes, does it make sense to claim that a reader can approach literary work from internal perspective and, given that she then enters a scenario that is, from her perspective like a TE, learn something new? Up to a certain point, this proposal seems plausible. Imagining what it would be like for me to be in this situation can be a source of knowledge. Barys Gaut offers a powerful account along these lines, in developing his theory of learning from the imagination. What he wants to show is that imagination should be recognized as a source of knowledge, and that literature is particularly powerful in governing the process of imagining.<sup>37</sup> But, on the other hand, it seems that approaching literary fiction in this manner would seriously undermine the overall value these works have, including cognitive values. One reason for that is that some literary fiction invites us to reflect on something, and in order to do that we do not have to think about what it would be like to be in that kind of situation. For example, in John Fowles’ novel *The Collector*, we are invited to think – among other things – about the value of art, creativity and artistic expression. In order to do that, I do not have to put myself in Miranda’s shoes and think about what it would be like to be held hostage. Readers who approached this work only to consider what it feels like to be held hostage in the manner described by Fowles would miss something terribly important that the work offers.

This strategy will also be futile with those works that offer a visual image of something. For example, Coetzee’s novels often depict post-apartheid circumstances in Africa. I do not think that readers necessarily have to put themselves in the perspective internal to the narrative in order to grasp the political regime and social consequences his works depict, in order to see them and reach conclusions about them.

So, we started with the assumption that literary fiction is cognitively valuable and that this value is in fact realized in several different cognitive benefits or payoffs. Then we agreed that each of them should have different justificatory grounds; in the case of factual knowledge and affective knowledge, as Davies

claims, we need to have a reliable author. In the case of general principles, insightful ways of classifying reality and categorization, fiction functions in the same way as TEs. My worry was to see whether all literary fiction can be seen as analogous to TEs and, after a brief analysis, I concluded that this analogy generally can only be extended to those works pertaining to science fiction, due to the nature of their content and the underlying motivation for that particular kind of description. However, if we focus on particular cognitive contributions – namely elaboration of general principles and categorical understanding – we can conclude that the analogy with TEs explains how fiction can be the source of cognitive values. For literary cognitivists such as Davies, this eliminated the hypotheses problem. However, I think there's more to be said about the hypotheses.

### 3. The value of hypotheses

Let us go back to the hypotheses problem. The claim put forward by those who deny literature cognitive value is that given that there is no evidence and no argument within the literary work, the best we can get out of the work are different hypotheses about general principles (moral, sociological, metaphysical) and potentially insightful ways of organizing experience. The problem then is how to test these hypotheses; David Novitz claimed that this process is done by applying the hypothesis to the world and, if proven correct, it can be accepted into one's system of belief and constitute knowledge. But, the argument goes, this cannot be right because it makes the whole process of testing external to the process of reading. Again, Carroll's claim, accepted by Davies, was that the analogy with TE's settles the matter in that this testing is done not by projecting the hypotheses to the external world, but by the fact that the hypotheses engage cognitive resources we already have and in virtue of reflective processes they generate in the reader get to be accepted or denied.

Let us pursue this a bit further, in order to try to avoid one more argument put against literary cognitivism. Namely, one possible worry that I have that is not settled by the analogy is the problem of different, perhaps even contradictory theses that the reader might end up negotiating, or even contradictory beliefs he might end up with after engaging with literature. Many literary anti-cognitivists insist upon the problem of literary works which advance contradictory claims. To give one rather general example, we can point to the way that slavery was presented in American literature. There is a tendency to talk in terms of pro-slavery vs. anti-slavery writers. While Harriet Beecher Stowe's *Uncle Tom's Cabin* is seen as arguing for recognition of moral invalidity of slavery, Caroline Lee Hentz's *Plantern's Northern Bride* aims to convince readers that slavery follows the just social order and should not be abolished.<sup>38</sup> Another interesting example of how literature puts forward different, even contradictory views and perspectives can be traced if we analyse the way the concept of free will is developed by American writers. While characters such as Huckleberry Finn embody the possibility of free will, Sister Carrie contradicts the very notion of independent agency. While in Mark Twain we still have the idea that it is possible to guide one's life according to one's own determination, choices,

35  
Ibid., 55.

36  
Davies 2010b, 390.

37  
See Gaut 2007.

38  
For interpretation along these lines see Crane 2007.

preferences and morality, in Theodore Dreiser and Stephen Crane characters are completely unable to control themselves and their choices and are utterly under the influence of the forces they cannot control or understand.<sup>39</sup> Examples like these show that the cognitive values of literature are not delivered in terms of propositions, in many cases hypotheses and questions we extract from works (“Are we free?” “Is there such a thing as moral obligation toward slaves?”) will not provide definite answers, but will certainly influence the way we think about our world, our experience and our place in the world, thus deepening our understanding of it. In that sense, even if different works advocate different perspectives, it is, I believe, up to the reader to consider them and evaluate them and see how they fit with one’s system of belief. For example, if one tries to make sense of the behaviour of Isabel in *The Portrait of a Lady*, one has to assume that in the end, she stays with her abusive and insensitive husband and rejects help because she believes she has to fulfil her duty as a wife. But in *A Doll’s House*, when trapped in unhappy marriage, Nora rejects the principle of duty and her obligations as a wife in order to pursue her own happiness. So what is a reader to conclude? For one thing, I think this question is wrong; whatever hypotheses reader picks up from the novel will not automatically generate new knowledge.

