Fictional Reality*

Kyle Blumberg and Ben Holguín

December 2023

Abstract

This paper defends a theory of fictional truth. According to this theory, there is a fact of the matter concerning the number of hairs on Sherlock Holmes' head, and likewise for any other meaningful question one could ask about what's true in a work of fiction. We argue that a theory of this form is needed to account for the patterns in our judgments about attitude reports that embed fictional claims. We contrast our view with one of the dominant approaches to fictional truth, which originates with David Lewis. Along the way we explore the relationship between fiction, counterfactuals, and vagueness.

1 Introduction

Sherlock Holmes first meets his assistant Dr. Watson in Sir Arthur Conan Doyle's *A Study in Scarlet*. Holmes has some number of hairs on his head at the time of this meeting. Must it follow, then, that one of the following claims expresses a truth (Lewis, 1978, p. 42)?

(1) a. At the moment he first meets Watson, Holmes has an odd number of hairs on his head.

b. At the moment he first meets Watson, Holmes has an even number of hairs on his head.

A Study in Scarlet does not explicitly state the exact number of hairs on Holmes' head. Moreover, it seems safe to assume that, at the time of writing,

*Thanks to Sam Carter, Sinan Dogramaci, Cian Dorr, Simon Goldstein, Jeremy Goodman, Al Hájek, John Hawthorne, Robert Hopkins, Harvey Lederman, Matt Mandelkern, Ian Phillips, and Hanna Pickard for very helpful discussion. We would also like to thank audiences at Dianoia, USC, UT Austin, Johns Hopkins, the New York Philosophy of Language Workshop, and the Australasian Association of Philosophy conference for their thoughtful feedback.

Sir Arthur Conan Doyle had no thoughts or intentions with respect to this matter either. As such, the question of whether Holmes has an odd or even number of hairs on his head at the time of meeting Watson seems to admit no determinately true answer. One might conclude that it has no true answer at all.

(1a) and (1b) are paradigm cases of *fictional indeterminacy*, the central topic of this paper. For the most part we'll leave the notion at an intuitive level, and allow our judgments about examples to guide our conception of the phenomenon. But the rough idea is that a fictionally indeterminate claim is a claim about the events of a work of fiction that is left "unsettled" by what's explicitly said or depicted in the work of fiction, what the author intended or believed to be true in the work of fiction, and so on.[1]

We assume that fictional indeterminacy—like any other form of indeterminacy—carries an epistemic constraint: there are in-principle barriers to knowing that a fictionally indeterminate claim expresses a truth. Given the actual historical facts concerning Sir Arthur Conan Doyle and *A Study in Scarlet*, no amount of textual or psychological research will reveal which of (1a) or (1b) is true. So neither can be known to express a truth.

But knowledge is one thing, truth another. Could a fictionally indeterminate claim like (1a) or (1b) be true *simpliciter*, even if not knowably so?

This is the question that will occupy us for most of this paper. We seek to make progress on it by considering what other rational attitudes we can take toward fictionally indeterminate claims. For example, can one believe the propositions expressed by these claims? Assign meaningful subjective probabilities to them? Wish they were true? Wonder whether they are true?

In what follows, we provide detailed answers to these questions. We will argue that the best way to capture the patterns in our answers is to adopt an account on which the semantic value of a fictional statement is determined by how things go at a single, unique world—the relevant "world of the fiction". Since worlds settle every (precise) question, on the theory we defend there *is* a fact of the matter as to how many hairs are on Holmes' head. It's just that no one—perhaps not even Sir Arthur Conan Doyle himself—is in any position to know what these facts are. As such, fictional indeterminacy entails neither falsity nor the absence of a truth-value; the question of whether Holmes has an odd number of hairs on his head is as factual as the question of whether he is a detective.

Philosophical orthodoxy takes such an account of fictional truth to be obviously untenable. For instance, both Lewis (1978, p.42) and Proudfoot (2006, p.11) call views akin to the one we defend "absurd".[2] Nonetheless,

[1] We'll say more to sharpen this notion in §4.

[2] As far as we know, Woods (2018) is the only other proponent of a view on which

we believe that such an account is the best available when it comes to explaining the patterns in our rational attitudes toward indeterminate fictional claims. These claims are legitimate objects of curiosity and wonder, and it is far from clear how to make sense of this and other related facts on the standard assumption that claims like (1a) and (1b) are invariably false or truth-valueless.

Most of the paper will be spent explaining and defending these ideas. But we will also argue that our account offers an interesting perspective on a number of issues in the semantics and metaphysics of fiction. These include questions about the grounds of fictional indeterminacy; the connections between fictional truth and counterfactual truth; whether fictional indeterminacy is a species of vagueness; and whether fictional worlds are created or discovered.

2 Varieties of fictional claims

By way of making our subject matter clear, we'll start with some distinctions between statements about fiction.

As Kripke (2013) noted, in making a statement about a work of fiction, there seem to be two different kind of claims we might be making: either a claim about matters *external* to the fiction, or a claim about matters *internal* to it.[3] As a rough gloss, the first kind of claim is a claim about the fiction itself—the fiction as an abstract entity created by a particular author. The second kind of claim, by contrast, is in some sense a claim about what's true *in* or *by the lights of* a certain work of fiction.

The distinction is best grasped through examples. Paradigm examples of fictional claims that are true when taken externally include:

(2) a. Sherlock Holmes is a fictional character.
b. Hamlet isn't a person, he's the protagonist of *Hamlet*.
c. Michael Mann's *Heat* is based loosely on the true story of Detective Chuck Adamson's pursuit of Neil McCauley.

And paradigm examples of fictional claims that are true when taken internally include:

(3) a. Sherlock Holmes lives on Baker Street.

exactly one of (1b) or (1a) expresses a truth. However, Woods doesn't provide a general theory of fictional discourse (since he claims that it is not possible to do so), and is motivated by quite different considerations from our own.

[3]See also van Inwagen (1977); Lewis (1978); Salmon (2011).

b. [*Hamlet* features a fictional play called *The Murder of Gonzago*.] Hamlet is a person, but Gonzago isn't.[4]

c. Tony Soprano is the most powerful criminal in New Jersey.

Supposing, as we do, that only flesh-and-blood people actually live on Baker Street, (2a) and (3a) cannot both be uttered truly in the same breath. Still, they both seem to express truths on their natural readings—hence the external/internal distinction.

In addition to the practice of making external and internal fictional claims, there is also a practice of making fictional claims using explicit natural language "fiction operators", like 'in the fiction', 'according to the fiction', and so on.[5] For example, consider the following claims:

(4) a. According to *A Study in Scarlet*, Sherlock Holmes lives on Baker Street.

b. By the lights of *Hamlet*, Hamlet is a person, but Gonzago isn't.

c. In *The Sopranos*, Tony Soprano is the most powerful criminal in New Jersey.

The kinds of facts that would seem relevant to the truth of these claims look a lot like the kinds of facts that would seem relevant to the truth of internal claims like (3) (and unlike the kinds of facts relevant to the truth of external claims like (2)). However, we will *not* be assuming that our intuitive judgments about (3) and (4) are reflective of some unitary phenomenon.[6] As we intend to use the term 'internal fictional claim', it is to be understood exclusively via ostension to our intuitive judgments about sentences like (3)—sentences that do not contain any explicit natural language operators. As such, we leave it as an open possibility that a certain fictional claim p could be true on its internal reading even when a sentence like ⌜According to the relevant work of fiction, p⌝ is false on all of *its* available readings.

With these distinctions in mind, our interest in this paper is with the nature of *internal* fictional truth. Why is it true that Sherlock Holmes lives on Baker Street, but not true that Sherlock Holmes is from Canada? And how should we think about internal claims that seem entirely indeterminate, like (1a) and (1b)?

We believe that these sorts of questions can be explored fruitfully from a position of neutrality on the various semantic, epistemological, and metaphysical issues raised by external fictional claims like (2), as well as questions

[4]This example is due to Kripke (2013).

[5]For some recent work on the semantics of explicit fictional operators, see, e.g., Sainsbury (2014), Dohrn (2015), Voltolini (2019), Semeijn (forthcoming).

[6]Cf. Bowker (2021).

concerning the semantics of (lexicalized) natural language fiction operators. So unless explicitly stated otherwise, for the rest of the paper the reader should interpret our talk of 'fictional claims' and 'fictional truth' exclusively in terms of the internal notions.[7]

3 The modal approach to fictional truth

Consider a straightforward fictional claim like (5):

(5) Gandalf is a wizard.

(5) seems to express a truth that is both knowable and assertable. However, standard assumptions in semantic theory imply that (5) is true only if (i) the entity denoted by 'Gandalf' exists and (ii) has the property expressed by 'is a wizard'. We are comfortable thinking (i) holds: it is plausible that anything that is in the extension of 'is a fictional character' is something that exists, and Gandalf is in the extension of 'is a fictional character' if anything is.[8] It's (ii) that creates a problem. There are no wizards in actuality, so plausibly nothing is in the extension of 'is a wizard'—not even Gandalf the fictional character. So the relevant interpretation of (5) cannot be one in which we look to the actual extension of 'is a wizard' in determining its truth-value.

To solve this problem, theorists have adopted a framework for modeling truth in fiction that goes back to Lewis (1978).[9] The central idea is that internal fictional claims are *implicitly modalized external claims*. That is to say: on its internal reading, a fictional claim is prefixed by a silent operator that shifts the evaluation of its (externally interpreted) prejacent to some number of non-actual possibilities.

Where f is a fictional story—for example *The Lord of the Rings* as written by J. R. R. Tolkein—the corresponding operator can be represented as $\Box_f$. To a first approximation, $\Box_f$ shifts evaluation to the worlds "compatible with the fiction f". These worlds can be represented by the set $\mathrm{Fic}_{f,w}$: the set of worlds compatible with the fiction f at w. Regarding the metaphysical status of these worlds, it is left open whether they represent genuine metaphysical

[7] Another notational point: for the sake of vivacity we will sometimes talk about a certain claim p being true *in* a work of fiction, as we did when we introduced the notion of an internal fictional claim above. In speaking this way, the reader should understand us simply as claiming that p is an internal fictional truth, and not as in claiming that an English sentence like ⌜In the relevant fiction, p⌝ expresses a truth.

[8] This is not to say this assumption is incontestable—see Bacon (2013) for discussion.

[9] For recent uses of the framework, see, e.g., Hanley (2004); Badura & Berto (2019); García-Carpintero (2022).

possibilities. However, they are usually assumed to be both *consistent* and *complete*. That is: for any world w and proposition p, exactly one of p or $\neg p$ is true at w.[10]

The modal approach raises an issue that is worth addressing immediately. Given that $\Box_f$ is an implicit operator, a surface-form sentence like 'Gandalf is a wizard' is systematically ambiguous: either it could denote an unmodalized claim (on its external reading), or it could denote a modalized claim prefixed by $\Box_f$ (on its internal reading). To get around this problem, we will represent logical forms with a designated sans serif font, e.g. p. So, although 'Gandalf is a wizard' is ambiguous, $\mathsf{Gandalf\ is\ a\ wizard}$ is not: the latter always expresses the "unmodalized" proposition that Gandalf is a wizard. We will also freely abuse notation so that an expression in sans serif will stand both for a (disambiguated) logical form, as well as the proposition expressed by that form. Given these conventions, the internal reading of (5) can be represented by (6):

(6) $\Box_f\mathsf{Gandalf\ is\ a\ wizard}$

Here, then, is the bare-boned analysis of the modal operator $\Box_f$:

Modal analysis

$\Box_f\mathsf{p}$ is true at w iff $\forall w' \in \mathrm{Fic}_{f,w}$: p is true at w'.

