

A Pragmatic Framework for Truth in Fiction*

Andrea Bonomi Sandro Zucchi

Dipartimento di Filosofia
Università degli Studi di Milano

Abstract

According to R. Stalnaker, context plays a role in determining the proposition expressed by a sentence by providing the domain of possible worlds that propositions distinguish between: a sentence expresses a proposition by selecting a subset of the set of possible situations given by the context. This is also true for embedded sentences, but these sentences express propositions by selecting subsets out of contexts derived from the basic one. In this paper we propose a semantic analysis of sentences of the form “In fiction x , p ” based on this picture of context. We argue that the derived contexts for sentences in the scope of “In fiction x ” are determined by three factors: what the beliefs of the author are taken to be, the conventions established for the fiction, and a defeasible presumption of reliability of the narrator. We develop a formal implementation based on the notion of a system of spheres centered on a set of worlds.

1 Fiction and communication

When we communicate with others, we do so against a background of common beliefs, a body of information shared by the participants in the communicative exchange and recognized by them to be shared. The background of shared information may help us to determine the content of what is said and is updated in the course of the communication process as a result of the speech acts performed by the participants in the process. This picture of how communication works has been

*We thank Horacio Arló-Costa and Paolo Casalegno for comments on a previous version of the paper. We also thank the participants in the workshop *Do Ficta Follow Fiction?* for discussion. Finally, we thank an anonymous referee whose comments have prompted a substantial revision of our proposal in the last part of the paper.

developed by R. Stalnaker in a number of papers.¹ While the type of discourse most often discussed in these papers is the spoken discourse, the picture is applicable to written texts as well. For example, when one reads an article in a newspaper, the content of the article is determined on the basis of the information that is taken to be shared by the reader and the writer, and, if what the article says is accepted, the common background is updated by adding the content of the article to it. The same may happen when the information is conveyed by other media, for example by a documentary: the information provided by the documentary is assessed on the basis of the information shared by the author and the audience and, if accepted, is added to the common background.

Works of fiction, be they oral, written, or visual, are not meant to update the common background in the same way. If someone reads *A Study in Scarlet* and understands that it's a work of fiction, the outcome of the process, if the communicative exchange is successful, is not the updating of the common background with the information that there is a detective called Sherlock Holmes who lives in London at 221B Baker St. Yet, the picture of communication as occurring against a background of shared information is relevant to account for our understanding of works of fiction as well. When we read, hear or watch a work of fiction, the body of information that is presumed to be shared by the audience and the narrator matters for the understanding of the work. Grasping the content of a work of fiction involves grasping both what is explicitly said in the story and what is implicit in it, and what's implicit in a fiction may depend, among other things, on information that is taken to be shared by the narrator and the audience to which the story is told. We know that the earth is round in the stories of Sherlock Holmes, because, although this is never mentioned in the stories, we take the assumption that the earth is round to be part of the information that is supposed to be available to the narrator and the audience. This information together with the information explicitly presented in the fiction determines the content of the fiction. It is the fiction's content in this broad sense that is relevant to evaluate p in sentences of form (1):

(1) In fiction x , p .

The purpose of this paper is to develop an account of the semantics of (1) based on this intuitive picture. In order to pursue this goal, we have to be more explicit on how the interpretation of sentences is determined by context, we have to say more on how the shared background in the context of utterance of a sentence may contribute to determine what the sentence expresses. We take up this task in the

¹See Stalnaker (1999). There are some problems with identifying the shared background with common beliefs. In our discussion, we'll ignore these problems. See Stalnaker (2002) for some proposed refinements.

next section.

2 Basic contexts and derived contexts

Stalnaker has defended the view that contexts should be represented as sets of possible worlds or *context sets*. A context set is the set of possible worlds compatible with the beliefs shared by the participants in the communicative exchange and recognized by them to be shared. The context plays a role in determining the proposition expressed by a sentence by providing the domain of possible worlds that propositions distinguish between, namely propositions are built out of the set of possible worlds provided by the context set.² To illustrate how this works, consider (2):

(2) John ran

A context may contain information about which individual called “John” is most salient and which past interval is relevant. For example, context C_1 may be a set of worlds in which a is the most salient individual called “John” and t is the relevant past interval, C_2 a set of worlds in which b is the most salient individual called “John” and t' the relevant past interval. In these contexts, (2) will express different propositions: in C_1 it will express the proposition that a ran during t (the set of worlds in C_1 in which a runs during t), in C_2 the proposition that b ran during t' (the set of worlds in C_2 in which b runs during t').

One way of implementing this picture formally is to assume that sentence meanings are functions from contexts (in Stalnaker’s sense, sets of worlds) to propositions built out of these sets. For example, if we ignore tense, the meaning of (2) may be represented by (2)', where “ C ” is a variable ranging over context sets (characteristic functions of sets of worlds) and w a variable ranging over worlds:

(2)' $\lambda C \lambda w [C(w) \wedge \text{run}(j)(w)]$

According to (2)', the meaning of (2) is a function that assigns to each context set the set of worlds in this set in which John runs. The notion of truth in a context C may be spelled out in this way:

(T) A sentence φ is true in a context C if and only if the proposition expressed by φ in C is true in the real world (more formally: φ is true in C (relative to an assignment g) iff $\llbracket \varphi \rrbracket_g (C)(w_r) = 1$).

²Stalnaker (1988).

