Truth in Fiction, Underdetermination, and the Experience of Actuality

Abstract

It seems true to say that Sherlock Holmes is a detective, despite there being no Sherlock Holmes. When asked to explain this fact, philosophers of language often opt for some version of Lewis's view that sentences like 'Sherlock Holmes is a detective' may be taken as abbreviations for sentences prefixed with 'In the Sherlock Holmes stories…'. I present two problems for this view. First, I provide reason to deny that these sentences are abbreviations. In short, these sentences have aesthetic properties that we should not expect of abbreviations. Second, I argue that the apparent truth of these sentences would not be explained even if they were abbreviations. An alternative is presented that avoids these problems. Following Walton, talk about fiction is viewed as a game of make-believe; following Lewis, interpretations of fiction are modelled using possible worlds.

Keywords: truth, fiction, aesthetic experience, underdetermination, Lewis, Walton, make-believe

1. Truth and Fiction

[1] Nicola Sturgeon is a politician.

This sentence seems true. Why is that? The obvious answer is that the sentence *seems* true because it *is* true. The name 'Nicola Sturgeon' picks out Nicola Sturgeon. The predicate 'is a politician' assigns her the property of being a politician. The sentence is therefore true if Nicola Sturgeon is a politician. Because we understand the sentence and we know that Nicola Sturgeon is a politician, we know that the sentence is true.

Compare the following sentence.

[2] Sherlock Holmes is a detective.

This sentence also seems true. Why is that? Following the previous example, we might suggest the name 'Sherlock Holmes' picks out Sherlock Holmes, the predicate 'is a detective' assigns him the property of being a detective, and the sentence is therefore true if Sherlock Holmes is a detective. Because we understand the sentence and we know that Sherlock Holmes is a detective, we know that the sentence is true. The problem is that there is no such individual as Sherlock Holmes for the name to pick out. Nor is it true that Sherlock Holmes is a detective. If Sherlock Holmes were a detective, then there would be a detective named 'Sherlock Holmes', but there is not.

David Lewis offered an explanation that brings [2] back in line, to some extent, with the explanation of [1]. Lewis (1978, p. 37) suggests that sentences like [2] seem true because they are "abbreviations for longer sentences beginning with an operator 'in such-and-such fiction…'.".[1] When evaluated as a simple subject-predicate sentence, [2] cannot be true and, Lewis (1978, p. 38) says, "we may abandon it to the common fate of subject-predicate sentences with denotationless subject terms: automatic falsity or lack of truth value, according to taste." But Lewis suggests that we ordinarily intend statements like [2] to be interpreted as embedded under the 'In the fiction' operator, which is left implicit as a harmless means of abbreviation. Sentence [2], for example, is an abbreviation of [3]:

[3] In the Sherlock Holmes stories, Sherlock Holmes is a detective.

Sentence [2] *seems* true because the sentence it abbreviates, [3], *is* true. The seeming truth of [2] is explained in similar fashion to the seeming truth of [1], albeit with the slight complication that [2] abbreviates a true sentence, rather than being true itself.

How is it that [3] can be true, even though it includes the same denotationless term as [2]? Lewis construes operators like 'In the Sherlock Holmes stories' as intensional operators. A complex sentence of the form 'In the Sherlock Holmes stories ϕ' is evaluated for truth by evaluating the truth of ϕ, not at the actual world, but at other possible worlds. The fact that [2] is not actually true is therefore no barrier to the truth of [3]. We can draw an analogy here to other intensional operators like 'John believes that'. 'John believes that he was abducted by aliens' can be true, even though the embedded sentence 'he was abducted by aliens' is not.

[1] Lewis is explicit that this analysis doesn't apply to all uses of 'Sherlock Holmes'. It doesn't apply, for example, when we say, 'Sherlock Holmes is a fictional character', nor should we take it to apply to the sentences that compose the stories. This paper follows Lewis in excluding these cases.

Rather, the truth of the complex sentence requires that the embedded sentence be true at all worlds consistent with John's belief. Likewise, [3] requires that the embedded sentence [2] is true at all worlds consistent with the Sherlock Holmes stories.

So, [3] is true if and only if [2] is true at every world consistent with the Sherlock Holmes stories. Which worlds are these? As a starting point, Lewis considers the set of all worlds at which the fiction is told as known fact.[2] The set includes every possible world in which the fiction is told just as it is at our world but where the teller *knows* that the story is true, rather than merely *pretending* that it is true, as they do at our world. Were the stories told as known fact, there would be a detective answering to 'Sherlock Holmes' who is known by the storyteller to have gone on the adventures recounted in the stories. At every world in the relevant set, therefore, 'Holmes is a detective' will be true. The stories don't tell us when Holmes's birthday is, however. The stories could be told as known fact were Holmes's birthday January 1st, or January 2nd, or any day of the year. The relevant set will therefore include worlds at which his birthday is 1st January, worlds at which his birthday is 2nd January… and so on.

Much seems true in a story beyond what we are told by the storyteller, however. There are no purple gnomes in the Holmes stories, even though we're never told that explicitly. Lewis offers two ways of explaining these additional facts. First, he suggests that 'In the fiction, ϕ' requires the evaluation of ϕ relative to the closest worlds at which the fiction is told as known fact, resulting in Analysis 1:

> Analysis 1: A sentence of the form "In the fiction f, ϕ" is non-vacuously true iff some world where f is told as known fact and ϕ is true differs less from our actual world, on balance, than does any world where f is told as known fact and ϕ is not true. It is vacuously true iff there are no possible worlds where f is told as known fact.
>
> (Lewis, 1978, p. 42)

Lewis presents a problem for Analysis 1, however. It seems that there are no purple gnomes in the Holmes stories. If there are no purple gnomes in the actual world, then Analysis 1 entails that there are no purple gnomes in the Holmes stories. If it should turn out, however, that there

[2] Lewis extends the position in his footnote 7: "consider worlds where the act of storytelling really is whatever it purports to be—ravings, reliable translation of a reliable source, or whatever—here at our world". For further discussion of this point, see (Hanley, 2004).

are actually a few undiscovered purple gnomes living on the banks of Loch Ness, then Analysis 1 entails that there are purple gnomes in the stories. To avoid this problem, Lewis suggests Analysis 2 on which we should consider the worlds at which the Holmes stories are told as known fact and which adhere, as closely as possible, to the overt beliefs of the community in which the story is written. A belief is *overt* in the community "iff more or less everyone shares it, more or less everyone thinks that more or less everyone else shares it, and so on" (Lewis, 1978, p. 44).