Given the modal analysis, (6) is true at w if and only if every world in $\mathrm{Fic}_{f,w}$ is a world in which $\mathsf{Gandalf\ is\ a\ wizard}$ is true, i.e. is a world in which it's externally true that Gandalf is a wizard.

We will be taking the modal analysis for granted in what follows. By our lights, the main attraction of the modal framework is that it allows us to state and assess predictions of various theories of fictional truth in a manner that is reasonably precise. As we will see in the next few sections, for example, a number of debates around the nature of fictional truth can be modeled as debates about the *cardinality* of the set $\mathrm{Fic}_{f,w}$.

But the claim that indeterminate fictional claims can often be true—our central thesis—is not tied in any essential way to the modal approach.[11] The argument for this thesis is grounded in the patterns in our intuitive judgments about fictional claims—patterns that any good theory of fictional truth ought to be able to explain. The modal view just happens to be a

[10] We will revisit this assumption in some detail in §11.

[11] Alternatives to the modal approach include Meinongian views (Parsons, 1980; Sylvan, 1980; Zalta, 1983; Jacquette, 1989), contextualist theories (Predelli, 2008; Antonsen, 2020), and "make-believe" theories (Currie, 1990; Walton, 1990; Byrne, 1993; Gatzia & Sotnak, 2014; Bowker, 2021).

particularly useful way of modeling the metaphysical principles that underly them.

These points aside, the modal approach *on its own* does not make all that much headway on the issue of fictional truth. For all it says is that an internal fictional statement is true just in case the proposition expressed by the statement (on its external reading) is true at all the worlds "compatible with the relevant fiction". But which set of worlds is that? To make progress on this question, we'll start by introducing a notion that will be important in much of what's to come.

4 Principal fictional truths

Taking inspiration from Lewis (1978) and Walton (1990), let us say that that for any given work of fiction f, there is a set of *principal truths* of f.[12] Intuitively speaking, these are claims the fiction is in some sense *straightforwardly committed to*.

It is difficult to give a precise characterization of the notion of a principal fictional truth.[13] Like the distinction between external and internal fictional claims, we believe the notion is best grasped through examples. So here are some paradigm cases of sentences that report principal fictional truths:[14]

(7) a. Sherlock Holmes is a detective.
b. Halmet has two nostrils.
c. Ignatius Reilly believes the modern world is lacking in theology and geometry.
d. Elizabeth Bennett was born many years after the death of Caeasar.
e. The disease Katerina Ivanovna suffers from is tuberculosis.

[12] One word of caution about the nomenclature here. We use the term 'principal' rather than the oft-used 'primary', since we think theorists like Lewis and Walton are tracking a slightly different notion with their use of the term 'primary truth' than the one we're interested in, and so would like to avoid confusion. On standard usage, the set of primary fictional truths is the set of claims that are *explicitly* stated or depicted in the fiction. For instance, relative to *A Study in Scarlet*, it is a primary truth that Holmes lives on Baker Street, but *not* a primary truth that Holmes has two nostrils, since nowhere does the book mention Holmes' nostrils. However, as we intend to use the term 'principal', these are both principal truths.

[13] For some recent attempts see, e.g., Friend (2017), Motoarc (2017), Badura & Berto (2019), Stokke (2021), Franzén (2021), and Skow (2022).

[14] These examples are about, respectively, *A Study in Scarlet*, *Hamlet*, *A Confederacy of Dunces*, *Pride and Prejudice*, and *Crime and Punishment*.

And here are some paradigm cases of sentences that do *not* report principal fictional truths:[15]

(8) a. Holmes has an odd number of hairs on his head.
b. Hamlet was born on a Tuesday.
c. Ignatius Reilly has tried reading *The Critique of Pure Reason*.
d. Elizabeth Bennett eventually dies in her 70s.
e. Raskolnikov has eczema.

These examples suggest some heuristics for the application of the distinction. Generally speaking, if a work of fiction explicitly says or depicts some claim, then that claim is a principal fictional truth of that fiction—hence (7a) and (7c).[16] But explicitness is not a precondition on principality. Some claims might count as principal fictional truths not because they are explicitly stated or depicted, but because the author of the fiction *intended* or *believed* them to be true in the fiction, or because they are obvious consequences of the things the author intended or believed to be true in the fiction. We take (7b) and (7d) to be suggestive of this possibility. Finally, it is plausible that some claims are principal fictional truths simply because they are needed to make the story "sufficiently realistic", modulo the constraints imposed by the other conditions. This seems to be the lesson of examples like (7e), given that Dostoevsky had no idea that consumption was caused by the tuberculosis bacterium (Franzén, 2021).[17]

These cursory remarks aside, we intend to treat the category of a principal fictional truth essentially as a black box, and trust that the motivating examples are enough to make it theoretically useful. The only substantive assumption we make about the principal truths is that they are all and only the *determinate* fictional truths. By extension, we assume that if one can know that a fictional claim is true, then that claim reports a principal fictional truth.

We are now in a position to state a minimal constraint on $\text{Fic}_{f,w}$. Letting $\text{P-Fic}_{f,w}$ be the set of worlds compatible with all the *principal* truths of f at w, we can say that $\text{Fic}_{f,w} \subseteq \text{P-Fic}_{f,w}$. So anything that is entailed by a principal fictional truth is a fictional truth simpliciter.

[15] Williams & Woodward (2019) discuss a number of other examples of indeterminate fictional claims (what they call "fictional incompleteness"), including the claim that Deckard is a replicant (in the original *Blade Runner* film) and that Juliet has blue eyes (in *Romeo and Juliet*).

[16] We say 'generally speaking' because some works of fiction have unreliable narrators, while others describe moral impossibilities and other sources of so-called "imaginative resistance". For more on the first issue, see Currie (1990); Walton (1990); Maier & Semeijn (2021). And for more on the second, see Gendler (2000); Weatherson (2004); Altshuler & Maier (2020).

[17] Cf. Friend (2017).

The relevant question is whether $\text{Fic}_{f,w} = \text{P-Fic}_{f,w}$. That is: whether the fictional truths are all and *only* those which are entailed by the principal fictional truths.

5 The standard analysis

Proponents of what we will call the *standard analysis* of fictional truth answer this question in the affirmative: the fictional truths are all and only those which are entailed by the principal fictional truths.

> **Standard analysis**
>
> $\Box_f \mathsf{p}$ is true at w iff $\forall w' \in \text{P-Fic}_{f,w}$: p is true at w'.

The canonical defense of the standard analysis is due to Lewis (1978), though there have been a number of refinements since.[18] Our arguments target rather general features of the standard view, so we'll paint with a broad brush in what follows.

Lewis, like many other theorists of fictional truth, writes as if it's obvious that indeterminate fictional claims like (1) and (8) cannot report truths (1978, p. 42):[19]

> Is the world of Sherlock Holmes a world where Holmes has an even or an odd number of hairs on his head at the moment when he first meets Watson? What is Inspector Lestrade's blood type? It is absurd to suppose that these questions about the world of Sherlock Holems have answers.

What seems to motivate Lewis' thinking here is the combination of two ideas: (i), the basic observation that nothing about *A Study in Scarlet* or the psychological state of Sir Arthur Conan Doyle could plausibly settle the answer to these questions one way or the other. And (ii), the philosophical intuition that if the truth of a fictional claim cannot be grounded in these

[18] See, e.g., Proudfoot (2006); Badura & Berto (2019); García-Carpintero (2022) for a representative sample. Note that much of Lewis' paper is devoted to (i) articulating and defending the modal analysis and then (ii) giving a broadly counterfactual theory of the principal fictional truths. The idea that the fictional truths are all and only the principal fictional truths is all but taken for granted by Lewis and many others in the subsequent literature.

[19] See also Proudfoot (2006, p. 11).

factors, then it cannot be among the truths of the relevant work of fiction.[20] Hence the identification of fictional truth with principal fictional truth.

To see how the standard analysis is working in more detail, consider (1a). By hypothesis, (1a) is true if and only if $\Box_{SS}$Holmes has an odd number of hairs on his head is true.[21] Thus, according to the standard analysis, (1a) is true if and only if: $\forall w' \in$ P-Fic$_{SS}$, Holmes has an odd number of hairs on his head is true at w'. But Holmes has an odd number of hairs on his head is not among the principal fictional truths of *A Study in Scarlet*. So $\exists w' \in$ P-Fic$_{SS}$ such that Holmes has an odd number of hairs on his head is false at w'. This entails $\neg\Box_{SS}$Holmes has an odd number of hairs on his head. Thus, (1a) is false.

The analysis goes through in essentially the same way with (1b) and the examples in (8). Each of these is a claim of the form $\Box_f \mathsf{p}$, where p is not among the principal truths of f. So, by the lights of the standard analysis, each of these claims must be false.

In fact, the standard analysis is committed to something stronger: not only are the claims in (1) and (8) all false, they are all *knowably* false. In the case of (1a)/(1b), anyone sufficiently familiar with *A Study in Scarlet* knows that the question of how many hairs are on Holmes' head is left unsettled by the sorts of factors that determine the principal truths: what's explicitly said in the story, Sir Arthur Conan Doyle's beliefs and intentions, and so on. So anyone sufficiently familiar with *A Study in Scarlet* should know that neither the claim that Holmes has an odd number of hairs on his head nor the claim that he has an even number of hairs on his head is a principal truth. Thus, if the standard analysis correct, then such a person can know that both (1a) and (1b) are false.

The question is whether this is a good prediction. We will now present a range of observations suggesting it is not. And the moral we will draw from these observations is that any good theory of truth in fiction will have to allow that the fictional truths vastly outstrip the principal fictional truths.

6 Attitudes toward fiction

In this section, we present the data that ultimately motivates our account of fictional truth. Our primary example comes from the television series *JAG*.

[20] There are moments where Lewis seems to flirt with the idea that fictional claims that do not report principal truths are neither true nor false (see especially p. 43). We find it difficult to square what he says in these passages with the actual analyses he gives later in the paper. In any event, in §7.2 we consider and argue against variants on the standard analysis that pursue this line.

[21] Here 'SS' picks out *A Study in Scarlet*.

In the final episode, the two protagonists Mac and Harm are engaged to be married, but have jobs in different countries. They agree that one of them will resign from their position in order to join the other, depending on the flip of a fair coin tossed by their friend, Bud. If the coin lands heads, Mac will join Harm; and if it lands tails, Harm will join Mac. The coin is flipped, but the show dramatically ends with the coin in mid-air.[22]

Given these facts about *JAG*, we take it that both (9a) and (9b) fail to express principal fictional truths:

(9) a. Bud's coin landed heads.
b. Bud's coin landed tails.

After all, neither is settled by any of what's explicitly depicted in the show; nor is either plausibly an obvious consequence of anything that's depicted; nor is either needed to make the events of the show sufficiently realistic. And we will assume, perhaps counterfactually, that the writers of *JAG* never had any intentions or beliefs with respect to the outcome of the flip either: they finished writing the final scene, and then never thought about the events of *JAG* again.[23]

The question we now wish to ask is this. Given that you know that neither (9a) nor (9b) reports a principal fictional truth of *JAG*—that both are thoroughly unsettled by that which is explicitly stated or otherwise depicted in the show—what attitudes can you rationally hold towards these claims?

(For brevity's sake we'll focus on attitudes toward (9a) in particular. Given the symmetries of the example, we see no harm in assuming that the range of attitudes you can rationally take toward (9a) are exactly those you can take toward (9b).)