Some comments are in order concerning this definition of truth in a context. According to (T), if the context C representing the set of assumptions shared by the conversational participants does not contain the real world, no sentence can be true in C . This means that, if we share some mistaken assumptions about the world, nothing we say can be true in that context. This is an unwelcome result. For one thing, given that we often share mistaken beliefs about the world, it would have the effect of making most of what we say false. Intuitively, if someone is happy now, it seems reasonable to claim that he says something true by asserting “I’m happy now”, whether or not when he is making this assertion he shares some false belief with his audience concerning which team won the soccer world cup in 2002.³ Let’s say that a context is true if the real world is a member of it, and false otherwise. Notice that we cannot avoid the problem posed by (T) by saying: φ is true in a true context if the proposition φ expresses in it is true in the real world and φ is true in a false context if there is a true context in which φ is true. Suppose we utter “he is happy now” in a context in which we are talking of John. This sentence is true in this context iff John is happy now. Suppose, however, that the set of shared beliefs that make up the context contains the false assumption that the earth is flat. By claiming that we said something true in this context, we do not mean that there is a true context in which the reference of “he” is happy at the time when the sentence is uttered. A true context may contain no shared assumption to the effect that we are talking of John when we utter “he is happy”. Worse, it may contain only propositions about Bill, in which case “he is happy now” would probably be understood as a claim about Bill. Clearly, the existence of such a true context in which “he is happy now” is true is not sufficient to regard the sentence as true in the original false context. We might require that, if the context is false, φ is true in it iff φ is true in the context $\{w_r\}$. But this raises other problems. The context set $\{w_r\}$ corresponds to the context in which we share complete knowledge of the world. By requiring that a sentence is true in a false context iff it is true in $\{w_r\}$, we predict that exactly the same sentences are true in every false context. Again, this is an unwelcome result. Take a false context in which the participants believe that the earth is flat while they are aware that John has recently been to Afghanistan and a false context in which the participants believe that the earth is flat and also mistakenly believe that Bill has been to Afghanistan. The sentence “He was in danger recently” may express the true proposition that John was in danger recently in the former context but not in the latter, as the referent of the pronoun “he” is likely to be John in the first case and Bill in the second. A more plausible definition of truth in the case of a false context is one that minimizes the differences between the false context and the true one in which we check for truth: if the context C is

³See Heim (1982) for a discussion of this point.

false, φ is true in C iff φ is true in a context as close as possible to C that includes the real world as a member. Here, we will not pursue this issue further and we'll leave the task of spelling out this proposal formally for another occasion.

In the analysis sketched above, a sentence expresses a proposition by selecting a subset of the set of possible situations given by the context. This is true also for embedded sentences, but these sentences express propositions by selecting subsets out of contexts *derived* from the basic one, namely from the set of worlds representing the shared background. In some constructions, the derived context is a subset of the basic context set,⁴ in other constructions this need not be the case. Sentences ascribing beliefs provide an instance of the latter type of construction. Belief report (3) may be translated as (3)' (where the denotation of $D_{mb,w}$ is specified as in (4)):

- (3) Mary believes that it's raining
 (3)' $\lambda C \lambda w [C(w) \wedge \forall w' [D_{mb,w}(w') \rightarrow \text{rain}(w')]]$
 (4) $D_{mb,w}$ abbreviates $\lambda w' [w'$ is compatible with Mary's beliefs in $w]$

According to (3)', the proposition expressed by (3) in a context C is the set of worlds meeting this condition: a world w is in this set iff w is in C and the worlds compatible with Mary's beliefs in w are worlds in which it's raining. For every world w in the basic context set C , the set denoted by $D_{mb,w}$ (where w is assigned to w) is the derived context set relative to w for the complement of "believe" in (3). This derived context is not included in the basic context if at w Mary holds beliefs that are not compatible with w . The translation schema yielding (3)' is (5) (where φ translates p):

- (5) x believes that $p \Rightarrow \lambda C \lambda w [C(w) \wedge \forall w' (D_{xb,w}(w') \rightarrow \varphi(D_{xb,w})(w'))]$
 (6) $D_{xb,w}$ abbreviates $\lambda w' [w'$ is compatible with x 's beliefs in $w]$

We can specify the derived context for sentences in the scope of the modifier "In fiction x " in a similar fashion. Before we do this, however, we need to dispel a confusion that might arise from the use of certain phrases in ordinary talk about fiction. Sometimes, we use expressions like "the world of the novel," "the world of *David Copperfield*," and so on. This might generate the impression that each novel, each work of fiction, identifies exactly one possible world in our sense. But possible worlds, in the sense of possible world semantics, leave no room for indeterminacy. At any given time, an individual existing at a world will have a certain height, a certain weight, a certain eye color, a certain spatial position, and so on. So, if a novel describes exactly one world in this sense, it must contain

⁴This is the case, according to Stalnaker, for derived contexts of consequents of indicative conditionals.

information about height, weight, eye color, spatial position of its characters at any given time. This is clearly false. We don't exactly know how tall is Sherlock Holmes, we don't know his exact weight, we don't know his spatial coordinates at different times. Thus, we must conclude that, from the point of view of possible world semantics, a novel does not describe exactly one possible world, but a set of possible worlds. In all the worlds in the set described by Doyle's stories, Holmes will have all the properties explicitly or implicitly attributed to him in the stories. As the stories are silent on his exact height and weight, he will have different heights and weights in these worlds. Now that we have made this clear, let's turn to the task of specifying the translation rule for (1):

(1) In fiction x , p .