The cognitive value of these hypotheses is not in generating new knowledge or beliefs, but in – what I will roughly call – their instrumental function, that is, in the impact they have for the way we account for the experience – cognitive and emotional – with the world. Peter Kivy has offered an analysis of the cognitive values of literature in these terms. Drawing on the work of William James, he draws a distinction between live and dead hypotheses. The difference is roughly that live hypotheses are recognized by the audience as worthy of further consideration and thought, while dead are recognized as of no interest and probably untrue. Kivy advances what he calls a theory of literary plausibility, according to which:

“... one of the purposes of some literary works is to express propositions, frequently moral or, broadly speaking, philosophical ones, which present to us live hypotheses concerning matters of deep and abiding significance. When a literary work succeeds in doing this, it possesses thereby literary value, which I might as well call propositional value. This is by no means the only kind of literary value it possesses, and many literary works do not possess propositional value at all”.<sup>40</sup>

Although Kivy is mostly concerned with the banality problem, his analysis can easily be applied to the problem of the value of hypotheses. Here is again a quote from him:

“Where the banality is alleged to result from the lack of argument and analysis in the literary works, as it would in many novels, the reply is that argument and analysis occur in the gaps and afterlife, *in the reader’s mind*, as part and parcel of a legitimate literary experience”.<sup>41</sup>

Although Kivy does not go into any detail as to how this happens in the reader’s mind, obviously, with Kivy as well as with Davies, we have this idea that the process of testing is internal. Kivy, however extends this internality to include what he calls gaps and afterlife in the reading experience, during which “the images and content linger on in the mind to be savoured and thought about”.<sup>42</sup> I think this helps us settle the problem of diverse, even contradictory hypotheses.

To conclude, literary fiction is saturated with different cognitive benefits and its contribution to the way we see the world, make sense of it, and negotiate our relations with others is certainly cognitively valuable. This, however,

does not mean that this cognitive aspect should take precedence over overall aesthetic value. In this paper, I simply wanted to ground cognitive dimension of literature from epistemic point of view. That should not take away from – if anything, it should add to – the reasons we value literature aesthetically.<sup>43</sup>

## References

- Botting F. (2005), “‘Monsters of the Imagination’: Gothic, Science, Fiction”, in D. Seed (ed.), *A Companion to Science Fiction*, Blackwell, pp. 112–126.
- Brown J. R. (2007), “Thought Experiments in Science, Philosophy, and Mathematics“, *Croatian Journal of Philosophy*, 7, 19, pp. 3–26.
- Carroll N. (1990), *The Philosophy of Horror: Or, Paradoxes of the Heart*, Routledge.
- Carroll N. (2002), “The Wheel of Virtue: Art, Literature and Moral Knowledge“, *The Journal of Aesthetics and Art Criticism*, 60, 1, pp. 3–26.
- Carter R. and J. McRae (2001), *The Routledge History of Literature in English*, Routledge.
- Clark C. R. L. (2005), “Science Fiction and Religion“, in D. Seed (ed.), *A Companion to Science Fiction*, Blackwell, pp. 95–110.
- Crane G. (2007), *The Cambridge Introduction to Nineteenth-Century American Novel*, Cambridge University Press.
- Davies D. (2007a), *Aesthetics and Literature*, London: Continuum.
- Davies D. (2007b), “Thought Experiments and Fictional Narratives“, *Croatian Journal of Philosophy*, 7, 9, pp. 29–45.
- Davies D. (2010a), “Learning through Fictional Narratives in Art and Science“, in R. Frigg and M. Hunter (eds.), *Beyond Mimesis and Convention: Representation in Art and Science*, Boston Studies in the Philosophy of Science 262, Dordrecht: Springer, pp. 51–70.
- Davies D. (2010b), “Eluding Wilson’s ‘Elusive Narrators’“, *Philosophical Studies* 147.3, pp. 387–394.
- Dohrn D. (2009), “Counterfactual Narrative Explanation“, *The Journal of Aesthetics and Art Criticism*, 67, 1, pp. 37–47.
- Gaut B. (2006), “Art and Cognition“, in M. Kieran (ed.), *Contemporary Debates in Aesthetics and the Philosophy of Art*, Wiley–Blackwell, pp. 115–126.
- Gaut B. (2007), *Art, Emotion, and Ethics*, Oxford University Press.
- Gibson J. (2007), *Fiction and the Weave of Life*, Oxford University Press.
- Goldie P. (2003), “Narrative, Emotion, and Perspective“, in M. Kieran and D. Lopes (eds.), *Imagination, Philosophy and the Arts*, Routledge, pp. 55–69.
- Kieran M. (1996) “Art, Imagination, and the Cultivation of Morals“, *The Journal of Aesthetics and Art Criticism* 54:4, pp.337–351.
- Kieran M. (2005), *Revealing Art*, Routledge.