6.1 Epistemics

Unsurprisingly, knowledge ascriptions sound uniformly terrible:

(10) a. ✗ I know that Bud's coin landed heads.
b. ✗ I know whether Bud's coin landed heads.

[22] The final shot of the show can be seen here: https://en.wikipedia.org/wiki/JAG_(TV_series)#/media/File:JAGcoin.png.

[23] We also assume that the writers intended that Bud's coin landed either heads or tails, and thus that it didn't land on its side, wasn't destroyed mid-air, etc. In other words, we are assuming that it's a principal fictional truth of *JAG* that Bud's coin landed either heads or tails.

Our intuitive judgment about (10a) is to be expected given that (9a) is a paradigm case of an indeterminate fictional claim. Whatever else we know about such claims, we know that we cannot know them to express truths. So (10a) must be false.

Likewise, to our ears, (10b) seems problematic for essentially the same reason (10a) is. You can't know that Bud's coin landed heads, and you can't know that it landed tails. So you can't know how it landed. But if you can't know how it landed, then how could know whether it landed heads?[24]

Notice as well that the reasoning here is transparent: anyone who knows that it's indeterminate how Bud's coin landed can reason in the way we just did to the conclusion that they don't know whether Bud's coin landed heads. So (10a) and (10b) aren't just false, they're *knowably* false.

6.2 Doxastics

So much for knowledge ascriptions. What about belief ascriptions? This depends on what kind of belief you're talking about (Goodman & Holguín, forthcoming). If 'believes' expresses a strong doxastic attitude along the lines of what is expressed by 'is sure' or 'is certain', then the belief ascriptions pattern in the way of the knowledge ascriptions:

(11) a. ✗ I am sure that Bud's coin landed heads.
b. ✓ I'm not sure whether Bud's coin landed heads.[25]

These judgments can be further bolstered given a natural assumption about the relationship between knowledge and surety. And that is the assumption that if one knows that one doesn't know whether p, then if one is rational, one is not sure whether p, and thus not sure that p (Unger, 1975; Williamson, 2000; Goodman & Holguín, forthcoming). Thus, since one knows that (10b) is false, it follows that (11a) is false and (11b) is true.

Consider now the weaker interpretations of 'believes', on which it expresses a doxastic attitude along the lines of what is expressed by 'thinks' (Hawthorne *et al.* , 2016; Dorst, 2019; Holguín, 2022). Holding fixed the details of the *JAG* story, the relevant report continues to seem problematic:

[24]The claim that no one knows whether Bud's coin landed heads is further supported by a piece that appeared in *Country Living* whose headline was '*NCIS: LA* Will Finally Reveal Who Won the Coin Flip on the *JAG* Finale' (https://www.countryliving.com/life/entertainment/a27396842/jag-who-won-coin-flip/). As the author Megan Stein points out, '. . . the show decided to leave us in suspense, as we never found out if it was heads or tails'.

[25]Here we use a (felicitous) negated surety ascription, since 'I'm sure whether Bud's coin landed heads' is ungrammatical.

(12) ✗ I think that Bud's coin landed heads.

But here the diagnosis is different. Unlike the attitude of being sure, the attitude of thinking is compatible with known ignorance: even if I know that for all I know I'll win the upcoming lottery, I can still rationally think that I won't. But, intuitively, thinking that a proposition is true requires having *some* reason to favor that proposition over the relevant alternatives.[26] And in the case of Bud's coin, one has no reason to think the coin landed heads, and also no reason to think it didn't.

So now suppose *JAG* had instead ended with the flip of a coin that has a 2:1 bias in favor of heads, but where again the result is neither explicitly nor implicitly depicted. In this case the analog of (12) seems perfectly acceptable:

(13) ✓ I think that Bud's biased coin landed heads.

So can one *believe* indeterminate fictional claims? Yes and no. If 'believe' is interpreted strongly, then no. If 'believe' is interpreted weakly, then in at least in some cases, yes.

Similar points go for the attitude of doubting:

(14) a. ✗ I doubt that Bud's coin landed heads.
b. ✓ I doubt that Bud's biased coin landed tails.

Here we assume that to doubt that p is to think that $\neg p$ (Anand & Hacquard, 2013). Since I have no reason to think Bud's fair coin didn't land heads, I cannot rationally doubt that Bud's fair coin landed tails. But since I do have reason to think that Bud's *biased* coin landed heads, I *can* rationally doubt that Bud's biased coin landed tails.

6.3 Subjective probability

What about some of the more quantitative doxastic attitudes? There appears to be no barrier to rationally assigning non-zero subjective probabilities to indeterminate fictional claims.

(15) a. ✓ It's .5 likely that Bud's coin landed heads.
b. ✓ It's twice as likely that Bud's biased coin landed heads as it is that it landed tails.

[26] For further discussion of the exact form of these sorts of norms, see Holguín (2022); Dorst & Mandelkern (2023); Skipper (forthcoming).

To our ears both reports are perfectly felicitous. In the case of (15a), I know that in *JAG*, Bud's fair coin is flipped. I know that *JAG* itself does not explicitly settle the question of how it landed, as well as that the writers of the show had no intentions with respect to this question either. And I know that a fair coin—whether here or in the world of *JAG*—is just as likely to land heads as it is to land tails. So it seems perfectly reasonable to have .5 credence that Bud's coin landed heads.

Note too that even in cases where it's difficult to assign particular credences to the propositions expressed by fictionally indeterminate claims, often we can still make comparative likelihood judgments about them. For example, although it's difficult to say what likelihood ought to be assigned to claims like (8c), ('Ignatius Reilly has tried reading *The Critique of Pure Reason*') and (8d) ('Elizabeth Bennett eventually dies in her 70s'), it is to our ears perfectly felicitous to say things like:

(16) a. ✓ It is more likely that Ignatius Reilly has tried reading *The Critique of Pure Reason* than that he has tried reading *Fact, Fiction, and Forecast.*

b. ✓ It is more likely that Elizabeth Bennett eventually dies in her 70s than that she lives to be 110.

6.4 Inquisitives

Next we turn to inquisitive attitudes. We agree with others in finding it reasonable to wonder or be curious about indeterminate fictional claims:

(17) a. ✓ I wonder whether Bud's coin landed heads.

b. ✓ I wonder whether Deckard is a replicant.[27]

(18) a. ✓ I'm curious how Bud's coin landed.

b. ✓ I'm curious about what day of the week Holmes was born on.

6.5 Wishing

Finally, we turn to wishing. Interestingly, indeterminate fictional claims seems to be uniformly *un*acceptable under 'wish':

(19) a. ✗ I wish that Bud's coin had landed heads.

b. ✗ I wish that Bud's coin had landed something other than heads.

[27] Williams & Woodward (2019, p. 6) say that the claim that Deckard is a replicant 'allow[s] for speculation and wonder'.

This is not because there's some general prohibition on having bouletic attitudes toward claims about what happens in a work of fiction. It is perfectly felicitous to report oneself as wishing that a *principal* fictional truth were otherwise not true:

(20) ✓I wish that Bud's coin hadn't been flipped at all.

Instead, the infelicity of (19a) and (19b) seems to be due to the fact that you can only (rationally) wish that p if you are sure that $\neg p$.[28] I'm not sure how Bud's coin landed, so I can't be sure it landed heads or be sure that it landed something other than heads, so neither (19a) nor (19b) can be true if I am rational.

6.6 Summary

Our observations from this section are summarized in the following table:

Attitude	FICTION
Know	✗
Sure	✗
Think	✓
Doubt	✓
Credence	✓
Wonder	✓
Wish	✗

We will now argue that the interpretation of the modal operator $\Box_f$ given by the standard analysis is incapable of predicting this pattern of judgments. That will set us up for the presentation of our positive account in §8.

7 Against the standard analysis

7.1 The core problem

The standard analysis identifies fictional truth with principal fictional truth. Thus, Frege puzzles aside, the standard analysis implies that one's attitudes toward the proposition that a fictional claim is true in f are nothing over and above one's attitudes toward the proposition that that claim is a principal truth of f. This is just to say that the standard analysis validates:

[28] See, for example, Heim (1992); von Fintel (1999); Blumberg (2018, forthcoming); Grano & Phillips-Brown (fc). Note that this constraint on (rational) wishing is normally articulated in terms of "belief", though it's clear in context of these discussions that the constraint is meant to be interpreted as a kind of strong belief, akin to being sure.

Attitude Equivalence S Φs $\neg\Box_f \mathsf{p}$ iff S Φs P-Fic$_f$ $\not\subseteq \mathsf{p}$

But Attitude Equivalence mischaracterizes the data in two directions. First, it implies that a number of intuitively false attitude reports are in fact true; and second, it implies that a number of intuitively true attitude reports are in fact false.

Starting with the first sort of mischaracterization, consider again (10b):

(10b) ✗ I know whether Bud's coin landed heads.

On the modal approach, the logical form of the complement of (10b) is given by $\Box_{JAG}$Bud's coin landed heads. So (10b) is true if and only if you know whether $\Box_{JAG}$Bud's coin landed heads. For any p, you know whether p if and only if you know that p or you know that $\neg p$ (Uegaki, 2015). Thus, (10b) is true if and only if you know that $\Box_{JAG}$Bud's coin landed heads or you know that $\neg\Box_{JAG}$Bud's coin landed heads. But given Attitude Equivalence, you *do* know that $\neg\Box_{JAG}$Bud's coin landed heads—i.e., that it's not true in JAG that Bud's coin landed heads. After all, you know that this claim is not among the principal truths. Thus, the standard analysis predicts (10b) is true, contrary to the appearances.

The problem quickly spreads. Anyone who is sure that it's not a principal truth of JAG that Bud's coin landed heads is sure that it's not true that Bud's coin landed heads. This implies that (11b) ('I'm not sure whether Bud's coin landed heads') is false, contrary to intuitive judgment. Likewise, since you know (and are sure) that it's not true that Bud's coin landed heads, you must doubt that it did, and there is no barrier to your wishing that it had. Thus, the standard analysis struggles to explain the infelicity of (14a) ('I doubt that Bud's coin landed heads') and (19a) ('I wish that Bud's coin had landed heads').

The standard analysis also predicts that various intuitively true attitude reports are false. For instance, I think (and indeed know) that it's not a principal truth of the version of JAG in which Bud's coin is biased 2:1 in favor of heads that Bud's biased coin landed heads. By Attitude Equivalence, it follows that I think that it's not true that Bud's biased coin landed heads. But then reports such as (13) ('I think that Bud's biased coin landed heads') cannot be true.

Similarly, Attitude Equivalence implies that one's credence that a fictional claim is true is identical to one's credence that that claim is among the principal truths of the relevant fiction. So anyone who assigns credence 1 to the claim that it's not a principal fictional truth of JAG that Bud's coin landed heads must assign credence 0 to the claim that Bud's coin landed heads. But this runs afoul of the intuitive judgments about (15a) ('It's .5

likely that Bud's coin landed heads'). *Mutatis mutandis* for the comparative likelihood judgments like (16): on the standard analysis, any two claims that are equally (un)likely to be principal fictional truths are equally (un)likely to be fictional truths simpliciter. The standard analysis obliterates intuitive probabilistic distinctions between indeterminate fictional claims.

Finally, assuming that a rational agent can only wonder or be curious about whether p when they aren't sure whether p (and certainly not when they *know for sure* whether p), it follows immediately that rational agents cannot wonder or be curious about indeterminate fictional claims. When one knows that the principal truths fail to settle the matter, there's nothing left to wonder or be curious about. This prediction is in stark contrast with the intuitive judgments about reports like (17) and (18).