> Analysis 2: A sentence of the form "In the fiction f, ϕ" is non-vacuously true iff, whenever w is one of the collective belief worlds of the community of origin of f, then some world where f is told as known fact and ϕ is true differs less from the world w, on balance, than does any world where f is told as known fact and ϕ is not true. It is vacuously true iff there are no possible worlds where f is told as known fact.
> (Lewis, 1978, p. 45)

Here, the 'collective belief worlds' are those worlds at which the overt beliefs of a community are true. It may turn out that there are actually a few purple gnomes. Yet, it was an overt belief of Conan Doyle's community (as it is in ours) that there are no purple gnomes, so there are no purple gnomes in the stories.

Lewis also mentions, but never fully works out, a third analysis, to allow for what he calls intra- and inter-fictional carry-over. We'll return to this analysis, and the problems it was intended to solve, in section 2.2. For our purposes, however, the difference between these analyses will not matter. The key point is that [3] is true iff [2] is true at every world in some set of possible worlds determined by the fiction.

Lewis's view involves two key claims. First, that sentences like [2] are abbreviations of sentences like [3]. Second, that sentences like [3] are true iff the embedded sentence [2] is true at every world in some set determined by the fiction. I consider these two claims in turn. Section 2.1 argues that sentences like [2] should not be considered as abbreviations of sentences like [3] because the two can evoke very different aesthetic experiences. Section 2.2 then argues that the truth of sentences like [3] cannot be explained by the truth of sentences like [2] at every world in

some set determined by the fiction. Section 3 presents an alternative account of fiction that avoids both problems.

2. Problems for Lewis's Account

2.1 The Experience of Actuality

Lewis's discussion of [2] is based purely on alethic considerations. He wants to explain why [2] seems *true* and suggests that the apparent truth of [2] is inherited from the truth of [3]. When we consider aesthetic features of these sentences, however, significant differences emerge. Sentences like [2] are often experienced as though they are true at the actual world. Call this *the experience of actuality*. When the 'In the fiction' operator is present, however, the embedded sentence is no longer experienced in the same way. I propose that the experience evoked by the complex sentence is explained by the semantics of the 'In the fiction' operator. If the same operator were present in our interpretation of sentences like [2], therefore, the experience of actuality would not arise.

For a particularly clear example of the experience I am interested in, consider the following plaque that apparently sits in St Bartholomew's Hospital in London:

> At this place New Years Day, 1881 were spoken these deathless words: "You have been in Afghanistan, I perceive." by Mr. Sherlock Holmes in greeting to John H. Watson, MD. at their first meeting.

The inscription on the plaque seems true. Holmes did meet Watson in that location on New Year's Day, 1881. According to Lewis, the sentence on the plaque seems true because it abbreviates a sentence prefixed with the operator 'In the Sherlock Holmes stories…'. When we judge that the sentence on the plaque is true, Lewis suggests that we actually evaluate the truth of the sentence it abbreviates. Truth-values are not the only significant aspects of interpretation, however. When we consider aesthetic aspects, we see significant differences between the two versions of the sentence.

The plaque is written as though it commemorates a true historical encounter, and this is no accident; the plaque is formulated to evoke a particular aesthetic experience. When reading

the plaque, we feel connected to Holmes, Watson, and to their initial encounter. The feeling is rather like the feeling one might get when standing at the site of a true historical encounter. Standing at the site of the Academy, we can marvel at the fact that, had we only been there at the right time, we could have seen Plato converse with Aristotle. Reading the plaque generates a similar romantic feeling. We feel that we stand where Holmes and Watson actually met and that, had we only been present at the right time, we could have witnessed this historic encounter. The plaque presents the meeting of Holmes and Watson as though it took place, not merely in some fiction, but in the actual world. Of course, we are not intended to believe that Holmes actually existed or that he really met Watson. The plaque is only playfully to be treated *as though* it documents a true historical encounter, but we miss something central to the purpose of the plaque if we cannot explain the aesthetic experience that it is intended to provoke. Call it *the experience of actuality*.

Consider a version of the plaque with the Lewisian prefix made explicit: 'In the Sherlock Holmes stories, at this place New Years Day…'. This version evokes a very different experience from the unprefixed version of the plaque. It is quite true that Holmes and Watson met at that place and time in the Sherlock Holmes stories, but we no longer feel connected to that encounter. While the actual plaque can be interpreted as though it commemorates an encounter that took place in the actual world, the prefixed version of the plaque makes explicit that the event was merely fictional and it is clear that we could not have witnessed it, even if we had been at the right place and time.

Why is the experience of actuality blocked when the 'In the fiction' operator is made explicit? I suggest that the answer lies in the semantics of intensional operators. According to Lewis's semantics 'In the Sherlock Holmes stories' is an intensional operator that functions to shift evaluation of the embedded sentence from the actual world to some set of worlds determined by the fiction. The difference induced by the 'In the fiction' operator is strongly analogous to the difference induced by other intensional operators. Consider the difference between 'John ate soup' and 'I believe that John ate soup'. The former presents John's soup-eating as an actual event, while the latter presents it as a believed event. This is explained by the semantics of the operator, which shifts evaluation from the actual world to the worlds consistent with my beliefs. I suggest that this explanation extends from familiar intensional operators to

Lewis's 'In the fiction' operators. Just as the experience of actuality is blocked by other intensional operators, it is blocked by the 'In the fiction' operator.

If the experience of actuality is blocked due to the semantics of the 'In the fiction' operator, this poses a problem for Lewis, who suggests that we interpret [2] as prefixed with the same operator. If the operator blocks the experience of actuality, the experience of actuality should also be blocked in the case of [2]. This suggests that the operator is not in fact present in our interpretation of [2].