The derived context for p in (1) is given by instruction (7):

(7) The derived context for p relative to a world w is the set of worlds described by fiction x in w .

Sentence (8) is translated as (8)':

(8) In *A Study in Scarlet*, Holmes meets Watson in a lab.

(8)' $\lambda C \lambda w [C(w) \wedge \forall w' (D_{\text{Scarlet}, w}(w') \rightarrow \text{meet-in-a-lab}(wts)(h)(w'))]$

(9) $D_{\text{Scarlet}, w}$ abbreviates $\lambda w' [w' \text{ is described by } A \text{ Study in Scarlet in } w]$

According to (8)', the proposition expressed by (8) in a context C is the set of worlds meeting this condition: a world w is in this set iff w is in C and the worlds described by *A Study in Scarlet* in w are worlds where Holmes meets Watson in a lab. For every world w in the basic context set C , the set denoted by $D_{\text{Scarlet}, w}$ (where w is assigned to w) is the derived context set relative to w for "Holmes meets Watson in a lab" in (8). This means that (8) is true in a context C iff the worlds described by *A Study in Scarlet* in the real world are all worlds in which Holmes meets Watson in a lab. The translation schema yielding (8)' is (10) (where φ translates p):⁵

⁵Translation schema (8)' does not allow for truth-value gaps. If Holmes is not 6.2 ft. tall in some worlds of Doyle's stories, (8)' predicts that (i) is false.

(i) In Doyle's stories Holmes is 6.2 ft. tall.

Lewis (1978) suggests that it may be preferable to make (i) neither true nor false rather than simply false. After all, even if the stories do not tell us that Holmes is 6.2 ft. tall, they do not exclude this either. If Lewis is right, translation schema (8)' must be modified by requiring that "In fiction x , p " be true if p is true in all the worlds described by the fiction, false if p is false in all such worlds, and

- (10) In fiction x , $p \Rightarrow \lambda C \lambda w [C(w) \wedge \forall w' (D_{fx,w}(w') \rightarrow \varphi(D_{fx,w})(w'))]$
 (11) $D_{fx,w}$ abbreviates $\lambda w' [w' \text{ is described by fiction } x \text{ in } w]$

As it stands, translation schema (10) provides truth-conditions for sentences of form (1) only insofar as we already know under which conditions, given a work of fiction x , a world w is in the set of worlds described by x . Unless we say more on how the set of worlds described by a fiction is determined, the theory has little predictive power. If I assert (8) or (12), I have said something true. On the other hand, if I assert (13) or (14), I have said something false.

- (12) In *A Study in Scarlet*, the earth is round.
 (13) In *A Study in Scarlet*, Holmes solves the Boscombe Valley mystery.
 (14) In *A Study in Scarlet*, the earth is flat.

In our terms, this means that the worlds described by *A Study in Scarlet* are worlds in which Holmes meets Watson in a lab and the earth is round, but not worlds in which Holmes solves the Boscombe Valley mystery or worlds in which the earth is flat. Why are the worlds described by *A Study in Scarlet* made up in this way? To answer this question, we need to be more explicit on how the derived set denoted by $D_{\text{Scarlet},w}$ in (8)' is determined. We take up this issue in the next section.

3 Worlds of fiction

A work of fiction describes a set of worlds. Which worlds are they? Are there general principles at work that determine this set? Can we provide a general characterization for it of the form: given a fiction x , the set of worlds x describes is the set that meets the following conditions . . . ?⁶ For the purpose of describing the logical form of sentences of the form “In fiction x , p ,” all is needed is to suppose that some such set exists. In the next sections, nonetheless, we’ll discuss some problems raised by the attempt to characterize the worlds of a fiction and we’ll propose some strategies to deal with them.

Implicit content At first blush, one might claim that the worlds described by a fiction are those in which the content that is explicitly conveyed by the fiction is true: if the fiction is a written text, the worlds in which the text is true, if it’s a movie, the worlds in which the events represented in the movie take place, if it’s

neither true nor false otherwise. For sake of simplicity, we’ll hold on to the formulation in (8)'.

⁶This issue has been discussed by several authors, among them Lewis (1978, 1983), Ross (1997), Currie (1990), Walton (1990).

a play, the worlds in which the events represented in the play take place. One immediate problem with this first characterization is that it doesn't account for truths that are implicit in a fiction. Consider the following sentences:

- (8) In *A Study in Scarlet*, Holmes meets Watson in a lab.
- (12) In *A Study in Scarlet*, the earth is round.
- (15) In *A Study in Scarlet*, Watson has blood in his veins.

All these sentences are true, yet only (8) is true in virtue of what the text of *A Study in Scarlet* explicitly says. The truth of (12), as well as the truth of (15), is only implicit in the novel. To account for the truths in (12) and (15), the worlds described by the novel must then be worlds in which Watson has blood in his veins and the earth is round. How can this result be obtained?