We conclude that the standard analysis is in a bad way. The identification of fictional truth with principal fictional truth may seem reasonable enough when intuitions about fictional claims are considered in isolation. But when such claims are embedded under attitude verbs and probability operators, the predictions of the standard analysis become untenable. Knowing that a claim is not a principal fictional truth is one thing; knowing that it is not a fictional truth simpliciter is another.

7.2 Fictional indeterminacy as undefinedness?

Could some of these problems be solved if one maintained that fictionally indeterminate claims are not necessarily false, but are instead merely *undefined*?

The answer to this question depends on what exactly is entailed by a claim's being "undefined". Perhaps undefinedness is to be understood as a kind of presupposition failure, or as involving a third truth-value. We're happy to let proponents of the view fill in the details how they like. If it's to be an improvement on the standard analysis, the status of undefinedness needs to have the right sorts of connections to judgments of infelicity. In particular, it needs to be that a claim's being known to be undefined makes an assertion of that claim infelicitous:

Assertability If S knows that p is undefined, then S should not assert p.

For without this connection, we'd lack any explanation of the infelicity of bare assertions of fictional claims that are known to be indeterminate, like (9a) and (9b).

Similarly, if a proponent of this view wants to explain why reports like (10b) ('I know whether Bud's coin landed heads') are infelicitous, it needs to be

that undefinedness *projects* under attitude verbs:[29]

Attitude Projection If S knows that p is undefined, then S $\Phi\mathsf{s}$ p is undefined.

This is a familiar idea.[30] For example, it is widely assumed that a sentence like 'The King of France is happy' is defined only if there exists a King of France. It is common knowledge that there is no such individual. So consider the following reports:

(21) a. ✗ I know whether the King of France is happy.
b. ✗ I'm sure that the King of France is happy.
c. ✗ I think that the King of France is happy.

These examples are robustly infelicitous, which is exactly what is predicted by the combination of Assertability and Attitude Projection.

Given these principles, those who maintain that fictionally indeterminate sentences are undefined can respond to the overgeneration concerns raised above. For instance, $\Box_{JAG}$Bud's coin landed heads is known to be indeterminate and thus, by hypothesis, undefined. So, given Attitude Projection, I know whether $\Box_{JAG}$Bud's coin landed heads is undefined, and thus, given Assertability, (10b) is unassertable. Similar points go through for each of the other infelicitous examples.

However, the view still undergenerates. This is because a number of attitude reports that embed fictionally indeterminate claims are perfectly felicitous: for example, certain of the thinking and doubting reports, as well as reports about subjective probability, wonder, and curiosity. Yet given Attitude Projection all such reports are undefined and so, by Assertability, unassertable.

To get around this problem, Attitude Projection must be revised to say that the known undefinedness of the complement entails undefinedness of the attitude report as a whole—but only for *certain* attitudes. For example: if p is undefined, then so is S knows whether p, but not S wonders whether p.

This strikes us as uncomfortably *ad hoc*. But there is also a more serious challenge here, which is that with respect to more familiar kinds of undefinedness—say the (known) undefinedness of 'The King of France is happy'—there is no such variation in how these claims embed under attitude verbs. All such embeddings are equally infelicitous:[31]

[29]Note that one could have a view on which an attitude report with an undefined complement clause is *false* rather than undefined. However, such a view would be subject to exactly the kinds of worries we are about to raise for Attitude Projection as stated.

[30]See, e.g., Heim (1983, 1992); Schlenker (2009); Sudo (2014); Blumberg & Goldstein (forthcoming).

[31]As we will observe in §10, similar points arise for attitude reports that embed claims that are knowably vague.

(22) a. ✗ I doubt that the King of France is happy.
b. ✗ There is some chance that the King of France is happy.
c. ✗ I wonder whether The King of France is happy.
d. ✗ I'm curious about whether The King of France is happy.

So why should the projection properties of undefined fictional claims project pattern so differently?

Lastly, we note that there is a general challenge for any theory of fictional truth that is (a) designed to deliver the relevant patterns in the attitude judgments surveyed above, while (b) being committed to a trivalent conception of undefinedness:

Trivalence If p is undefined, then p is neither true nor false.

And this is that any such view must predict that abominable conjunctions of the following form can in fact be uttered truly:

(23) a. ✗ I know whether p is true, but I don't know whether p.
b. ✗ I wonder whether p, but I don't wonder whether p is true.

After all, if I know that p is fictionally indeterminate, then I know whether p is true (it isn't), and can't wonder whether p is true (because I know it's not). But as we've seen, the attitude data suggest that all this is no barrier to my not knowing whether p, or to my rationally wondering whether p. So both (23a) and (23b) would have to have true instances. It is hard to see how this could be so.

In light of these considerations, we think that our observations involving attitudes count strongly against the standard analysis—whether it treats fictionally indeterminate claims as false or as undefined.

8 A positive account

8.1 Scopelessness

The lesson of the failures of the standard analysis is that the correct account of fictional truth will block the inference from P-Fic$_f \not\subseteq \mathsf{p}$ to $\neg\Box_f\mathsf{p}$. Accordingly, the set of worlds compatible with what's true in a work of fiction needs to be smaller than the set of worlds compatible with the principal truths of that work. This means that Fic$_f$ is not just a subset of P-Fic$_f$, but a *strict* subset of it.

In fact, we can place even stronger constraints on Fic$_f$. To help motivate them, consider an example such as (24):

(24) ✗ I know that Bud's coin failed to land heads.

Presumably the only available logical form for this sentence is equivalent to one where negation takes narrow scope with respect to $\Box_{JAG}$ (we use parentheses simply to aid readability):[32]

(25) I know that: $\Box_{JAG}\neg$(Bud's coin landed heads)

But this raises a puzzle, for the following argument appears to be valid:

(26) I don't know that Bud's coin landed heads. And I don't know that it failed to land heads. So I don't know whether Bud's coin landed heads.

This inference strike us as unimpeachable. However, you don't know whether $\Box_{JAG}$Bud's coin landed heads if and only if you neither know that $\Box_{JAG}$Bud's coin landed heads nor know that $\neg\Box_{JAG}$Bud's coin landed heads. But all the premises establish is that you fail to know that $\Box_{JAG}$Bud's coin landed heads and, given (25), that you fail to know that $\Box_{JAG}\neg$Bud's coin landed heads. What this means is that failing to know $\Box_{JAG}\neg$Bud's coin landed heads implies failing to know $\neg\Box_{JAG}$Bud's coin landed heads. Thus, (27) must imply (25):

(27) I know that: $\neg\Box_{JAG}$Bud's coin landed heads

In fact, (25) and (27) are equivalent. For so long as Fic_{JAG} is non-empty, it follows from the modal analysis itself that (25) implies (27): if you know that every world in Fic_{JAG} is a $\neg$p-world, then you know that it's not true that every world in Fic_{JAG} is a p-world.

These points suggest that the correct analysis of $\Box_f$ will validate the following:

Scopeless Inference S Φs $\neg\Box_f$p iff S Φs $\Box_f\neg$p

Which is in turn tantamount to requiring that $\Box_f$ commutes with negation:

Scopelessness $\neg\Box_f\text{p} \leftrightarrow \Box_f\neg\text{p}$

We will now argue that Scopelessness is the key to explaining the patterns from §6.

[32] Although 'Bud's coin failed to land heads' is *semantically* equivalent to 'Bud's coin did not land heads', the former is not *syntactically* equivalent to the latter in the sense that the former does not feature a scope-taking negation operator.

8.2 Explaining the data

For now let us continue to assume that if p is fictionally indeterminate in f—i.e., if neither p nor $\neg\mathsf{p}$ is a principal truth of f—then one can neither know that $\Box_f\mathsf{p}$ nor know that $\Box_f\neg\mathsf{p}$. (We'll suggest an explanation for this in §9.)

Given this assumption, it follows that you can't know that $\Box_{JAG}$Bud's coin landed heads or that $\Box_{JAG}\neg$Bud's coin landed heads. So (10a) ('I know that Bud's coin landed heads') is false. (10b) ('I know whether Bud's coin landed heads') is true if and only if either (i) you know that $\Box_{JAG}$Bud's coin landed heads—which we just saw is not the case—or (ii) you know that $\neg\Box_{JAG}$Bud's coin landed heads. Given Scopelessness, (ii) holds if and only if you know that $\Box_{JAG}\neg$Bud's coin landed heads. But (24) is false, so (ii) does not hold either. So (10b) must be false, as desired.

To account for the surety reports, we'll once again invoke our earlier principle: if you know that you don't know whether a certain claim is true, then, if you are rational, you are not sure whether that claim is true. We just gave an argument for why you don't know whether Bud's coin landed heads. But notice that the reasoning there is entirely transparent: anyone who knows that it's fictionally indeterminate how Bud's coin landed can easily know that they don't know whether Bud's coin landed heads, by reasoning roughly in the way we just did. So (10b) isn't just false, it's knowably false. So (11b) ('I'm not sure whether Bud's coin landed heads') must be true. And this in combination with Scopelessness entails that both (11a) ('I'm sure that Bud's coin landed heads') and (28) are false:

(28) ✗ I'm sure that Bud's coin didn't land heads.

Given that you neither know nor are sure whether Bud's coin landed heads, there is no barrier to you wondering or being curious about whether Bud's coin landed heads. This accounts for inquisitive reports like (17) and (18). Likewise, it also explains the infelicity of the wish reports: if you wish that Bud's coin had landed heads, then you must be sure that it's not true that Bud's coin landed heads. You're not sure of this, so you can't wish it had been otherwise—hence the infelicity of (19).

Accounting for the weak doxastic attitude and subjective probability reports requires our second assumption (or really set of assumptions), which is about how credences are modeled. None of these ideas are particularly controversial, but still we think it's worth stating them explicitly.

We assume that an agent's state of surety can be represented by a set Dox_S, understood as the set of worlds compatible with what S is sure of.[33] S is

[33]Strictly speaking, Dox_S should be parametrized to a world. We leave this implicit in what follows.

sure that p iff $\forall w' \in \text{Dox}_S$: p is true at w' (Hintikka, 1962). We assume that an agent's credences can be represented by a subjective probability function, C_S.[34] We also assume that C_S is uniform over Dox_S. Thus, if every world in Dox_S is a p-world (i.e., S is sure that p), then S's credence that p is 1; if exactly half the worlds in Dox_S are p-worlds, then S's credence that p is .5; if S finds it more likely that p than that q, then there are more p-worlds in Dox_S than q-worlds; and so on.

Given these assumptions, (15a) ('It's .5 likely that Bud's coin landed heads') is true just in case exactly half the worlds in Dox_S are worlds in which $\Box_{JAG}$Bud's coin landed heads is true, while (15b) ('It's twice as likely that Bud's biased coin landed heads as it is that it landed tails') is true just in case there are twice as many worlds in Dox_S in which $\Box_{JAG}$Bud's biased coin landed heads is true as there are worlds in which $\Box_{JAG}$Bud's biased coin landed tails is true. We don't have a direct argument for the conclusion that Dox_S *must* have this structure. But if you are rational, then plausibly you will be doxastically indifferent to the various possibilities compatible with the principal truths of *JAG*. And such indifference should provide Dox_S with the relevant structure to witness the truth of (15a) and (15b).