On what general principles does this argument rest? Does it, for example, assume a principle like "if the truth value of a sentence S is ultimately determined by the truth value of a proposition P, then one has to have the same type of aesthetic experience when entertaining S that one has when entertaining P"?[3] This principle is false. 'I like to start the day with a big caffeinated drink' evokes a very different experience from 'I like to start the day with a big drink containing 1,3,7-trimethylxanthine' but this is not sufficient to refute the view that the sentences express the same proposition and so the identity of caffeine with 1,3,7-trimethylxanthine.[4] Indeed, that principle is unable to accommodate even simple differences of rhyme and rhythm. 'Wear a cummerbund at the United Nations Children's Fund' has rhyme and rhythm not shared by 'Wear a cummerbund at UNICEF' but that difference is far from sufficient to show that the two sentences express different propositions.

The argument above is not intended to rely on any general principle that would lead to these problematic conclusions. Insofar as any general principle is at play, it is simply the principle that any difference in the experience evoked by two sentences calls for some explanation, but different explanations will be called for in different cases. In the case of rhyme and rhythm, for example, we need a phonetic explanation, rather than a semantic explanation.

So there is no general principle at play which entails that the case of the plaque requires a semantic explanation. Rather than attempting a deductively valid argument through some such general principle, I argue that a semantic explanation is appropriate through a form of inference to the best explanation, motivated by analogy with other intensional operators. I assume that the analogous difference between 'John ate soup' and 'I believe that John ate soup' should be explained in terms of the intensional semantics associated with the 'believes that'

[3] This principle was suggested, and rightly criticised, by an anonymous reviewer.
[4] This example was suggested by the reviewer.

operator and suggest that the same explanation be extended to the 'In the fiction' operator. I'll note that Lewis's own argument is a form of inference to the best explanation: "The way of the Meinongian is hard," he writes and so he explores "a simpler alternative" (Lewis, 1979, p. 37).

The difference between the prefixed and unprefixed versions of the plaque requires an explanation and I suggest that one can be given in terms of the 'In the fiction' operator that is analogous to the correct explanation of an analogous phenomenon for other intensional operators. If this is the correct explanation of the prefixed version of the plaque, this poses a problem for Lewis, as he explains the apparent truth of the unprefixed version of the plaque in terms of the 'In the fiction' operator. If that operator blocks the experience of actuality, then the experience of actuality should also be blocked for the unprefixed version of the plaque.

Why should we suppose that interpreters will be aware that sentences like [2] are associated with an 'In the fiction' operator? As an analogy, consider Russell's analysis of definite descriptions, which is plausibly intended as an account of the conditions under which sentences involving definite descriptions are *true*, not as an account of what speakers consciously take these sentences to mean.[5] The degree to which ordinary speakers should be aware of Russell's analysis of definite descriptions is a difficult question that I will not try to settle here. The analogy, however, is mistaken in an important respect: Lewis's analysis of statements like [2] is explicitly not a semantic analysis of the conditions under which those statements are literally true. For Lewis, when a statement like [2] is interpreted without the prefix, the empty name renders it ineligible for truth: "we may abandon it to the common fate of subject-predicate sentences with denotationless subject terms: automatic falsity or lack of truth value, according to taste" (Lewis, 1978, p. 38).

Lewis's analysis differs from Russell's analysis of definite descriptions in that it is designed precisely to explain what we take statements like [2] to mean, rather than their literal semantic meaning. Lewis wants to explain why statements like [2] *seem true*, despite the fact that they cannot be, and suggests that we interpret them as abbreviations of prefixed statements like [3], which are literally true: "Thus if I say that Holmes liked to show off, you will take it that I have asserted an abbreviated version of the true sentence" (Lewis, 1978, p. 38). Lewis's account is intended to explain precisely what we 'take it' as asserted by utterances like [2]. That

[5] This problem, and the analogy to Russell's account of definite descriptions, were also suggested by the anonymous reviewer.

is not to say that speakers must be consciously aware that they interpreted statements like [2] as statements like [3] but it is to say that they interpret them as prefixed with the 'In the fiction' operator. If they are interpreted as prefixed, and that operator blocks the experience of actuality, then it is mysterious why they are often associated with that experience.

This fact, I think, explains why examples involving necessary a posteriori identities (such as the caffeine and 1,3,7-trimethylxanthine example) are not a promising line of explanation. These sentences can rationally be assigned different truth-values, which suggests that they are interpreted by different processes. Whatever the difference in the way we interpret sentences about 'caffeine' and the way we interpret sentences about '1,3,7-trimethylxanthine', we can hope to leverage this difference in order to explain differences in the experiences they evoke. Lewis's takes the truth-value assigned to a sentence like [2] as parasitic on the truth-value assigned to a sentence like [3], however, so we cannot appeal to an analogous difference.

As an inference to the best explanation, the argument above can be defeated by a superior explanation that retains Lewis's analysis. I'll note two ways in particular in which the argument could be defused. First, by an alternative explanation of why explicitly prefixed 'In the fiction' sentences block the experience of actuality. The explanation will have to be one that doesn't then extend to the unprefixed sentences. This, it seems to me, rules out any explanation in terms of the semantics of the operator. If the semantics of the operator are responsible for blocking the experience, and our intuitions about the truth-value of unprefixed sentences rely on interpreting them with the prefix, then we cannot take the semantics of the operator to block the experience of actuality. Second, by an explanation of why implicitly prefixed sentences like [2] allow for the experience of actuality, despite our intuitions about truth depending on interpreting them with the prefix. Here, perhaps, we might put pressure on the notion of *implictness*, suggesting that leaving the operator implicit somehow allows us to ignore its semantics. According to Lewis, however, our intuitions about the truth of the sentence rely on the semantics of the operator and it is not clear to me how the prefix could be significant enough to explain our intuitions about truth-value but insignificant enough to have no effect on aesthetic experience.

One might question whether the sentence on the plaque is a viable target of Lewis's analysis in the first place.[6] Lewis is explicit that his analysis is not intended to apply to all statements about fiction. Perhaps statements that induce these interesting aesthetic experiences should be set aside, leaving Lewis's analysis to deal with more purely descriptive cases. On the one hand, I am happy if the cases identified here are ultimately decided to fall beyond the scope of Lewis's analysis. I do not, after all, want to claim that sentences like [2] are *never* used as abbreviations of sentences like [3]. The account is plausible, at the very least, when [2] is uttered in response to 'What do you know about the Sherlock Holmes stories?'. Even if they are not suitable targets for Lewis's analysis, cases like the plaque stand in need of an explanation, and one will be offered in section 3.