A way to make sure that the worlds of the novel meet this condition is to assume that the worlds described by a fiction are those closest to the real world in which the content that is explicitly conveyed by the fiction is true. Since in the real world men have blood in their veins and the earth is round, this condition requires that (12) and (15) be true. Indeed, for any world w where the explicit content of the fiction is true, Watson does not have blood in his veins, and the earth is not round, there is a world where the explicit content of the fiction is true which is closer than w to the real world: a world like w except that Watson has blood in his veins, the earth is round and the consequences of these facts hold in this world. Yet, as several authors have pointed out, closeness to the real world, as way to yield a fiction's implicit truths, leads to some counterintuitive results. Here's a case described in Walton (1990) which makes the point:⁷

A storyteller, in a culture in which it is universally and firmly agreed that the earth is flat and that to venture too far out to sea is to risk falling off, invents a yarn about bold mariners who do sail far out to sea. No mention is made in the story of the shape of the earth or of the danger. (p. 150)

Walton observes that, if truth in fiction depends on what is true in those worlds closest to the real one in which the explicit content of the fiction is true, we are led to conclude that (16) is true:

- (16) In the yarn, the earth is round and there is no danger of falling off the edge.

This is not a desired result: "to insist on this. . . would ruin a good adventure story; the tale . . . depends on the danger for its dramatic effect."

⁷See Lewis (1978) for other problematic cases.

Lewis's suggestion, for those who find this consequence unacceptable, is that implicit truths in fiction depend on closeness not to the real world, but to the set of worlds that represents the overt shared beliefs of the community where the fiction originates.⁸ Lewis calls these worlds the *collective belief worlds* of the community. If we accept this suggestion, we may characterize the worlds described by a fiction as those worlds in which the explicit content of the fiction is true that are closest to the collective belief worlds of the community of origin of the fiction. Facts like the earth's being round, that are unknown in the community where the yarn about the bold mariners is produced, are thus irrelevant to establish what is true in the yarn. In *A Study in Scarlet*, on the other hand, it is true that the earth is round and that Watson has blood in his veins, although this is not explicitly said in the novel, precisely because it is a collective belief of the community in which the novel was written that the earth is round and human beings have blood in their veins.

However, the view that, in evaluating implicit fictional truths, we should stay close to the overt beliefs of the community in which the fiction originates has some problematic consequences of its own. Suppose an author writes a novel in Nazi Germany. A side character in the novel is a Jewish composer. Not much is said about him and the novel does not express racist views about Jews or about anyone else. Shall we conclude that (17) is true?

(17) In the novel, the composer belongs to an inferior race.

The answer is no. If (17) were true, the novel would be a racist work, which is not. Yet, the truth of (17) is what Lewis seems to predict here, since arguably racist views about Jews were overt shared belief in Nazi Germany. One might protest that in Nazi Germany racist views about Jews were no quite part of overt collective beliefs. It is not true, one might say, that more or less everyone in the community held those racist view. Suppose we change the example a bit. Suppose it had been true that in Nazi Germany more or less everyone held racist views about Jews. Would (17) have been true in this case? The answer is still no: the novel does not become racist because racist beliefs become more widespread. Writing becomes a dangerous business if what counts for establishing implicit fictional truths is closeness to the overt collective beliefs of the community in which the fiction originates. You write a novel, do not address the issue of racism in it, but if you happen to belong to a racist community, unsavory propositions come true in your novel. If you think implicit fictional truths depend on closeness to overt collective beliefs in the community of origin, maybe you are willing to bite the bullet and accept this consequence. Maybe some authors would be inclined to sue.

⁸These are the beliefs that more or less everyone holds in that community, that more or less everyone believes that more or less everyone else holds in that community, and so on.

Walton (1990, 1994) claims that, when it comes to works of fiction, we are less willing to let the worlds of the fictional works deviate from the real world in moral respects.⁹ If Walton is right, the fact that (17) is intuitively false may depend on the particular nature of the proposition that is asserted to be fictional in (17), namely on the fact that this proposition expresses a prejudice we find morally objectionable. Yet, a similar problem arises also in the case of fictional propositions that do not involve objectionable moral judgement. A case described by P. Casalegno makes the point. Suppose that K writes a novel in a society in which almost everyone believes in god. Although the novel does not question the existence of god, it is known to almost everyone that K does not believe in god, and K knows that almost everyone knows that he is an atheist. Intuitively, (18) below is false. Yet, if the worlds relevant for evaluating (18) are those compatible with the novel's explicit content that are closest to the collective belief worlds of the society in which the novel is written, we predict that (18) is true, as the belief in god is an overt shared belief in this society.¹⁰

(18) In K's novel, god exists.

Where do things go wrong here? Intuitively, (18) is false because K does not intend that his readers take for granted that god exists in the novel, and his readers, knowing that he doesn't believe in god, do not assume that he intends that they take for granted that god exists in the novel. Regarding as relevant the collective belief worlds of the society in which the novel is written means ignoring the author's beliefs, at least when these beliefs, although publicly known, do not coincide with the overt shared beliefs of the society in which the author is writing. A possible way out that lets the author's beliefs come into play is this (we owe the suggestion to D. Marconi). Let's say that *p* is an overt belief by an author in a community iff the author believes that *p* and, moreover, it is a widespread belief in the community that the author believes that *p*. Let's assume that the implicit truths in a work of fiction depend on the beliefs of the author that are overt in the community where the fiction originates.¹¹ According to this proposal, implicit truths in fiction depend

⁹On the difficulties we encounter in imagining fictional worlds that we take to be morally deviant, see also Gendler (2000).

¹⁰For those who still find this example morally laden, suppose there is a society in which almost everyone believes in the existence of trolls, and it is known to everyone in that society that a certain novelist belonging to that society does not believe in trolls. We take it that there is no moral issue at stake for us concerning the existence of trolls. Yet, we would be inclined to reject the claim that it is implicitly true in this novelist's work that trolls exist.