Finally, the explanation of the data concerning the thinking and doubting reports falls out of the explanation of the rest of the data. If you are rational and you find p more likely than not, then you can think that p; else you cannot. You find it exactly as likely that Bud's coin landed heads as that it didn't, so (12) ('I think that Bud's coin landed heads') is false. But you do find it more likely than not that Bud's biased coin landed heads, so (13) ('I think that Bud's biased coin landed heads) is true. And doubting that p is just thinking that $\neg\mathsf{p}$, hence why (14a) ('I doubt that Bud's coin landed heads') is false while (14b) ('I doubt that Bud's biased coin landed tails') is true.

We have shown how Scopelessness in combination with some other fairly uncontroversial assumptions about the workings of the attitudes allows us to capture the full range of patterns in the attitude data surveyed above. We take this to constitute strong abductive evidence in favor of theories of $\Box_f$ that validate Scopelessness, and against those like the standard analysis that invalidate it. We now want to turn to exploring some of the logical implications of Scopelessness, chief among them what we will call the principle of "fictional excluded middle".

[34] C_S is a function from $\mathcal{A}$, an algebra of subsets of W, to the unit interval. $C_S(\text{Dox}_S) = 1$; and for disjoint $\mathsf{p}, \mathsf{q} \in \mathcal{A}$, $C_S(\mathsf{p} \cup \mathsf{q}) = C_S(\mathsf{p}) + C_S(\mathsf{q})$.

8.3 The logic of fiction

Given that $\Box_f$ quantifies universally over the worlds in Fic_f, Scopelessness imposes strong constraints on the cardinality of Fic_f. In fact, it entails that it is a singleton:[35]

Uniqueness $|\mathrm{Fic}_f| = 1$

It also entails:[36]

FEM $\Box_f \mathsf{p} \vee \Box_f \neg \mathsf{p}$

From FEM it follows that exactly one of (9a) ('Bud's coin landed heads') or (9b) ('Bud's coin landed tails') is true. So too for Holmes' hair: either (1a) is true and he has an odd number of hairs on his head, or (1b) is true and he has an even number. There is an exact list of all the books Ignatius Reilly has tried to read in his life; it's an open question whether *The Critique of Pure Reason* is on it. Perhaps Hamlet wasn't born on a Tuesday; but if so, then it's true that he was born on a non-Tuesday. There is some precise number of years Elizabeth Bennett lives to be—we hope it's at least in the 70s. And the full extent of Raskolnikov's dermatological maladies may never be known, but the facts are nonetheless out there. Fictional indeterminacy is no barrier to truth. This is because for any given work of fiction, there is a single world that determines what's true in it, and that world is as rich in detail as reality itself.

We suspect that some will find the picture of fictional truth we have argued for difficult to accept. Indeed, we acknowledge that our account raises many questions that need to be addressed. In the remaining sections of the paper, we focus on those that we deem to be most urgent. We begin with two in particular. First, what could possibly determine which world gets to be *the* world of a fiction? And second, we've been assuming from the outset that fictionally indeterminate claims are unknowable—but why should this be? We will now argue that the answers to both of these questions plausibly have something to do with counterfactuals.

9 Counterfactual indeterminacy

Suppose Flippy is a fair coin that existed briefly in the year 2010 but was never flipped. Now consider the following counterfactual conditionals:

[35] Suppose Fic_f contained at least two worlds w_1 and w_2. Since these worlds are distinct, there would have to be some p such that p is true at w_1, and $\neg\mathsf{p}$ is true at w_2. But then we'd have (i) that $\neg\Box\mathsf{p}$ is true, since not every world in Fic_f is a p-world; but also (ii) that $\neg\Box\neg\mathsf{p}$ is true, since not every world in Fic_f is a $\neg\mathsf{p}$-world either. From there one application of double negation elimination yields a violation of Scopelessness.

[36] The left-to-right direction of Scopelessness yields $\neg\Box_f\mathsf{p} \rightarrow \Box_f\neg\mathsf{p}$, which is equivalent to $\neg\neg\Box_f\mathsf{p} \vee \Box_f\neg\mathsf{p}$ and thus FEM given double negation elimination.

(29) a. If Flippy had been flipped once, it would have landed heads.
b. If Flippy had been flipped once, it would have landed tails.

It's hard to see which facts about the physical world could explain why one rather than the other of these claims is true. As such, the question of how Flippy would have landed had it been flipped seems to admit no determinately true answer. In this sense (29a) and (29b) are *counterfactually indeterminate*.

Here are two further examples of indeterminate counterfactuals:[37]

(30) a. If Ann (an only child) had had a sibling, that sibling would have been exactly six feet tall.
b. If Bizet and Verdi had been compatriots, Bizet would have been Italian (Lewis, 1973; Stalnaker, 1980).

It seems obvious that one cannot know that an indeterminate counterfactual expresses a truth. But how do indeterminate counterfactuals pattern with attitudes more generally?

9.1 Attitudes towards indeterminate counterfactuals

For concreteness we'll focus on attitudes toward the Flippy counterfactuals, but we believe the judgments go through just as well with attitudes towards counterfactuals like (30a) and (30b). The relevant judgments, stated in one fell swoop, are as follows:

(31) Knowledge
a. ✗ I know that Flippy would have landed heads, had it been flipped once.
b. ✗ I know whether Flippy would have landed heads, had it been flipped once.

(32) Surety
a. ✗ I'm sure that Flippy would have landed heads, had it been flipped once.
b. ✗ I'm sure that Flippy wouldn't have landed heads, had it been flipped once.
c. ✓ I'm not sure whether Flippy would have landed heads, had it been flipped once.

[37] For further examples and discussion, see Lewis (1973), Stalnaker (1980), Schulz (2014, 2017), Khoo (2021), Goodman (manuscript).

(33) Thinking

a. ✗ I think that if Flippy had been flipped once, it would have landed heads.

b. ✓ I think that if a version of Flippy that was biased 2:1 in favor of heads had been flipped once, it would have landed heads.

(34) Doubting

a. ✗ I doubt that if Flippy had been flipped once, it would have landed heads.

b. ✓ I doubt that if a version of Flippy that was biased 2:1 in favor of heads had been flipped once, it would have landed tails.

(35) Credence[38]

a. ✓ It's .5 likely that Flippy would have landed heads, had it been flipped once.

b. ✓ It's twice as likely that biased Flippy would have landed heads as it is that it would have landed tails, had it been flipped once.

(36) Inquisitives[39]

a. ✓ I wonder whether Flippy would have landed heads, had it been flipped once.

b. ✓ I curious about whether Flippy would have landed heads, had it been flipped once.

(37) Wishing

a. ✗ I wish that Flippy wouldn't have landed heads, if it'd been flipped once.

b. ✗ I wish that Flippy would have landed something other than heads, had it been flipped once.

The parallels between the two sets of attitude data is striking. Combining these observations with those from §6, we get the following table:

[38] Judgments about credences in indeterminate counterfactuals has been discussed extensively in the literature. See, e.g., Moss (2013); Schulz (2014, 2017); Mandelkern (2018) for a representative sample.

[39] Moss (2013) notes that wonder reports felicitously embed indeterminate conditionals.

Attitude	FICTION	CONDITIONALS
Know	✗	✗
Sure	✗	✗
Think	✓	✓
Doubt	✓	✓
Credence	✓	✓
Wonder	✓	✓
Wish	✗	✗

9.2 Selection function semantics

So counterfactually indeterminate claims seem to embed under attitudes in essentially the same way fictionally indeterminate claims do. This is significant, since theorists working on the semantics of counterfactuals (and conditionals more generally) have tried to give semantics for the counterfactual that explain these sorts of embedding patterns.[40] And by far the most popular approach for theorists driven by such concerns involves a *selection function* semantics for the conditional.

There are several ways of developing a selection function semantics.[41] We will present a fairly simple account inspired by Stalnaker (1968) in order to get the main ideas across.

Letting '>' abbreviate the counterfactual conditional, the basic idea is that $\mathsf{p} > \mathsf{q}$ is true just in case every member of a set of p-worlds is a q-world. This set of p-worlds is in turn determined by a selection function $\mathfrak{s}$. This function takes a proposition and a world as arguments, and yields a set of worlds as output. The semantics can be represented as follows:

> **Selection Function Analysis for Counterfactuals**
>
> $\mathsf{p} > \mathsf{q}$ is true at w iff: $\forall w' \in \mathfrak{s}(\mathsf{p}, w)$: q is true at w'.

There are a number of standard assumptions about the workings of the selection function $\mathfrak{s}$ that theorists tend to adopt.[42] But what is distinctive

[40] In fact, theorists working on conditionals tend only to focus on the way indeterminate conditionals embed under credence and likelihood judgments. Far less attention has been paid to the way fictional content embeds under attitudes in general.

[41] See, e.g., Stalnaker (1968, 1980); Schulz (2014, 2017); Bacon (2015); Mandelkern (2018); Santorio (2022); Schultheis (forthcoming).

[42] For example, it is standardly assumed that $\mathfrak{s}(\mathsf{p}, w) \subseteq \mathsf{p}$ (i.e., that the p-selected set contains only p-worlds). This helps guarantee the validity of $\mathsf{p} > \mathsf{p}$. Likewise, it is standardly assumed that if $w \in \mathsf{p}$, then $\mathfrak{s}(\mathsf{p}, w) = \{w\}$ (i.e., that if the world at which the counterfactual is assessed is one that makes the antecedent true, then the selected world is the world of assessment itself). This helps secure the validity of inference rules like modus ponens. See Starr (2022) for a discussion of the various constraints one can impose on selection functions, and how this impacts the logic of conditionals.

about the Stalnakerian approach is the assumption that $\mathfrak{s}(p, w)$ contains *exactly* one world:

C-Uniqueness $|\mathfrak{s}(\mathsf{p}, w)| = 1$

Just as Uniqueness entails the validity of FEM, C-Uniqueness entails the validity of the principle known as "Conditional Excluded Middle" (CEM):[43]

CEM $(\mathsf{p} > \mathsf{q}) \vee (\mathsf{p} > \neg\mathsf{q})$

CEM implies that one of (29a) ('If Flippy had been flipped once, then it would have landed heads') or (29b) ('If Flippy had been flipped once, then it would have landed tails') is true.[44] Thus, the selection semantics predicts that counterfactual indeterminacy is no barrier to counterfactual truth.

Notably, this theory also validates:

C-Scopelessness $\neg(\mathsf{p} > \mathsf{q}) \leftrightarrow (\mathsf{p} > \neg\mathsf{q})$

And given (i) C-Scopelessness and (ii) the assumption that if $\mathsf{p} > \mathsf{q}$ is counterfactually indeterminate, then one can neither know that $\mathsf{p} > \mathsf{q}$ nor know that $\mathsf{p} > \neg\mathsf{q}$, then the patterns surveyed above can be captured in a way that is perfectly analogous to the account we gave of the patterns involving fiction from §6.

9.3 The Reality Principle

This sets us up to answer the first question we asked at the end of the previous section: of all the worlds compatible with the principal truths of a given work of fiction, why does exactly one of them get to be *the* world of that fiction?

We suggest that an answer can be given in terms of the *Reality Principle*:[45]

The Reality Principle $\Box_{f,w}\mathsf{p} \leftrightarrow (\text{P-Fic}_{f,w} > \mathsf{p})$

[43] For further discussion of the motivations for CEM, see, e.g., Stalnaker (1980); Williams (2010); Cariani & Goldstein (2018); Mandelkern (2018).

[44] Here we assume that the only way for Flippy to fail to land heads is for it to land tails.

[45] Variants of the Reality Principle were initially proposed and tentatively defended by Lewis (1978) and Walton (1990). Note that our version is not exactly like any of the existing versions, since ours appeals directly to the notion of a principal fictional truth. As we understand Lewis' project, by contrast, he is using an analog of the Reality Principle to generate the principal fictional truths from the "explicit" fictional truths.