On the other hand, I think it is important to note that some of Lewis's primary examples are very much like the plaque. I use the example of the plaque to introduce the problem because it provides a very clear example of the aesthetic experience I want to highlight but that experience is also present in the cases Lewis wants to explain, as evidenced by Lewis's use of the past tense. Lewis opens his paper by telling us that "we can truly say that Sherlock Holmes *lived* in Baker Street" (Lewis, 1978, p. 37. My emphasis). Why the past tense? Richard Hanley (2004, p. 114, footnote 6) suggests that the past tense is appropriate because Holmes eventually leaves Baker Street in the stories. The suggestion, I take it, is that we use the past tense to describe the properties of fictional characters when there is a time in the fiction at which the characters no longer have these properties. Two problems with this account. First, the past tense seems appropriate even to interpreters who aren't aware that Holmes eventually retires to the countryside. Second, the past tense is equally appropriate even when we predicate properties that Holmes retains throughout the stories, as in 'Holmes was highly intelligent' or 'Holmes was an Englishman'.

These predicates, thought of as retained over time, are known as *individual-level* predicates and contrast with *stage-level* predicates that are thought of as restricted to particular times, such as *being in a foul mood* and *smoking on a pipe* (Kratzer, 1995). While it is entirely acceptable to use the past tense when predicating an individual-level predicate of Holmes in an unprefixed sentence, the present tense is preferable when the sentence is embedded under the 'In the fiction' operator; that is, 'In the Sherlock Holmes stories, Holmes is an Englishman' is

[6] As did a different anonymous reviewer.

preferable to 'In the Sherlock Holmes stories, Holmes was an Englishman'. This difference is explained by the phenomenon known as the *lifetime effect* (Musan, 1997). When an individual-level predicate is used in the past tense, as in 'Socrates was argumentative' or 'Holmes was an Englishman', it implies that the subject is now deceased. The lifetime effect is unobjectionable in the unprefixed sentence because the experience of actuality leads us to imagine Sherlock Holmes as an actual individual who lived in Victorian London and is now presumably dead. Adding the 'In the fiction' operator, however, blocks the experience of actuality. This renders the lifetime effect jarring and the present tense is preferred. It seems, therefore, that the experience of actuality is at play, even in Lewis's primary cases.

In closing this section, I want to draw attention to a way in which the experience of actuality is important for understanding philosophical argument. Philosophers often present thought experiments in the form of short fictions, and it is sometimes important that these fictions are interpreted as taking place within the actual world. We have already seen one example: Lewis's argument against Analysis 1.

What if, Lewis (1978, p. 44) asks, there are actually purple gnomes living secretly in a secluded cabin on the banks of Loch Ness?[7] According to Analysis 1, this bizarre fact will carry over into the Holmes stories. Given that purple gnomes actually exist, worlds with purple gnomes differ less from our actual world, on balance, than does any world without purple gnomes. For this thought experiment to establish its conclusion, we need to imagine, not only that there is some possible world at which purple gnomes reside by Loch Ness but that purple gnomes *actually* reside by Loch Ness. Here, it is not only important for aesthetic, but for philosophical reasons, that the purple gnomes are thought of as residents of the actual world.

Where do the purple gnomes live? The purple gnomes live by Loch Ness. According to Lewis's analysis of fiction, the previous sentence strikes us as true (when talking about Lewis's short fiction) because it is interpreted as 'In Lewis's fiction, the purple gnomes live by Loch Ness', where the implicit prefix leads us again to shift evaluation of the embedded sentence from the actual world to some set of non-actual worlds. It is certainly true that the gnomes reside by Loch Ness in worlds consistent with Lewis's fiction but this isn't what the thought experiment

[7] We might quibble as to whether Lewis presents the Scottish gnomes as a fiction, as a hypothesis, as a supposition, or something else. Hypotheses and fiction are not mutually exclusive but whatever the fact about Lewis's presentation, he could have presented the gnomes in the form of a short fiction, giving them names, a backstory, and so on. Lewis, in his *Postscripts*, agrees that a philosophical example "is just a concise bit of fiction" (Lewis, 1983, P. 278).

requires us to imagine; we are supposed to imagine that the gnomes reside by Loch Ness *in the actual world*.

For Lewis, the form of speech exemplified by [2] is merely a matter of locutionary convenience. We omit the 'In the fiction' operator simply because it is economical to do so. Whether the operator is explicitly voiced or inscribed, we evaluate the very same proposition. This section has argued, on the contrary, that there are aesthetically and philosophically significant differences between prefixed and unprefixed versions of these sentences, which speak against Lewis's account. I turn now to a further problem: even if these statements were abbreviations, this would not explain their apparent truth.

2.2 Indeterminate Indeterminacy

Lewis explains the apparent truth of statements about fiction by taking them as abbreviations of statements carrying an 'In the fiction' prefix. For the account to succeed, we need an explanation of why the prefixed statements are true. Lewis presents several analyses, but the present argument targets any account on which [3] is true just in case [2] is true at every world in some set of possible worlds determined by the fiction.

The problem with such an account is that it assumes an implausible level of determinacy regarding fiction. If the truth of 'In the Sherlock Holmes stories, ϕ' depends on the truth-value of ϕ at every world in some set, there are two options: there are no worlds in the set at which ϕ is false and the complex sentence is true, or ϕ is false at some worlds in the set and the complex sentence is false. According to Lewis's account, any sentence of the form 'In the Sherlock Holmes stories, ϕ' must be true or not true. There are many cases, however, in which 'In the Sherlock Holmes stories, ϕ' can be considered true or false with equal justification.

Just which sentences these are can be expected to be contentious. The reader may come up with their own examples. I suggest the following:

> [4] In the Sherlock Holmes stories, H. H. Asquith was Prime Minister when Holmes met Von Bork.[8]

[8]Further examples: 'In *The Final Problem*, Sherlock Holmes dies'. Disregard the later stories and ask whether this was true at the time of publication. There is evidence of a struggle that ended in Holmes falling over a cliff. Is the evidence strong enough to establish Holmes's death? Watson certainly thinks so,

Consider two interpreters, A and B, who disagree about the truth of [4]. Both make their judgments based on the entire corpus of the Holmes stories and (let us assume) all facts about the actual world on which the fictional truths might supervene. Taking all this evidence into account, interpreter A judges that [4] is true. It is clear from the stories that there is a Prime Minister when Holmes meets Von Bork, during the events of *His Last Bow*, but Conan Doyle never specifies who that Prime Minister is. As Holmes meets Von Bork in 1914, when H. H. Asquith was the actual Prime Minister, and no alternative Prime Minister is specified, Interpreter A judges that this fact carries over to the stories from the actual world. This is also the interpretation suggested by Lewis's Analyses 1 and 2.