¹¹In characterizing the relevant worlds, we need to refer to the author's overt beliefs, rather than simply to the author's beliefs, to avoid that an author's private and idiosyncratic beliefs may determine what is true in a fiction. As Lewis points out, if Doyle held a secret belief in the existence of purple gnomes, this would not be sufficient to make it true that purple gnomes exist in the stories of

on closeness to the set of worlds that represent the overt beliefs of the author in the community of origin. Sentences (12) and (15) are true since, presumably, it is an overt belief of Doyle's in Victorian society that human beings have blood in their veins and that the earth is round. Sentence (16) is false, since that the earth is flat is an overt belief of the author in the community where he invents his yarn. Finally, (17) and (18) are false, since the propositions in the scope of "In the novel" are not overt beliefs of the author in his community.

A possible alternative is to assume that implicit truths in a work of fiction depend on closeness to the set of worlds that represent the overt beliefs of the author relative to our community. This view would make the same predictions concerning (12), (15), (16), (17), and (18), as long as we agree with the author's contemporaries about the author's beliefs. However, if we believed that Doyle was convinced that the earth is flat, we could no longer claim that (12) is true. This view has the consequence that the implicit content of a work of fiction is not stable, it may change as the beliefs of the community about the author's beliefs change. According to some theorists, the interpretation of a text is a function of the interpretive model adopted by the reader.¹² If you are inclined to accept such a relativistic view of a text's content, you are probably inclined to accept the consequence that implicit content is unstable. We won't try to reason you out of it.

The role of conventions If our diagnosis is correct, reference to the author's overt beliefs is needed for the purpose of characterizing the implicit content of a fiction. But this is still not sufficient to capture what truths are implicit in a fiction. Take the case of the dragon Sculch described by Lewis:

Suppose I write a story about the dragon Sculch, a beautiful princess, a bold knight, and what not. It is a perfectly typical instance of its stylized genre, except that I never say that Sculch breathes fire. Does he nevertheless breathe fire in my story? Perhaps so, because dragons in that sort of story breathe fire. But the explicit content does not make him breathe fire. Neither does background, since in actuality and according to our beliefs there are no animals that breathe fire. (Lewis 1978, p. 274)

Lewis's conclusion here is that the fact that Sculch breathes fire in the story follows from what is true of dragons in other stories. How should we provide for this possibility in the truth-conditions for sentences of the form "In fiction x, p"? The

Sherlock Holmes.

¹²See Fish (1980), for example.

answer we suggest is that the other stories about dragons have established a convention: in fictions about knights, princesses and monstrous reptile-like creatures with wings, these monstrous creatures normally breath fire.¹³ In evaluating what is true in a fiction, we should only take into account those worlds maximally close to the author's overt belief worlds in which the explicit content of the fiction is realized *in accord with the conventions of the genre to which the fiction belongs*.¹⁴

Appeal to conventions to which fictions belong not only has the effect of yielding some implicit truths that could not be derived otherwise. It also allows us to rule out some undesired conclusions about what is implicitly true in a fiction. As Walton points out, Othello utters verses that would qualify anyone as a first rank poet. Does he, thus, have an unbelievable natural literary flair in the play? No. Does Violetta behave oddly when she sings in *La Traviata* as she is about to die? No. Yet, by staying close to the author's overt belief worlds we should probably be committed to answer yes in both cases. Verdi would think it odd to sing when one is about to die, and we believe that he would. Presumably, Shakespeare reckoned that Othello's verses were pretty good poetry, and the Elizabethans probably thought he did.¹⁵ The undesired conclusions are avoided once we only take into account those worlds closest to the author's overt belief worlds in which the explicit content of *La Traviata* and *Othello* is realized in accord with the conventions of operas and Elizabethan plays. By the conventions of operas, people who die singing need not be odd, by the conventions of Elizabethan plays a man who expresses himself like Othello need not be a first rank poet.

Explicit content The worlds a fiction describes, according to what we are suggesting, are those worlds closest to some set of overt beliefs in which the explicit content of the fiction is realized compatibly with certain conventions. This characterization takes as unproblematic the notion *world in which the explicit content of the fiction is realized*. Above, we said, as a first approximation, that, if the fiction is a novel, the worlds in which its explicit content is realized are those in which the text of the novel is true, if it's a movie, the worlds in which the events represented in the movie occur, and so on. This is a simplification. Let's see why.

¹³We need to say 'normally', since a fiction about knights, princesses, and monstrous reptile-like creatures may also explicitly say that the reptilian monster does not breath fire.

¹⁴The conventions ruling a fiction are not necessarily established interfictionally, they don't have to depend on membership in a previously established genre. In some cases, the convention may be established by the text of a fiction even though there is no precedent for it. The fact, observed by Lewis, that we take the characters of the *Threepenny Opera* to be treacherous on the ground that everyone put to test proves treacherous would be a convention established intrafictionally.

¹⁵As Walton points out, the same problem arises for Lewis's suggestion that the worlds we should stay close to are the collective belief worlds of the community where the fiction originates.

Take the following passage from *The Adventures of Huckleberry Finn*:

... This is the speech – I learned it, easy enough, while he was learning it to the king:

To be, or not to be; that is the bare bodkin
That makes calamity of so long life. . .