39

See Crane 2007 for interpretation along these lines.

40

Kivy 2006, 103.

41

Ibid., 112.

42

Ibid., 108.

43

I would like to thank David Davies, Snježana Prijić Samaržija, Elvio Baccharini and two anonymous referees for their valuable comments and suggestions regarding this paper.

Kivy P. (2006), *The Performance of Reading: an Essay in the Philosophy of Literature*, Oxford: Blackwell.

Novitz D. (1984), "Fiction and the Growth of Knowledge", in J. Margolis (ed.), *The Worlds of Art and the World*, Amsterdam: Rodopi.

Person L.S. (2007), *The Cambridge Introduction to Nathaniel Hawthorne*, Cambridge University Press.

Shippey T. (2005), "Hard Reading: The Challenges of Science Fiction", in D. Seed (ed.), *A Companion to Science Fiction*, Blackwell Publishing, pp. 11–26.

Swirski P. (2007), *Of Literature and Knowledge: Explorations in Narrative Thought Experiments, Evolution, and Game Theory*, Routledge.

Wegner P. (2005), "Utopia", in D. Seed (ed.), *A Companion to Science Fiction*, Blackwell, pp. 79–94.

Iris Vidmar

### Misaoni eksperimenti, hipoteze i kognitivna dimenzija književne fikcije

#### **Sažetak**

*Neki autori brane književni kognitivizam – stajalište da književna fikcija ima spoznajnu vrijednost – povlačeći analogiju između spoznajne vrijednosti misaonih eksperimenata i književne fikcije. U ovome ću radu analizirati razloge za ovu analogiju te vidjeti koliko se njen doseg može proširiti. U drugome dijelu razmotrit ću tvrdnju književnih antikognitivista prema kojoj književnost u najboljem slučaju može biti samo izvor hipoteza, a ne znanja. Ovu ću tvrdnju dovesti u pitanje pokazujući da hipoteze same mogu imati vrijedne spoznajne koristi, u nadi da se na taj način mogu obnoviti spoznajne koristi koje čitatelji pronalaze u književnosti.*

#### **Ključne riječi**

spoznajne vrijednosti, hipoteze, književni kognitivizam, realizam, znanstvena fantastika

Iris Vidmar

### Gedankenexperimente, Hypothesen und kognitive Dimension der literarischen Fiktion

#### **Zusammenfassung**

*Einige Verfasser verteidigen den literarischen Kognitivismus – einen Standpunkt, nach welchem die literarische Fiktion erkenntnismäßig ersprießlich ist – indem sie eine Parallele zwischen den Erkenntniswerten der Gedankenexperimente und der Literaturfiktion ziehen. In dieser Schrift fasse ich den Vorsatz, die Ursachen für diese Parallelisierung auszukundschaften wie auch zu ermitteln, inwieweit sich diese Analogie ausdehnen lässt. In dem zweiten Teil wende ich mich der vonseiten der literarischen Antikognitivisten vorgebrachten Behauptung zu, welcher zufolge die Literatur günstigstensfalls eine Hypothesen- und keineswegs eine Wissensquelle zu sein vermag. Ich hinterfrage diese These, indem ich die Anschauung erhärte, dass Hypothesen über eigene wertvolle erkenntnisbezogene Vorzüge verfügen können; all dies in der Hoffnung, kognitive Vorteile herzustellen, welche die Leserschaft aus der Literatur herausliest.*

#### **Schlüsselwörter**

kognitive Werte, Hypothesen, literarischer Kognitivismus, Realismus, Science-Fiction

Iris Vidmar

**Expériences mentales, hypothèses  
et dimension cognitive de la fiction littéraire**

**Résumé**

*Certains auteurs défendent le cognitivisme littéraire – position selon laquelle la fiction littéraire a une valeur cognitive – en établissant une analogie entre les valeurs cognitives de l'expérience mentale et la fiction littéraire. Dans cet article, j'analyserai les raisons de cette analogie et verrai jusqu'où cette analogie peut s'étendre. Dans la seconde partie, j'examinerai l'affirmation des anti-cognitivistes littéraires selon lesquels la littérature peut tout au plus être une source d'hypothèses et non pas du savoir. Je mets cette affirmation au défi en montrant que les hypothèses à elles seules peuvent avoir de précieux avantages cognitifs, dans l'espoir de restaurer ainsi les avantages cognitifs que les lecteurs tirent de la littérature.*

**Mots-clés**

valeurs cognitives, hypothèses, cognitivisme littéraire, réalisme, science-fiction