Interpreter B, however, judges that [4] is false. According to B, the identity of the Prime Minister when Holmes meets Von Bork is left an open question, like the date of Holmes's birthday. Conan Doyle explicitly specifies, during *The Adventure of the Second Stain*, that some Lord Bellinger was Prime Minister, though no such person actually held the position. As some facts about the identities of Prime Ministers are explicitly different in the Holmes stories, Interpreter B judges that no facts about the identities of Prime Ministers carry over from the actual world to the fiction. The problem resembles one that Lewis (1978, p. 45) considers explicitly: intra-fictional carry-over. In the *Threepenny Opera*, several principal characters are supposedly tested and found to be treacherous, which Lewis takes to suggest that other characters in the play would also prove to be treacherous, were they put to the test. Likewise, according to interpreter B, the fact that one Prime Minister in the Holmes stories is explicitly someone other than the actual Prime Minister of the time shows that the identities of Prime Ministers do not carry over to the fiction from the actual world.

Interpreters A and B disagree about the truth of [4]. A considers the sentence true and B considers it false. Both interpretations are perfectly reasonable and there is no reason to suppose that one interpretation is superior to the other. If Lewis's view were correct, however, then the truth of [4] would be determined by the truth of the embedded sentence at each world

---

as did many readers at the time. The body is never found, however, so there may be room to treat Holmes's death as underdetermined. 'In *Blade Runner*, Deckard is a replicant'. The movie (on at least some cuts) allows for the interpretation on which his humanity is left an open question and for the interpretation in which he is clearly shown to be a replicant. 'In *A Feast for Crows*, the gravedigger is Sandor Clegane'. There are some clues as to the identity of the gravedigger. We might think that these clearly establish the gravedigger's identity or that it leaves their identity left open. 'In *Inception*, Cobb is reunited with his children'. The ending might be left open, or there might be enough hints to establish that Cobb sees his children again. Some of these are famous debates (at least within certain circles) but, as the Asquith example shows, we can formulate more mundane cases. My thanks to Mark Thakkar, Tom Hunter, and Martin Jonathan Philip Hendry for some of these examples.

in some set. Either there are no worlds in the set at which the embedded sentence is false, rendering [4] true, or there are some worlds in the set at which the embedded sentence is false, rendering [4] false. If Lewis's view were correct, therefore, either Interpreter A is right and Interpreter B is wrong, or Interpreter B is right and Interpreter A is wrong. Neither Interpreter A nor Interpreter B is uniquely right, however, so Lewis's view is not correct.[9]

An anonymous reviewer suggests that the problem can be dissolved by Lewis's (1983, p. 277) *method of intersection*. Watson has a single war wound that, at different points in the Holmes stories, is said to be in his leg and in his shoulder.[10] The method of intersection runs as follows. We divide the Holmes stories into two maximally consistent fragments. According to one fragment, the wound is in Watson's leg; according to the other, the wound is in his arm. 'In the Sherlock Holmes stories, ϕ' is true for any ϕ that is true in *both* fragments. In the Holmes stories, therefore, it is true that Watson has a war wound because that is true whether the wound is in his shoulder or his leg, but it is not true that the wound is in his leg and it is not true that the wound is in his shoulder. The location of the wound is left open, like the date of Holmes's birthday.

It is not clear how the method would apply to the case at issue. Both interpreters A and B base their judgments about [4] on the entire corpus of the Holmes stories. There is not one fragment in which the Prime Minister is clearly Asquith when Holmes meets Von Bork and another in which the Prime Minister is clearly not Asquith. What is clear, however, is that the method of intersection cannot be used to dissolve the problem. There are two possibilities. Either, in every fragment, Asquith is Prime Minister when Holmes meets Von Bork, or it is not the case that, in every fragment, Asquith is Prime Minister when Holmes meets Von Bork. In the first case, [4] is true, Interpreter A is correct, and B is incorrect. In the second case, [4] is false and only interpreter B assigns it the correct truth-value. Either way, one of the interpreters is correct and the other is wrong. However, neither interpreter is uniquely correct.

[9] The moral of the argument can be put in the terms used by Proudfoot (2006, pp. 25-26), who considers the difficulty of identifying the boundary of a bounded representation: the boundary between those sentences that a fiction makes true or false (like 'Holmes is a detective' and 'Holmes is not a detective') and those sentences the fiction makes neither true nor false (like 'Holmes's birthday is 1st January). Proudfoot's concern is that they may be no non-circular way of identifying the boundary between what is and is not represented. This argument goes further. The problem is not that the boundary is difficult to identify but that there is no such boundary. If there were such a boundary, [4] would either fall on one side of it or the other and either A or B would be correct.

[10] See (Badura and Berto, 2019) for a recent attempt to deal with inconsistent fictions by appeal to impossible worlds.

In his *Postscript*, Lewis presents an alternative method of resolving contradictions. According to this *method of union*, 'In the Sherlock Holmes stories, ϕ' is true for any ϕ that is true in *either* of the maximally consistent fragments. It is true in one fragment that Watson's wound is in his shoulder, so it is true in the fiction that the wound is in his shoulder. It is true in another fragment that the wound is not in his shoulder, so it is true in the fiction that the wound is not in his shoulder. It is not true in either fragment, however, that the wound is and is not in his shoulder, so this contradiction is not true in the fiction.