The novel is told in first person by Huck, and here Huck is telling us that Hamlet's famous soliloquy is different from the one we know. As Byrne (1993) points out, if the worlds the novel describes are those that make the text of the novel true, we should conclude that in *The Adventures of Huckleberry Finn* Hamlet's soliloquy is different from the way it is in reality. But this is clearly an incorrect prediction: what we are supposed to gather from the above passage is that Huck got Hamlet's speech wrong. Unreliable narrators then pose a problem for the simple minded view that the explicit content of a novel corresponds to those worlds in which the text of the novel is true.¹⁶

Lewis (1978) proposes the following truth-conditions for sentences of the form "In fiction x, p":

"In fiction x, p" is true iff p is true in all the worlds in which the fiction actually denoted by x is told as known fact that are closest to the collective belief worlds of the community in which the fiction originates.

In order for a story to be told as known fact in a world, what is narrated in the story must be true in that world, since, as Lewis points out, if what is narrated were not true it could neither be known nor truly told of. If what is narrated in *The Adventures of Huckleberry Finn* is true in the worlds of the novel, however, we face the problem of the unreliable narrator.¹⁷

According to Currie (1990), the truth-conditions for sentences of the form "In fiction x, p" should be stated thus:

¹⁶The problem posed by unreliable narrators is also pointed out in Currie (1990) and in Lewis (1983).

¹⁷Notice that the problem here cannot be dodged by proposing, as Lewis (1978) does in footnote 7, that the worlds realizing the story are those in which the story is told for whatever purpose the narrator intends: if the story is told for the purpose of telling a known fact, the worlds in which the story is told as known fact, if it is told for the purpose of telling a lie, the worlds in which the story is told as a lie, and, if it is told for the purpose of telling a mixture of facts and lies, the worlds in which the story is told as a mixture of facts and lies. In *The Adventures of Huckleberry Finn*, Huck, the narrator, is telling the story for the purpose of narrating a known fact, thus veridicality should apply in this case even if we adopt Lewis's proposal.

“In fiction x, p” is true iff it is reasonable for the informed reader to infer that the fictional author of the fiction denoted by x believes the proposition denoted by p (where the informed reader is a reader who knows the relevant facts about the community in which the fiction was written).¹⁸

The fictional author, according to Currie, tells the fiction as known fact. However, Currie avoids the problem posed by unreliable narrators by suggesting that the fictional author should not be identified with the explicit narrator, the character who tells the story. The beliefs of the fictional author in *The Adventures of Huckleberry Finn* would be different from Huck’s: while Huck believes that Hamlet says “that is the bare bodkin. . .,” the fictional author doesn’t. And it’s the fictional author’s beliefs that count for establishing what is true in the novel.¹⁹

The presence of a fictional author distinguished from the explicit narrator must be accepted in some cases. As Currie observes, in Kurosawa’s film *Rashomon*, there are four explicit narrators telling incompatible stories each from his own point of view. In this case, it seems plausible to postulate an unobtrusive narrator that tells us how these characters in the story report what happened. Yet, in the case of *The Adventures of Huckleberry Finn*, Huck is telling the whole story and, apart from the need to deal with his unreliability, there is no independent reason to posit a fictional author in the wings who is telling the story. If the fictional author is so unobtrusive in the text of *The Adventures of Huckleberry Finn*, how can we determine what is true in the novel on the basis of what his beliefs are?

Suppose we stick to the *prima facie* hypothesis that in *The Adventures of Huckleberry Finn* Huck is the teller, no hidden narrator. The case of the unreliable narrator then shows that, in stating the truth-conditions for sentences of the form “In fiction x, p”, we cannot require that the narrator’s story be true in the worlds relevant to establish what is true in the fiction. Thus, we have to give up the condition that the worlds of a fiction are worlds in which the fiction is told as known fact. We may go for a weaker condition: the relevant worlds must be worlds in which the narrator believes that the story he is telling is true. This formulation, by making it explicit that the relevant worlds are not necessarily worlds in which the content of the fiction is true, no longer predicts that (19) is true, but has the undesired effect of casting doubt on the truth of (20) as well:

- (19) In *The Adventures of Huckleberry Finn*, Hamlet’s soliloquy in Shakespeare’s play says “To be, or not to be; that is the bare bodkin. . .”.
- (20) In *The Adventures of Huckleberry Finn*, Huck and Jim find a raft.

¹⁸Here, we ignore Currie’s reference to interpretations.

¹⁹Currie’s fictional author is similar to what Eco (1994) calls “model author.”

If, among the worlds in which the narrator believes the story he is telling, we consider those that are closest to the worlds of the author's overt beliefs in the community where the fiction originates and compatible with the conventions of the genre, there is no reason to assume that in these worlds anything that is told in *The Adventures of Huckleberry Finn* takes place. Presumably, among Twain's overt beliefs is the belief that the story told in the novel is made up. Thus, among the worlds where Hucks tells his story and believes it to be true, those that are closest to Twain's overt belief worlds are worlds in which Huck, without realizing it, makes up the whole thing. Thus, (20) is false. To get the desired result that (20) is true and (19) is false, we need to do better than simply trading the requirement that what the narrator tells is true in the worlds of the fiction with the requirement that in these worlds the narrator believes what he says.