What the Reality Principle says, in words, is that a claim is a fictional truth of f iff the following counterfactual is true: if all the principal truths of f were externally true—i.e., true in reality, not just in f—then that claim would be externally true too. According to the Reality Principle, fictional truth is just a kind of counterfactual truth.

We lack the space to argue for the Reality Principle in detail here.[46] But the striking similarities in the attitude data for fictional and counterfactual indeterminacy provide strong abductive support for the hypothesis that fictional truth is a species of counterfactual truth. Indeed, we see no better way of explaining these similarities.[47]

What's of primary interest to us is the way The Reality Principle interacts with a selection function semantics for the counterfactual, on the one hand, and our analysis of fictional truth, on the other. For notice that the Reality Principle says that a fictional claim is true if and only if a certain counterfactual is true. The selection function semantics tells us that a certain counterfactual is true if and only if the *unique* world selected by the selection function is a world in which that counterfactual's consequent is true. So, on the assumption of The Reality Principle and the selection function semantics, one can explain why for some particular world w, $\text{Fic}_f = \{w\}$. And that is because w is the selected world—i.e., the unique world that would accurately represent how things are, were all the principal claims of f really true.

10 Arbitrariness and vagueness

Is this an entirely satisfying explanation? We suspect some will think not. Some might argue that all we have done is reduce one kind of indeterminacy to another. We claim that for any given work of fiction, there's a unique world compatible with everything that's true in that fiction—namely the world that would be the real one were the principal fictional claims an accurate description of reality. But this explanation involves appealing to an

[46] For criticisms, see, e.g., Currie (1990); Phillips (1999); Proudfoot (2006); Woodward (2011); Friend (2017); Badura & Berto (2019). But see also Franzén (2021) for a recent defense of the principle.

[47] Interestingly, Stalnaker (1980, p. 95) explicitly draws a parallel between his selection function semantics for conditionals and truth in fiction:

> It is not surprising, from the point of view of the analysis I am defending, that the possible situations determined by the antecedents of counterfactual conditionals are like the imaginary worlds created by writers of fiction. In both cases, one purports to represent and describe a unique determinate possible world, even though one never really succeeds in doing so.

indeterminate counterfactual. Consequently, some may wonder why a particular world gets to be selected by the selection function, rather than one of the many other worlds compatible with the principal fictional truths.

Sometimes facts about intuitive similarity between worlds will be of some help in fleshing out the story: if one of two antecedent worlds is sufficiently more similar to the world of assessment than the other, then we should expect the selection function to select that world if it's selecting either (Stalnaker, 1968, 1980). However, so long as 'similarity' is meant to to be understood independently of the workings of the selection function, it's unclear that this kind of explanation can work in the general case. There seems to be no intuitive sense in which worlds where Flippy lands heads after being flipped are more similar to actuality than worlds in which Flippy lands tails after being flipped, or in which worlds where Holmes has an even number of hairs on his head are closer to actuality than worlds in which he has an odd number.

So perhaps the workings of the selection function are fundamentally *arbitrary*, at least modulo distinctions in intuitive similarity: when choosing between two sufficiently similar worlds compatible with the antecedent of a given counterfactual, the selection function selects one at random, so to speak.[48]

Supposing this is right, we have a straightforward answer to the question of why fictionally indeterminate claims are beyond our epistemic reach. And the answer is that if p is not a principal fictional truth of f, then whether p is true in f is determined by an arbitrary selection process.[49] Unless one has "direct" evidence as to the result of this process—the word of an oracle, say—one's evidence won't be able to rule out any of the possible results. Thus, it is not possible to know whether p is true in f.

There are many who find this conception of the workings of counterfactuals intolerably counterintuitive. We admit it takes some getting used to, though given the striking patterns in attitude data we think it is well worth taking seriously as a hypothesis about how the counterfactual (and thus fiction) functions. In any case, some proponents of the selection function semantics have tried to appeal to vagueness to help soften the blow. They claim that although any given selection function does select a single unique world, there are many selection functions compatible with our ordinary use of counterfactuals, and in any given context it will generally be a vague matter which of these selection functions is operative. By extension, indeterminate coun-

[48] Views of this sort have been defended by Schulz (2014, 2017) and Bacon (2015). See especially Bacon's discussion of his deflationary notion of "random" selection (p. 146), as well as fn.49 below.

[49] One could make this process more explicit by appealing to the so-called "epsilon operator", which is discussed in detail by Schulz (2017, Ch. 6).

terfactuals are vague (Stalnaker, 1980, 89-90), and so too are indeterminate fictional claims, given the Reality Principle.

There are choice points as to how to develop a theory of vagueness. But regardless of the precise shape of this response, we argue that it should not be taken up by proponents of FEM. This is because the indeterminacy generated by canonically vague sentences patterns quite differently from counterfactual and fictional indeterminacy.

First, indeterminacy due to vagueness embeds differently under attitudes. To see this, suppose that we have a long sequence of people such that the first person in the series is clearly bald, while the last person is clearly not bald. Also suppose that adjacent members in the sequence differ by exactly one strand of hair, and that George is some middle-ish member of the sequence. Then (38) is vague (and thereby indeterminate):

(38) George is bald.

Now, as has been widely observed, vague sentences are uniformly unacceptable under propositional attitudes:[50]

(39)
a. ✗ I know whether George is bald.
b. ✗ I'm sure that George is bald.
c. ✗ I think that George is bald.
d. ✗ I doubt that George is bald.
e. ✗ I find it $.x$ likely that George is bald.
f. ✗ I wonder whether George is bald.
g. i. ✗ I wish that George was bald.
   ii. ✗ I wish that George wasn't bald.

These data contrast quite strikingly with our observations from §6 and §9.1:

Attitude	Fiction	Conditionals	Vagueness
Know	✗	✗	✗
Sure	✗	✗	✗
Think	✓	✓	✗
Doubt	✓	✓	✗
Credence	✓	✓	✗
Wonder	✓	✓	✗
Wish	✗	✗	✗

[50]It is widely accepted that vague sentences are unacceptable under epistemics and doxastics, and Field (2010) notes that vague content also does not felicitously embed under 'wonder'. But these judgments are not entirely uncontroversial: Bacon (2018), for example, argues that credence and comparative confidence judgments in vague claims are perfectly acceptable, while Spencer (forthcoming) argues that certain vague sentences embed felicitously under bouletics like 'hope'. We demur, but we lack the space to engage with these arguments here.

Thus, attempts to assimilate fictional and counterfactual indeterminacy with vague indeterminacy will fail to capture the distinctive way in which fictional and counterfactual claims embed under attitudes.

Second, many philosophers maintain that vague indeterminacy arises due to some kind of semantic indecision. One piece of evidence for this comes from the ways in which one can respond to questions about vague matters:

(40) [You don't know anything about the state of Bill's scalp, but I know that he is borderline bald.]
  a. You: Is Bill bald?
    i. ✓ Me: Yes and no.
    ii. ✓ Me: In some sense yes, in some sense no.

But as [redacted] (p.c.) has pointed out, one cannot give analogous answers to questions about indeterminate conditionals:

(41) [You know that Flippy wasn't flipped, but you don't anything about the weighting of the coin. I know that Flippy is fair.]
  a. You: If Flippy had been flipped, would it have landed heads?
    i. ✗ Me: Yes and no.
    ii. ✗ Me: In some sense yes, in some sense no.

Mutatis mutandis for indeterminate fictional claims. These sorts of replies are utterly bizarre when given as answers to the question 'Does Holmes has an even number of hairs on his head?'.

In short: the kind of indeterminacy involved with vagueness seems quite different from the kind of indeterminacy involved with counterfactuals and fictional claims. The former kind seems to have something to do with semantic indecision; the latter kinds do not. So we are inclined to reject an account of conditional indeterminacy—and by extension an account of fictional indeterminacy—that assimilates it with vague indeterminacy. Vagueness is no remedy to the seemingly ineliminable arbitrariness involved in determining which world in P-Fic$_f$ gets to be the unique world of f.

11 Further objections

We imagine some readers will be inclined to reject the premises that have led us to our conclusions sooner than they would embrace such surprising sounding claims. Those who feel the pull of this reaction are invited to see

the argument of the paper as fundamentally conditional in nature: *if* we are not systematically mistaken in our judgments about the kinds of attitudes that are appropriate to take toward indeterminate fictional claims, *then* there must be facts of the matter about any meaningful question one can ask about what's true in a work of fiction—facts whose grounds will in many cases be almost entirely arbitrary. We would sooner accept the consequent of this conditional than deny the antecedent. But we must leave a proper defense of our methodological stance for another time.

What we want to do in the remaining parts of the paper is discuss objections to our account of fictional truth that are *not* premised on concerns of arbitrariness and the like. We will consider three: an objection from putatively impossible fictions, an objection from putatively incomplete fictions, and an objection concerning fictional change. We want to stress in advance that we lack the space to give each of them more than a fairly brief treatment. Still, we hope that the discussion will bring out some interesting theoretical upshots of our account of fictional truth.

11.1 Inconsistent fictions

The first objection we consider concerns the possibility of *inconsistent fictions*—cases where for some work of fiction f, both $\Box_f \mathsf{p}$ and $\Box_f \neg \mathsf{p}$ are true.

Here's the problem in schematic form. Our theory of fictional truth says that $\Box_f \mathsf{p}$ is true if and only if $\forall w' \in \text{Fic}_f$: p is true at w'. We assume that for any world w, at most one of p or $\neg \mathsf{p}$ is true at w. It follows immediately that if both $\Box_f \mathsf{p}$ and $\Box_f \neg \mathsf{p}$ are true, then Fic_f is the empty set. But if Fic_f is the empty set, then it follows trivially that no matter what p is, $\Box_f \mathsf{p}$ is true. Our account thus predicts that an inconsistent fiction is a fiction in which every claim is true. By extension, as far as fictional truth is concerned, there is no distinguishing inconsistent works of fiction.[51]

This prediction looks problematic if one believes in the possibility of works of fiction that are merely "locally" inconsistent. For example, a crucial plot point of Priest's 1997 'Sylvan's Box' is the discovery of a box that is described as being simultaneously (and without equivocation) empty and non-empty. Given that the box is so-described, it's plausible not only that $\Box_{SB}$The Box is empty and $\Box_{SB}\neg$The box is empty are both true, but that these claims report *principal* fictional truths. Nonetheless, Priest (1997, pp. 580–581) insists that 'Sylvan's Box' is a "coherent story" in which "not everything happens". For instance, it is *not* meant to be a truth of 'Sylvan's Box' that at the end of the story the box is shot off to the moon.

[51]Note that this implication arises for *any* version of the modal analysis, not just the kinds that validate FEM.

The problem of squaring an account of fictional truth with the possibility of (merely locally) inconsistent fictions is an old one.[52] One might take such cases to motivate the use of "impossible" worlds in one's semantics for fictional truth, thereby abandoning the assumption that at any world w, at most one of p or $\neg\mathsf{p}$ is true. Or one might argue that these stories are in fact globally consistent, with the inconsistency arising only in the narrator's *reporting* of the events of the story. Or one might adopt Lewis's 1978 "method of union", representing an inconsistent fiction f as a collection of fragments of consistent (but mutually incompatible) fictions, with truth in f amounting to truth in any of the fragments.