Again, it is unclear how the stories can be divided into two appropriate fragments. Supposing that the method could be applied to the problematic sentence, however, we get one of two results. Either there is some fragment in which the embedded sentence of [4] is true, or there is not. If there is no such fragment, then [4] is not true. In that case, only Interpreter B assigns the sentence the right truth-value. If there is such a fragment, then [4] is true and only Interpreter A is assigns it the right truth-value. There might also be a fragment in which Asquith is not Prime Minister when Holmes meets Von Bork, in which case both [4] and 'In the Sherlock Holmes stories, it is not the case that Asquith was Prime Minister when Holmes met Von Bork' are true but this is not B's position. According to B, neither of these complex sentences is true. The identity of the Prime Minister when Holmes meets Von Bork is, according to B, like the date of Holmes's birthday: it is left completely open. Utilising the method of union, therefore, one of the interpreters is uniquely correct, which is not the case.

If the problem were restricted to statements like [4], then perhaps we might ignore it on the grounds that Lewis's theory is only intended to explain the apparent truth of statements like 'Holmes is a detective' and 'Holmes is a goblin' that seem clearly true or clearly untrue. The problem is not restricted to contentious sentences like [4], however, but extends to every single sentence that Lewis wants to explain. Everyone agrees that [2] seems true. Lewis suggests that [2] *seems* true because [3] *is* true, and that [3] is true because [2] is true at all worlds in some relevant set. Even in this uncontroversial case, the explanation relies on the assumption that there is a unique set of relevant worlds. If there is such a set, however, then [4] is either true or untrue and either A or B is uniquely correct. Neither of them is uniquely correct, so Lewis's explanation cannot be right. The status of this sentence is not a fringe problem that can be set aside but infects the explanation of even the least contentious cases.

3. An Alternative Account

3.1 Walton on Fiction

Lewis's account sharply distinguishes between the language used by the author and the language we use to describe the fiction they have created. In writing [2], Conan Doyle merely pretends to speak truly and thereby generates fictional truths. We readers, in uttering or inscribing [2], do not merely *pretend* to speak the truth, but really do speak the truth, because [2] abbreviates [3], which is literally true. The problems posed above speak against this distinction. The first problem suggested that unprefixed sentences like [2] are not always abbreviations of prefixed sentences like [3]. By uttering prefixed sentences like [3], perhaps we merely express truths about a fiction but, by uttering unprefixed sentences like [2], we can engage in an aesthetically more interesting and imaginative way. The second problem went further, suggesting that even utterances explicitly prefixed with an 'In the fiction' operator are not always simply true or false but can allow for further imaginative interpretation.

In his *Postscripts,* Lewis (1983, p. 276) agrees that our practices surrounding fiction constitute "a cooperative game of make-believe" in which audiences, as well as speakers, are engaged in pretence. Lewis credits Kendall Walton with inspiring this view of fiction. Lewis's view of sentences like [2], however, remains largely unchanged in light of this insight; they are still viewed as abbreviations of statements like [3]. In this section, I describe an alternative account of truth in fiction, also inspired by Walton's work, that avoids the problems identified in the previous section, while using Lewis's possible worlds machinery to expand on Walton's account.

For Walton, we engage with fiction through games of make-believe. Make-believe is presented by considering children's games. A child can play with a doll in various ways; they might pretend that the doll is a baby or pretend that it is a pillow. These two kinds of games differ in that only games of the first kind are *authorized* games. The function of a doll is to be used in games of the first kind; games in which the doll is imagined to be a baby (Walton, 1990, p. 51). When the doll is used in accordance with this function, we are playing an authorized game. When we pretend that the doll is a pillow, we are not.

Works of literature are also used in games of make-believe. Again, games can be authorized or unauthorized. When we treat the Sherlock Holmes stories as the journals of Dr John Watson, pretending that they recount the actual adventures of an individual from Victorian London named 'Sherlock Holmes', we play an authorized game. In an unauthorized game, I might pretend that the works contain the delusions of explorer driven mad by esoteric secrets learned from an expedition to the most remote corners of the Amazon. That is a perfectly fine game; just not the authorized one.

Lewis's *Postscripts* allow that we engage in games of make-believe when we read stories – we pretend to learn truths from the storyteller – but statements like [2] are still taken to be abbreviations of statements like [3] which are not *pretended* to be true but are *literally* true. For Walton, however, we engage in games of make-believe both when we read the stories and when we evaluate statements like [2]. Again, games can be authorized or unauthorized. How do we engage with [2] when we play an authorized game? We pretend that [2] is true, certainly, but it won't do to pretend that it is true on just any interpretation. We can pretend that the sentence is true by pretending that 'Sherlock Holmes' refers to detective Raymond C. Schindler, but this pretence isn't part of a game authorized by the stories. Walton agrees with Lewis that 'Sherlock Holmes' doesn't refer to anyone. We can't be pretending to assert the proposition that Sherlock Holmes is a detective, therefore, because there is no such proposition (Walton, 1990, p. 391).

Rather, Walton suggests that we engage in a complicated kind of pretence that is difficult to describe in purely descriptive terms. Roughly, we pretend to refer *de re* to an individual called 'Sherlock Holmes', to whom we attribute the property of being a detective. Walton recognizes that this falls short of specifying the precise conditions for engaging in an authorized pretence but takes the pretence to be defined ostensively by the very act of uttering the sentence. By uttering [2], I engage in a particular kind of pretence, call it K, and thereby demonstrate the kind of pretence that is authorized by the stories.

Though [2] semantically expresses no proposition, Walton suggests that it can be used to make truth-conditional assertions. To make this vivid, consider someone who mistakenly says, 'Sherlock Holmes is a policeman'. We can correct them: 'No, Sherlock Holmes is a detective'. Here, we are correcting our interlocutor by showing them what kind of pretence is authorized by the works. We communicate a truth just in case the kind of pretence we engage in is the one that is authorized by the stories. The maker of the plaque is likewise showing readers

how an authorized game of make-believe is played: we pretend that Holmes met Watson here, in St Barts, on a certain date. These utterances are therefore truth-evaluable; [2] is true just in case this kind of pretence is part of the authorized game of make-believe established by the Sherlock Holmes stories. Walton (1990, p. 400) therefore suggests the following paraphrase of [2]:

> [5] The Sherlock Holmes stories are such that one who engages in pretence of kind K in a game authorized for it makes it fictional of himself in that game that he speaks truly.
> (Adapted from (Walton, 1990, p. 400))

Like Lewis, Walton explains the apparent truth of [2] by associating it with another statement, [5], that is literally true. Walton's paraphrase of [2] is far more complex than Lewis's. While it seemed plausible that we interpret [2] as [3], it seems implausible that we interpret [2] in the complex terms of [5]. Walton (1990, p. 404) is clear, however, that we don't have anything like [5] "specifically in mind" when we utter a sentence like [2]. What do we have in mind? Simply [2] itself. We think in the simple terms of [2], rather than the complex terms of [5].