A fiction, by its nature, invites us to suppose that the world is different from the way we believe it is. Adopting a reasoning that leads us to conclude that no fictional truths are generated by works of fiction except that there is a narrator that invents an unlikely story amounts to defeating the purpose of fiction. For this reason, when we engage in the kind of supposition required by fiction, we follow this strategy: we do not question the narrator's claims on the general ground that, in view of the overt beliefs of the author, his tale is likely to be made up. In view of the overt beliefs of the author, the tale of any fictional narrator is likely to be made up. To question the narrator's claims in a fiction more specific reasons are required than reasons that apply to any fiction whatsoever, we need reasons related to the particular content of the fiction at hand, specific clues contained in the fiction that tell us which claims should be called into question. Let's call reasons of this sort *intrinsic to the fiction*. The fact that a fiction, however realistic, describes worlds that are different from those compatible with the overt beliefs of the author is not a reason intrinsic to the fiction to question the truth of the narrator's claims in the fiction. The fact that a character in a fiction is an uneducated boy may be a reason intrinsic to that fiction to question his claims about the contents of Shakespeare's plays in the fiction.

In view of these observations, we suggest that the following presumption of reliability is at work when we evaluate the claims of the narrator:

Presumption of reliability. Let w be a world in which fiction x is narrated and the narrator believes that x is true. If the narrator tells that q , q is true in w unless there are reasons intrinsic to the fiction that indicate otherwise.

In *The Adventures of Huckleberry Finn*, Huck only goes to school for three or four months, and can barely spell, read or write. So, there are reasons intrinsic to the story to question his claim about the text of Shakespeare's plays. But the story

gives us no reason to doubt that Huck and Jim went down the Mississippi river on a raft. As the fiction gives no indication that Huck is unreliable in this respect, we believe him on this point.

The intuitive proposal We are now in a position to summarize our proposal. We argued that the worlds relevant to establish what is true in a fiction are worlds that are as much as the narrator says as is required by the presumption of reliability: what the narrator says is true in these worlds unless there are reasons intrinsic to the fiction that indicate otherwise. We also argued that, in order to account for fictional truths, we need to pay attention to the conventions that are at work for fictions. Let's say that a world w is compatible with the conventions for a fiction x iff everything required by the conventions for x (whether these conventions are established interfictionally or intrafictionally) is true in w . The worlds relevant to establish what is true in a fiction are worlds meeting the presumption of reliability that are closest to the worlds that obey the conventions for the fiction. Finally, to single out the desired set of worlds, we must consider the worlds meeting the previous requirements that are closest to some set of worlds representing overt beliefs. We suggested that these are the beliefs of the author that are overt in the relevant community. Our proposed truth-conditions for "In fiction x , p " may now be informally stated thus:²⁰

- (21) "In fiction x , p " is true iff p is true in every world w meeting (a)-(d):
- a. the fiction is narrated in w by a teller who believes that x is true in w ,
 - b. w is as much as the teller believes it is as is required by the presumption of reliability,
 - c. among the worlds meeting conditions (a)-(b), w is closest to the set of worlds compatible with the conventions for the fiction actually denoted by x ,
 - d. among the worlds meeting conditions (a)-(c), w is closest to the set of worlds that represents the overt beliefs of the author in the community where the fiction originates.²¹

²⁰Clause (21-a) makes us side with Lewis and Currie in assuming that, for any fiction x , it is true in x that there is someone who tells x . This puts fictions that explicitly claim that there is no narrator in the class of impossible fictions (in our terms, this is the class of fictions such that there are no worlds meeting conditions (a)-(d)). By (21), this means that anything is vacuously true in fictions of this sort. See Lewis (1978, 1983) and Walton (1990) for some strategies to deal with this problem that are compatible with the account we propose.

²¹Or, if you are of a more relativistic persuasion, w must be closest to the set of worlds that represent the overt beliefs of the author relative to our community.

4 Going formal again

We are now in a position to refine the formalism sketched in section 2. In (21) above we make use of the notion of closeness to a set of worlds, where this set may correspond to the overt beliefs of the author of a fiction or to the conventions established for the fiction. Being close to a set of worlds C may be expressed in terms of a system of spheres centered on C of the kind described in Grove (1988). This system is similar to the one proposed in Lewis (1973) for the semantics of counterfactuals. One difference between them is that, while in Lewis's system the spheres are centered on a world, in the system we adopt the centering is on a set of worlds. Intuitively, each ring in the sphere system represents the worlds that count as equally close to the relevant background of overt beliefs. Let S be a collection of subsets of W (i. e., $S \subseteq \wp(W)$). S is a system of spheres centered on a context (set of worlds) C if the following conditions are met:

1. S is totally ordered by \subseteq .
2. $W \in S$ (thus, W is the largest element in S).
3. $C \in S$ and, for any $X \in S$, $C \subseteq X$ (i.e., C is the smallest element in S).
4. For any proposition p , if there is any sphere X in S such that $X \cap p \neq \emptyset$, then there is the *smallest* sphere Y such that $Y \cap p \neq \emptyset$. (This is the limit assumption discussed by Lewis).²²

In virtue of defining properties 1-4, given a system of spheres S centered on C , there is a function g_C defined in this way:

for any $p \in \wp(W)$,
 $g_C(p) = W$ if $X \cap p = \emptyset$ for every $X \in S$;
 otherwise, $g_C(p) = Y \cap p$, where Y is the smallest sphere in S such that
 $Y \cap p \neq \emptyset$.