Following Franzén (2021), our preferred approach to the problem is to draw on the Reality Principle. For note that inconsistent fictions are (putatively) *determinately* inconsistent: in each case the inconsistency arises from what is explicitly said or depicted in the fiction. The Reality Principle says that p is true in f if and only if, were all the principal fictional truths of f (externally) true, p would be (externally) true as well. So given the Reality Principle, the question of whether p is true in an inconsistent fiction is equivalent to the question of whether a certain *counterpossible*—i.e., a counterfactual with an impossible antecedent—is true.

Strikingly, the dialectic concerning the truth-conditions of counterfactuals with impossible antecedents in many ways resembles the dialectic concerning the truth-conditions of fictional claims in inconsistent fictions.[53] Standard accounts of the counterfactual predict that if p is inconsistent, then $\mathsf{p} > \mathsf{q}$ is true no matter what q is, and thus that all counterpossibles are trivially true.

Yet counterpossibles differ in their intuitive truth-value:

(42) a. ✓ If Hobbes had squared the circle, he would have performed a mathematical impossibility.

b. ✗ If Hobbes had squared the circle, he would have assassinated JFK.

And as in the case of inconsistent fictions, there is a question of how to square these intuitive judgments with the otherwise powerful and predictive orthodox semantics for the counterfactual conditional.

Various options have been explored in the literature, many of which resemble the standard treatments of inconsistent fiction.[54] We do not take a stand on which of these options is best, and are happy to adopt whichever approach

[52] For helpful discussion, see, e.g., Lewis (1978); Currie (1990); Byrne (1993); Poidevin (1995); Priest (1997); Hanley (2004); Woods (2018); Nolan (2021); Kim (forthcoming).

[53] For helpful discussion see Kocurek (2021) and the citations therein.

[54] See again Kocurek (2021).

to counterpossibles proves most promising. As such, the question of what's fictionally true in an inconsistent fiction reduces to the question of what's counterfactually true on the supposition of an impossible antecedent, which is where we're happy to leave the problem. We believe any reasonable view here will be consistent with our central thesis: that for any claim p and work of fiction f, there is a fact of the matter about whether p is true in f. After all, the problem raised by inconsistent fictions is that they make too many fictional claims *determinately* true, not that they make too many fictional claims true *simpliciter*.

11.2 Incomplete fictions

A more pointed objection to our account concerns the possibility of putatively incomplete fictions—cases where for some work of fiction f and claim p, it's a principal truth of f that $\neg(\mathsf{p} \vee \neg\mathsf{p})$.[55] One can imagine an author, say Borges, writing a novel called *The "Red" Cube*, in which there is a cube that is described as being "neither scarlet nor not scarlet". If it helps, imagine that this stipulation about the cube's color is meant to be integral to the plot of the story.

Taking the description of the case at face value, it seems to require that $\Box_{RC}\neg(\mathsf{The\ cube\ is\ scarlet} \vee \neg\mathsf{The\ cube\ is\ scarlet})$, which in turn entails both $\neg\Box_{RC}\mathsf{The\ cube\ is\ scarlet}$ and $\neg\Box_{RC}\neg\mathsf{The\ cube\ is\ scarlet}$. This is a direct counterexample to FEM.[56] And accounting for it would require revising the modal analysis so that $\Box_f$ is able to sometimes quantify over worlds that contain truth-value *gaps*—i.e., worlds at which, for some p, neither p nor $\neg\mathsf{p}$ is true. All of the claims we've made about the logic of fictional truth would then have to be restricted to works of fiction f for which none of the worlds in P-Fic$_f$ are incomplete. (We'll see an example of such a restriction momentarily.)

Ought cases like *The "Red" Cube* be taken at face value? We find it hard to say. This is partly because we're not convinced that there really are any works of fiction like *The "Red" Cube*. Or rather: we're not convinced that there are any works of fiction for which it's a *principal* fictional truth that for some p, neither p nor $\neg\mathsf{p}$. The mere fact that an author says 'In the room was a cube that was neither scarlet nor not scarlet' is not itself particularly strong evidence that it's literally true in the relevant work of fiction that there is a cube that is neither scarlet nor not scarlet. (By analogy: the opening line of *A Tale of Two Cities* does not make it an impossible work of fiction.) Indeed, we speculatively conjecture that many authors would

[55] Thanks to [redacted] for discussion here.

[56] It's also a counterexample to the standard analysis, since the standard analysis assumes that fictional worlds are complete.

have a hard time saying what they mean in saying things like 'It was red, but neither scarlet nor not scarlet', and, if pressed, would probably opt for a paraphrase that makes it equivalent to what analytic philosophers would express with claims like: 'It was a borderline shade of scarlet' or 'Its color was an unknowable shade of red'.

That said, we are reticent to rely in any important way on these sorts of hermeneutical speculations. Perhaps a trained philosopher could write a work of fiction like *The "Red" Cube* and really mean it. So suppose that it really is a principal truth of *The "Red" Cube* that there is a cube that is neither scarlet nor not scarlet. What then should we say?

We'll start by noting that truth-value gaps are no more possible than truth-value guts. So given the Reality Principle, fictional claims about *The "Red" Cube* are equivalent to certain kinds of counterpossibles. If counterpossibles are all trivially true, then the *The "Red" Cube* is no counterexample to FEM (since both disjuncts would be true). But if our best theory of counterpossibles requires us to make distinctions between counterfactuals whose antecedents express truth-value gaps and counterfactuals whose antecedents express truth-value gluts—with only the latter kind being trivially true—*if* we must accept all that, then we'd take ourselves to have a genuine counterexample to FEM.

In that case, we would be willing to endorse an account of fictional truth that validates the following weaker principle instead:[57]

Restricted FEM $(\forall \mathsf{p}\colon\ \forall w' \in \text{P-Fic}_f\colon\ w' \in (\mathsf{p} \vee \neg \mathsf{p})) \rightarrow \forall \mathsf{p}\colon\ (\Box_f \mathsf{p} \vee \Box_f \neg \mathsf{p})$

In words: if no truth-value gap is among the principal truths of f, then for any claim, either that claim or its negation is a truth of f.

Though weaker than FEM, Restricted FEM is still quite strong. Few works of fiction explicitly state or depict the presence of truth-value gaps. And we suspect that few authors intend for there to be any either. In fact, we suspect that none of the actually existing works of fiction we've discussed in this paper are ones in which for some p, it's a principal truth of that work that there's no fact of the matter about whether p. So we see no good reason to think that the possibility of works of fiction like *The "Red" Cube* should lead us to deny that for works of fiction like *A Study in Scarlet*, there is a fact of the matter as to the answer to any meaningful question one might ask of it.

11.3 Fictional change

The third and final objection we will consider concerns fictional change. Sometimes an author will start producing a work of fiction, get to a point

[57] Here we make the universal quantification over claims explicit.

that looks like partial or total completion, but then decide later to edit the work in various ways—say by changing certain character details or plot points. This is just to say that authors can revise their fictions. And when they do, the principal fictional truths of the fiction seem to change *non-monotonically*: some claims go from being principal fictional truths to being indeterminate, or even to being principal fictional falsehoods.

To give a working example, Vince Gilligan, the creator of *Breaking Bad*, originally intended for the character Jesse Pinkman to be killed off in a drug deal gone awry near the end of the show's first season.[58] But while filming the series' second episode, Gilligan was so impressed with actor Aaron Paul's performance as Jesse that he decided to revise the script so as to keep the character alive through the first season. Plausibly, then, at the time of the filming of *Breaking Bad*'s pilot, it was a principal fictional truth of *Breaking Bad* that Jesse would die in a drug deal gone awry. But by the time of the filming of the show's third episode, this was no longer the case.

There's an issue here for anyone like us who thinks that what's true in a fiction is what's true at a certain world: worlds are individuated by the propositions that are true at them. You can't change what's true at a particular world; at best, you can change which world you're talking about. So if works of fiction are individuated by the worlds that make their fictional claims true, then strictly speaking you can't change what's true in a work of fiction; at best you can change which work of fiction you're talking about. This might seem absurd, given that fictional change seems not just obviously possible, but obviously ubiquitous. Jesse Pinkman's fate in *Breaking Bad* is but one of an enormous number of vivid examples of this phenomenon.

We see two possible responses to this problem. One response is to deny the possibility of fictional revision. On this view, what looks like fictional change is in fact just the creation of a new work of fiction, one that will inevitably have quite a lot of qualitative overlap with the original (at least as far as the principal fictional truths are concerned). To put it picturesquely: fictional "change" is the author moving their cosmoscope from one fictional world to another, eventually landing on the one that best matches their creative vision. On this view, there are really (at least) two versions of *Breaking Bad*: the original version, in which it's true that Jesse Pinkman dies in a drug deal gone awry; and the final version, in which it's false that Jesse Pinkman dies in a drug deal gone awry. Plausibly most all of our ordinary thought and talk involving the name '*Breaking Bad*' (and the names of the show's characters) is directed at the final version. So when we say that p is a fictional truth of *Breaking Bad*, what we say is true if and only if p is true in the world of the final version. That said, in some special contexts—like in the interviews with Gilligan where he discusses his original authorial intentions—we can

[58] https://en.wikipedia.org/wiki/Jesse_Pinkman#Production.

think and talk about the first version of *Breaking Bad*, for which the fictional truths (whether principal or otherwise) can come quite dramatically apart from those of the final version.

Alternatively, we might deny that works of fiction are individuated by the worlds that make their claims true. Instead we might think of works of fiction as functions from times to sets of fictional truths. If a fiction f's principal fictional truths change over time, then there is no single world w that is the timeless world of f; instead, there are as many such worlds as there are times in which the principal fictional truths change. On this view, there is only one fiction *Breaking Bad*; but there are multiple worlds that are "the" world of *Breaking Bad*, at least across different times. At one time t_1 the world of *Breaking Bad* was w_1, and at w_1 it is (externally) true that Jesse Pinkman dies in a drug deal gone awry; but at t_2 the world of *Breaking Bad* is w_2, and there it isn't (externally) true that Jesse Pinkman dies in a drug deal gone awry. By default, when we say that p is true in *Breaking Bad*, we talk about what's true in the most recent version of *Breaking Bad*—i.e., what's true in *Breaking Bad* since the last revision to its principal fictional truths. But in special contexts—like in Gilligan's interview—we can talk about what was true in *Breaking Bad* at earlier times.

It is difficult for us to see any deep considerations in favor of one of these two ways of thinking about fictional change. But thankfully we think either is adequate for our purposes. The fact that what's true in a work of fiction can seem to evolve non-monotonically is no threat to the idea that there is a single world consistent with all that's true in a work of fiction. For either such appearances are misleading—what's changing is not the work of fiction itself, but which work of fiction we're talking about—or the appearances are genuine, and what's true in a work of fiction can evolve over time.

12 Conclusion

We have argued that fictional truth vastly outstrips determinate fictional truth. In creating a story, an author commits themselves to various claims—the principal fictional truths. There are an enormous number of worlds consistent with a story's principal fictional truths, and so if fictional truth were a function of what was in common to all such worlds, then the set of fictional truths would be highly impoverished. But our patterns of rational attitude formation suggest that fictional truth is not a function of what's in common to all such worlds, but is instead a function of what's true in exactly one of those worlds—in particular, the unique world that would correctly represent how things are, had the work of fiction in fact been an accurate work of non-fiction. And although we are in no position to know which world that is, we know that it is as rich in detail as reality itself.

References

Altshuler, Daniel, & Maier, Emar. 2020. Death on the Freeway: Imaginative Resistance as Narrator Accommodation. *In:* Frana, Ilaria, Benito, Paula Menendez, & Bhatt, Rajesh (eds), *Making Worlds Accessible: Festschrift for Angelika Kratzer*. UMass ScholarWorks.