Walton appeals to a complex kind of pretence he calls 'K' and notes that it is difficult to describe in purely descriptive terms. In the following section, I suggest that we can clarify this pretence by viewing it as a playful variant of act one engages in when one mistakes fiction for fact. To align with Walton's observations, the account will meet two criteria: it will explain why the utterance [2] is true if and only if Walton's suggested paraphrase [5] is true, and it will allow that the sentence is interpreted directly, without the audience deriving the suggested paraphrase.

3.2 Mistaking and Pretending

Lewis's strategy was to explain the apparent truth of [2] by taking it as an abbreviation of a statement that is literally true. This manoeuvre is clearly problematic for other cases of apparent truth. Consider someone who mistakenly takes the Holmes stories for works of history rather than fiction. For this interpreter, [2] seems true. Following Lewis's method, we might attempt to explain this apparent truth by taking [2] to be an abbreviation of some sentence that the interpreter can recognize as true. Perhaps

[6] According to my beliefs, Holmes is a detective.

The operator 'According to my beliefs' works like the 'In the fiction' operator, prompting interpretation of the sentence, not relative to the actual world, but relative to worlds consistent with the interpreter's mistaken beliefs. This analysis, however, cannot explain the experience of actuality noted in section 2.1. The mistaken interpreter experiences [2] as *actually* true, not only true according to their beliefs.

Of course, this explanation was very unnatural to begin with. There is little temptation to posit an unpronounced operator, as the case is very simply explained by the fact that the interpreter is wrong about the state of the actual world. The interpreter tries to evaluate [2] for truth at the actual world but they end up evaluating it relative to the way they mistakenly take the world to be. The interpreter will therefore judge the sentence to be true, though it actually is not.

The apparent truth of [2], for the mistaken interpreter, is not explained by taking it as an abbreviation but by considering the way the interpreter takes the world to be. I suggest that we explain the apparent truth of [2] in similar fashion for interpreters who understand that the statement concerns fiction. When interpreting a statement like [2], we do playfully what the mistaken interpreter does in all seriousness; we interpret the sentence as though it were true at the actual world. The assessor's interpretation of the Holmes stories can be represented as a set of worlds and it is relative to this set of worlds that a statement like [2] is evaluated. As there is a detective called 'Sherlock Holmes' at all worlds consistent with my reading of the Sherlock Holmes stories, I judge [2] to be true. As Walton suggests, [2] is interpreted directly, without first deriving an alternative paraphrase, whether of the form of [3] or [5].

How does this explain the truth-conditions that Walton identifies in their paraphrase [5]? The utterance will be judged true only if it is true at all worlds in the set I use to evaluate it, but the worlds themselves are not arbitrary. Just like the mistaken interpreter, the worlds are selected with a goal in mind. The mistaken interpreter intends the set to characterize the actual world. They are mistaken precisely because the actual world is not included in this set. The knowledgeable interpreter's behaviour is also goal-directed: they intend the set to characterize the 'world of the fiction', that is, they intend to evaluate the sentence relative to the worlds compatible with what is true in a game authorized by the stories. In this, they can also be

mistaken. One gets it wrong if one doesn't evaluate the sentence relative to an appropriate set of worlds, where an 'appropriate set of worlds' is one at which (among other things) [2] is true. The utterance is treated as true, therefore, only if the sentence is true relative to worlds that are consistent with the game authorized by the fiction.

Consider again our interlocutor who says, 'Holmes is a policeman'. They evaluate that sentence relative to a set of worlds W at which Holmes is a policeman. They might be playing some inventive and nonstandard game, or they might be intending to speak falsely in a game authorized by the stories, but if they are trying to speak truly in a game authorized by the stories then they have not succeeded. They pretend to speak truly – they pretend that some set of worlds W characterize the actual world and the sentence they utter is true at all the worlds in W – but they do not speak truly in a game authorized by the fiction. To do that, they would have to pretend that a different set of worlds characterize the actual world: a set of worlds at which [2] is true. We correct them: 'No. Holmes is a detective'. In so doing, we pretend that a different set characterizes the actual world, a set at which [2] is true, thereby demonstrating the kind of pretence that is authorized within the game and communicating the truth of [5].

3.3 Avoiding the Problems Faced by Lewis's Account

The view just outlined avoids the problems raised in the section 2. To speak truly within the game authorized by the stories, we must treat the right kind of worlds as actual, but there is no single set of worlds that fits the bill: There are many different ways to play the same game. It isn't clear, for example, whether [4] should be considered true or false. Different interpreters might evaluate sentences relative to different sets of worlds while still playing the same game. As long as the contentious sentence doesn't come up, the difference won't even be noticed, and we can continue playing together.

Compare playing with a doll. I imagine that the stony-eyed doll blinks intermittently, while you do not. Despite this difference in our pretence, we can play together perfectly well. It's only when this difference is made salient ('Hang on! The baby hasn't blinked! We need to take them to the doctor!') that this difference is even noticed. When the difference is noticed, one party can relent, adopting the pretence of the other ('You're right! Let's go to the doctor.') or they can dig in their heels ('No, she's OK. Look: she just blinked!'). If both parties dig in their

heels, the whole game can come apart. Likewise, in the case of fiction, we can play together even if our interpretations of the work diverge, so long as these differences don't come up or can be ignored. When they do come up, and neither party relents, we can have some very enjoyable discussions ('But H. H. Asquith *was* PM when Holmes met Von Bork. That happened in 1914!'), though we might no longer be playing a game authorized by the fiction.[11]

Interpreters may interpret stories differently without any criticism being possible. That is not to say that anything goes, however. While "there is nothing to stop us using a work however we like," (Walton, 1990, p. 397) people can be criticized on the grounds that they aren't using it as they would like. As noted previously, they will often intend to play an authorized game and, as such, they will be sensitive to certain kinds of correction. Interpreters intend to interpret fictions in particular ways. If an interpreter thinks that [2] is not true in the fiction, that is likely because they have made some identifiable mistake; perhaps they misread or forgot what they read in the past. Here, the interpretation on which [2] is false can be criticized without invoking a unique set of fictional truths. Quite generally, whenever someone argues that a certain proposition is true in a fiction, they will appeal to further considerations to justify their opinion. Whatever these considerations be – what is written in the text, what is true at the actual world (as in Analysis 1), or the overt beliefs of the community in which the fiction was produced (as per Analysis 2) – we can appeal to these same considerations to criticize that interpretation of fiction without assuming a unique set of fictional truths.