²²We make this assumption for simplicity sake. Lewis (1973) points out some problems for it. Suppose Theo is 1.75 cm tall and consider the worlds closest to the real one in which it is true that he is taller than that. Clearly, there is no smallest sphere Y centered on the real world in which he is taller than 1.75. Suppose there were such a sphere and that the number that measures Theo's height in the worlds in Y is n . However close he is to being 1.75 in these worlds, there will be a number m such that $1.75 < m < n$ and, for some world w , m is how tall he is in w . Any such world would be closer to the real world than those in Y . Thus, Y is not the smallest sphere centered on the real world in which Theo is taller than 1.75. As Lewis himself observes, examples of this sort may not be decisive to reject the limit assumption. In our case, we might suppose that worlds that are too finely individuated to be discernible from the point of view of the relevant overt beliefs all count as equally close with respect to similarity. Some such move might allow us to keep the limit assumption despite Lewis's counterexamples.

The function g_C associates to every proposition the set of worlds in which the proposition is true that are closest to the context C .

Given a fiction x , let's now define the sets B_x and R_x in this way:

$$B_x = \{w \in W \mid w \in p, \text{ where } p \text{ is a proposition which is an overt belief of the author of } x \text{ in the relevant community.}\}^{23}$$

$$R_x = \{w \in W \mid w \in p, \text{ where } p \text{ is a proposition which follows from the conventions for } x.\}$$

Given a sphere system centered on B_x , the function g_{B_x} associates to every proposition the set of worlds in which the proposition is true that are closest to the context corresponding to the overt beliefs of the author of fiction x in the relevant community. Given a sphere system centered on R_x , the function g_{R_x} associates to every proposition the set of worlds in which the proposition is true that are closest to the context in which everything required by the conventions for fiction x is true.

We will now say that F is the *directly generated content of a fiction x* if F is the set of worlds that meet conditions (a)-(b) in (21) above. Namely,

- (22) given a world w , the *directly generated content of a fiction x in w* is the set of worlds $F_{x,w}$ that meets these conditions: $F_{x,w} = \{w' \in W \mid \text{(a) } x \text{ is narrated in } w' \text{ by someone who believes that } x, \text{ (b) } w' \text{ is as much as the narrator believes it is as is required by the presumption of reliability.}\}$

The directly generated content in this sense is not the literal content of the work of fiction: as we saw in the case of *The Adventures of Huckleberry Finn*, the worlds that meet (a)-(b) need not be worlds in which the content narrated by Huck is true.

We may now refer to the derived context relevant to evaluate what is true in a fiction x in this way:

- (23) Let $F_{x,w}$ be the directly generated content of x in w . Given the systems of spheres S and S' centered, respectively, on B_x and R_x , the *derived context for x in w* (for short $\mathcal{D}_{F_{x,w}}$) is defined thus: $\mathcal{D}_{F_{x,w}} = g_{B_x}(g_{R_x}(F_{x,w}))$.

The translation rule for “In fiction x , p ” may now be stated in this way (where φ translates p and $\mathcal{D}_{F_{x,w}}$ denotes the derived context for the denotations of x , w):

- (24) In fiction x , $p \Rightarrow \lambda C \lambda w [C(w) \wedge \forall w' (\mathcal{D}_{F_{x,w}}(w') \rightarrow \varphi(\mathcal{D}_{F_{x,w}}(w')))]$

²³According to your favorite theory of interpretation, this will be the set of worlds that represents the overt beliefs of the author in the community where the fiction originates, or the set of worlds that represents the overt beliefs of the author relative to our community. Choose the theory you like best.

References

- Alex Byrne. Truth in fiction: the story continued. *Australasian Journal of Philosophy*, 71(1):24–35, 1993.
- Gregory Currie. *The Nature of Fiction*. Cambridge University Press, Cambridge, 1990.
- Umberto Eco. *Six Walks in the Fictional Woods*. Harvard University Press, Cambridge, 1994.
- Stanley Fish. *Is There a Text in This Class? The Authority of Interpretive Communities*. Harvard University Press, Cambridge, MA, 1980.
- Tamar Szabó Gendler. The puzzle of imaginative resistance. *The Journal of Philosophy*, (2):55–81, 2000.
- Adam Grove. Two modelings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.
- Irene Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts, Amherst, 1982. Published in 1989 by Garland, New York.
- David K. Lewis. *Counterfactuals*. Harvard University Press, Cambridge, Massachusetts, 1973.
- David K. Lewis. Truth in fiction. *American Philosophical Quarterly*, 15:37–46, 1978. Reprinted in D. K. Lewis, *Philosophical Papers*, Vol. 1, Oxford University Press, 1983.
- David K. Lewis. Postscript to “Truth in fiction”. In *Philosophical Papers*, volume 1, pages 276–280. Oxford University Press, Oxford, 1983.
- Jeff Ross. *The Semantics of Media*. Kluwer Academic Publishers, Dordrecht, 1997.
- Robert C. Stalnaker. Belief attribution and context. In R. Grimm and D. Merrill, editors, *Contents of Thought*, pages 140–156. University of Arizona Press, Tucson, 1988.
- Robert C. Stalnaker. *Context and Content. Essays on Intentionality in Speech and Thought*. Oxford University Press, Oxford, 1999.

Robert C. Stalnaker. Common ground. *Linguistics and Philosophy*, 25:701-721, 2002.

Kendall L. Walton. *Mimesis as Make-Believe*. Harvard University Press, Cambridge, MA, 1990.

Kendall L. Walton. Morals in fiction and fictional morality. *Proceedings of the Aristotelian Society*, 68:27–50, 1994. Supplementary volume.