Anand, Pranav, & Hacquard, Valentine. 2013. Epistemics and attitudes. *Semantics and Pragmatics*, **6**(8), 1–59.

Antonsen, Paal Fjeldvig. 2020. Fictional Reports. *Synthese*, 1–14.

Bacon, Andrew. 2013. Quantificational Logic and Empty Names. *Philosophers' Imprint*, **13**, 1–21.

Bacon, Andrew. 2015. Stalnaker's Thesis in Context. *Review of Symbolic Logic*, **8**(1), 131–163.

Bacon, Andrew. 2018. *Vagueness and Thought*. Oxford, England: Oxford University Press.

Badura, Christopher, & Berto, Francesco. 2019. Truth in Fiction, Impossible Worlds, and Belief Revision. *Australasian Journal of Philosophy*, **97**(1), 178–193.

Blumberg, Kyle. 2018. Counterfactual Attitudes and the Relational Analysis. *Mind*, **127**(506), 521–546.

Blumberg, Kyle. forthcoming. Wishing, Decision Theory, and Two-Dimensional Content. *Journal of Philosophy*.

Blumberg, Kyle H., & Goldstein, Simon. forthcoming. Attitude Verbs' Local Context. *Linguistics and Philosophy*.

Bowker, Mark. 2021. Truth in Fiction, Underdetermination, and the Experience of Actuality. *British Journal of Aesthetics*, **61**(4), 437–454.

Byrne, Alex. 1993. Truth in Fiction: The Story Continued. *Australasian Journal of Philosophy*, **71**(1), 24–35.

Cariani, Fabrizio, & Goldstein, Simon. 2018. Conditional Heresies. *Philosophy and Phenomenological Research*.

Currie, Gregory. 1990. *The Nature of Fiction*. Cambridge University Press.

Dohrn, Daniel. 2015. According to the Fiction. A Metaexpressivist Account. *Pages 37–60 of: Proceedings of the European Society of Aesthetics 7*.

Dorst, Kevin. 2019. Lockeans Maximize Expected Accuracy. *Mind*, **128**(509), 175–211.

Dorst, Kevin, & Mandelkern, Matthew. 2023. Good Guesses. *Philosophy and Phenomenological Research*, **105**(3), 581–618.

Field, Hartry. 2010. This Magic Moment: Horwich on the Boundary of Vague Terms. *In:* Dietz, Richard, & Moruzzi, Sebastiano (eds), *Cuts and Clouds: Vaguenesss, its Nature and its Logic*. Oxford University Press.

von Fintel, Kai. 1999. NPI Licensing, Strawson Entailment, and Context Dependency. *Journal of Semantics*, **16**(2), 97–148.

Franzén, Nils. 2021. Fictional Truth: In Defense of the Reality Principle. *Page 0 of:* Maier, Emar, & Stokke, Andreas (eds), *The Language of Fiction*. Oxford University Press.

Friend, Stacie. 2017. The Real Foundation of Fictional Worlds. *Australasian Journal of Philosophy*, **95**(1), 29–42.

García-Carpintero, Manuel. 2022. 'Truth in Fiction' Reprised. *British Journal of Aesthetics*, **62**(2), 307–324.

Gatzia, Dimitria Electra, & Sotnak, Eric. 2014. Fictional Truth and Make-Believe. *Philosophia*, **42**(2), 349–361.

Gendler, Tamar Szabó. 2000. The Puzzle of Imaginative Resistance. *Journal of Philosophy*, **97**(2), 55.

Goodman, Jeremy. manuscript. *Consequences of Conditional Excluded Middle*.

Goodman, Jeremy, & Holguín, Ben. forthcoming. Thinking and Being Sure. *Philosophy and Phenomenological Research*, 1–27.

Grano, Thomas, & Phillips-Brown, Milo. fc. *(Counter)factual want ascriptions and conditional belief*. MIT.

Hanley, Richard. 2004. Lewis on Truth in Fiction. *Page 113 of:* Jackson, Frank, & Priest, Graham (eds), *Lewisian Themes: The Philosophy of David K. Lewis*. Oxford University Press.

Hawthorne, John, Rothschild, Daniel, & Spectre, Levi. 2016. Belief is Weak. *Philosophical Studies*, **173**(5), 1393–1404.

Heim, Irene. 1983. On the Projection Problem for Presuppositions. *Pages 249–260 of:* Portner, P., & Partee, B. H. (eds), *Formal Semantics - the Essential Readings*. Blackwell.

Heim, Irene. 1992. Presupposition Projection and the Semantics of Attitude Verbs. *Journal of Semantics*, **9**(3), 183–221.

Hintikka, Jaakko. 1962. *Knowledge and Belief*. Ithaca: Cornell University Press.

Holguín, Ben. 2022. Thinking, Guessing, and Believing. *Philosophers' Imprint*, **22**(1), 1–34.

Jacquette, Dale. 1989. Mally's Heresy and the Logic of Meinong's Object Theory. *History and Philosophy of Logic*, **10**(1), 1–14.

Khoo, Justin. 2021. No Fact of the Middle. *Noûs*, **56**(4), 1000–1022.

Kim, Hannah H. forthcoming. Imagination and the Permissive View of Fictional Truth. *Australasian Journal of Philosophy*.

Kocurek, Alexander W. 2021. Counterpossibles. *Philosophy Compass*, **16**(11), e12787.

Kripke, Saul A. 2013. *Reference and Existence: The John Locke Lectures*. New York: Oxford University Press.

Lewis, David. 1973. Counterfactuals and Comparative Possibility. *Journal of Philosophical Logic*, **2**(4), 418–446.

Lewis, David K. 1978. Truth in Fiction. *American Philosophical Quarterly*, **15**(1), 37–46.

Maier, Emar, & Semeijn, Merel. 2021. Extracting Fictional Truth From Unreliable Sources. *In:* Maier, Emar, & Stokke, Andreas (eds), *The Language of Fiction*. Oxford University Press.

Mandelkern, Matthew. 2018. Talking About Worlds. *Philosophical Perspectives*, **32**(1), 298–325.

Moss, Sarah. 2013. Subjunctive Credences and Semantic Humility. *Philosophy and Phenomenological Research*, **87**(2), 251–278.

Motoarc, Ioan-Radu. 2017. A Bad Theory of Truth in Fiction. *British Journal of Aesthetics*, **57**(4), 379–387.

Nolan, Daniel. 2021. Impossible Fictions Part I: Lessons for Fiction. *Philosophy Compass*, **16**(2), 1–12.

Parsons, Terence. 1980. *Nonexistent Objects*. Yale University Press.

Phillips, John F. 1999. Truth and Inference in Fiction. *Philosophical Studies*, **94**(3), 273–293.

Poidevin, Robin Le. 1995. Worlds Within Worlds? The Paradoxes of Embedded Fiction. *British Journal of Aesthetics*, **35**(3), 227–238.

Predelli, Stefano. 2008. Modal Monsters and Talk About Fiction. *Journal of Philosophical Logic*, **37**(3), 277–297.

Priest, Graham. 1997. Sylvan's Box: A Short Story and ten Morals. *Notre Dame Journal of Formal Logic*, **38**(4), 573–582.

Proudfoot, Diane. 2006. Possible Worlds Semantics and Fiction. *Journal of Philosophical Logic*, **35**, 9–40.

Sainsbury, Mark. 2014. Fictional Worlds and Fiction Operators. *Pages 277–289 of:* García-Carpintero, Manuel, & Martí, Genoveva (eds), *Empty Representations: Reference and Non-Existence*. New York, NY: Oxford University Press.

Salmon, Nathan. 2011. Fiction, Myth, and Reality. *In:* Berger, Alan (ed), *Saul Kripke*. New York, NY: Cambridge University Press.

Santorio, Paolo. 2022. Path Semantics for Indicative Conditionals. *Mind*, **131**(521), 59–98.

Schlenker, Philippe. 2009. Local Contexts. *Semantics and Pragmatics*, **2**(3), 1–78.

Schultheis, Ginger. forthcoming. Counterfactual Probability. *Journal of Philosophy*.

Schulz, Moritz. 2014. Counterfactuals and Arbitrariness. *Mind*, **123**(492), 1021–1055.

Schulz, Moritz. 2017. *Counterfactuals and Probability*. Oxford, United Kingdom: Oxford University Press.

Semeijn, Merel. forthcoming. On the Difference Between the ?in? and ?According to? Operators. *Linguistics and Philosophy*, 1–26.

Skipper, Mattias. forthcoming. Good Guesses as Accuracy-Specificity Tradeoffs. *Philosophical Studies*.

Skow, Bradford. 2022. When (Imagined) Evidence Explains Fictionality. *Journal of Aesthetics and Art Criticism*, **80**(4), 464–476.

Spencer, Jack. forthcoming. Intrinsically Desiring the Vague. *Proceedings of the Aristotelian Society*.

Stalnaker, Robert. 1968. A Theory of Conditionals. *Pages 98–112 of:* Rescher, Nicholas (ed), *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*. Oxford: Blackwell.

Stalnaker, Robert C. 1980. A Defense of Conditional Excluded Middle. *Pages 87–104 of:* Harper, William, Stalnaker, Robert C., & Pearce, Glenn (eds), *Ifs*. Reidel.

Starr, W. 2022. Counterfactuals. *In:* Zalta, Edward N., & Nodelman, Uri (eds), *The Stanford Encyclopedia of Philosophy*, Winter 2022 edn. Metaphysics Research Lab, Stanford University.

Stokke, Andreas. 2021. Fiction and Importation. *Linguistics and Philosophy*, **45**(1), 65–89.

Sudo, Yasutada. 2014. Presupposition Satisfaction in Attitude Contexts and Modal Subordination. *Pages 175–199 of:* Crnič, Luka, & Sauerland, Uli (eds), *The Art and Craft of Semantics: A Festschrift for Irene Heim*, vol. 2. Cambride, MA: MIT Working Papers in Linguistics.

Sylvan, Richard. 1980. *Exploring Meinong's Jungle and Beyond: An Investigation of Noneism and the Theory of Items*. Research School of Social Sciences, Australian National University.

Uegaki, Wataru. 2015. Content Nouns and the Semantics of Question-Embedding. *Journal of Semantics*, **33**(4), 623–660.

Unger, Peter K. 1975. *Ignorance: A Case for Scepticism*. Oxford University Press.

van Inwagen, Peter. 1977. Creatures of Fiction. *American Philosophical Quarterly*, **14**(4), 299–308.

Voltolini, Alberto. 2019. Varieties of Fictional Operators. *Pages 199 – 210 of:* Capone, Alessandro, Carapezza, Marco, & Piparo, Franco Lo (eds), *Further Advances in Pragmatics and Philosophy: Part 2 Theories and Applications*. Springer Verlag.

Walton, Kendall L. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.

Weatherson, Brian. 2004. Morality, Fiction, and Possibility. *Philosophers' Imprint*, **4**, 1–27.

Williams, J. Robert G. 2010. Defending Conditional Excluded Middle. *Noûs*, **44**(4), 650–668.

Williams, J. Robert G., & Woodward, Richard. 2019. The Cognitive Role of Fictionality. *Philosophy and Phenomenological Research*.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.

Woods, John. 2018. *Truth in Fiction: Rethinking its Logic*. Cham, Switzerland: Springer Verlag.

Woodward, Richard. 2011. Truth in Fiction. *Philosophy Compass*, **6**(3), 158–167.

Zalta, Edward N. 1983. *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Dordrecht, Netherland: D. Reidel.