This account also avoids the deeper problem. Lewis explained the apparent truth of [2] through the literal truth of [3]. The semantics of Lewis's 'In the fiction' operator entailed that every sentence of the form 'In the Sherlock Holmes stories, ϕ' be either true or false. Lewis's analysis of even an uncontentious sentence like [2] assumed the truth or falsity of contentious statements like [4]. The present account of [2] doesn't rely on the 'In the fiction' operator, so the problem isn't so pressing. Still, what is the correct semantic account of the 'In the fiction' operator?

I suggest that we retain the shape of Lewis's semantics but revise our understanding of its role. According to the semantics, 'In the fiction, ϕ' is true iff ϕ is true at every world w consistent with the fiction. The semantics includes a variable ranging over worlds but doesn't

[11] As with "pursuing silly questions", we might "transform an authorized game into an unofficial one" (Walton, 1990, p. 406).

tell us which worlds they are; that is an issue for literary theory, not for semantics. Lewis thinks that the semantic theory should specify the value of the variable because he wants a semantic theory that can be used to tell us whether sentences of this form are true or false. I reject that notion of semantics, preferring to view semantics as explaining how interpreters generate interpretations and why they judge sentences to be true or false. Semantic theories can allow that different interpreters can assign different interpretations to a sentence while still interpreting semantically.

There is an analogy here between interpretations of fiction and interpretation of quantifier domain restriction. We might endorse a semantics on which 'All Fs are G' is interpreted with respect to a contextually-determined domain restricting variable but needn't think that the value of the domain variable is determined by the semantics. There may be an array of different restrictions that are available in the context and different interpreters can rely on different domain restrictions.[12] Likewise, the semantics tells us that truth-value of 'In the fiction ϕ' is determined relative to a set of worlds but doesn't tell us what that set is. Not any set of worlds will do, if we are trying to play the game authorized by the fiction, but it isn't for the semantics to identify the right set.

We can therefore explain why interpreter A and interpreter B can disagree, without assuming that either is uniquely correct. These interpreters have different conceptions of the worlds consistent with the fiction. This is a dispute over literary interpretation, rather than semantics. Feeding their interpretations into the schema given by the semantics, it predicts that interpreters A and B will assign [4] different truth-values. The semantics won't tell us whether [4] is actually true or false, but that isn't its purpose. It is supposed to feature in the explanation of why A judges that [4] is true and B judges that it is false, and it does that perfectly well.

Let's return now to the experience of actuality. On Lewis's account, statements like [2] are abbreviations of prefixed statements like [3]. The problem was that some such statements have aesthetic properties that we should not expect them to have if they are abbreviations. On the account offered here, [2] is interpreted as is, devoid of any 'In the fiction' operator. While [2] is assessed for truth relative to the assessor's interpretation of the Holmes stories, the view is not that statements like [2] are abbreviations of statements like:

[12] [redacted for anonymous review]

[7] According to my interpretation of the Sherlock Holmes stories, Sherlock Holmes is a detective.

or

[5] The Sherlock Holmes stories are such that one who engages in pretence of kind K in a game authorized for it makes it fictional of himself in that game that he speaks truly.

Statements like [2] are interpreted directly, without implicit operators. It is no problem, therefore, that they evoke very different aesthetic experiences to sentences like [3], where the operator is attached.

The experience of reading the plaque is explained by analogy with the mistaken interpreter who believes that the Holmes stories are historical accounts. Though the mistaken interpreter evaluates the plaque with respect to a set that excludes the actual world, they treat this set as though it included the actual world. The aesthetic experience evoked by the plaque is explained by our doing playfully what the mistaken interpreter does in all seriousness: we treat the plaque as though it depicts actuality. This generates the feeling, though not the belief, that we stand just where Holmes actually met Watson. Likewise, when considering Lewis's thought experiment, we treat worlds that incorporate purple gnomes as though they represent actuality, while knowing full-well that they do not.

4. Conclusion

Lewis wants to explain the apparent truth of statements about fiction by taking them to be abbreviations of literally true statements prefixed with an 'In the fiction' operator. I raised two problems. First, the aesthetic experience of interpreting statements about fiction suggests that they are not abbreviations of 'In the fiction' statements. Second, their apparent truth would not be explained even if they were abbreviations. Both problems are resolved by moving towards an alternative account on which statements about fiction are directly interpreted relative to a set of worlds determined by the interpreter's understanding of the fiction. This view is an extension of

Walton's view on which sentences about fiction are used to communicate truths about authorized games of make-believe.

Bibliography

Badura, C. and Berto, F. (2019) 'Truth in Fiction, Impossible Worlds, and Belief Revision'. *Australasian Journal of Philosophy*, 97, pp. 178-193.

Hanley, R. (2004) 'As Good as it Gets: Lewis on Truth in Fiction'. *Australasian Journal of Philosophy*, 82, pp. 112-128.

Kratzer, A. (1995) 'Stage-Level and Individual-Level Predicates', in Carlson, G. N. and Pelletier, F. J. (eds), *The Generic Book*, 125-175. Chicago: The University of Chicago Press, pp. 125-175.

Lewis, D. (1978) 'Truth in Fiction'. *American Philosophical Quarterly*, 1, pp. 37-46.

Lewis, D. (1983) 'Postscripts to "Truth in Fiction"'. *Philosophical Papers Volume 1*. Oxford: OUP, PP. 276-280.

Musan, R. (1997) 'Tense, Predicates, and Lifetime Effects'. *Natural Language Semantics,* 5, pp. 271–301.

Proudfoot, D. (2006) 'Possible Worlds Semantics and Fiction'. *Journal of Philosophical Logic*, 35, pp. 9-40.

Walton, K. L. (1990) *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA: Harvard University Press.