

# 1 The Secret Life of Literature

## 1.1 What It Looks Like

In a famous scene from an American novel, one twelve-year-old boy is hoodwinking another. The occasion is so iconic that, in 1972, the US Postal Service honored it with a special stamp. Designed by the artist Bradbury Thompson (figure 1.1), the stamp depicts Tom Sawyer pretending to be absorbed in whitewashing a fence, while Ben Rogers watches intently.

Only a minute ago, Ben was playing a game—impersonating a Missouri steamship—but now it has lost all charm for him. All he wants is to take over Tom’s chore, and, after appropriate hesitation and negotiation, Tom obliges, quietly exulting in his cleverness: “Tom gave up the brush with reluctance in his face, but alacrity in his heart. And while the late steamer Big Missouri worked and sweated in the sun, the retired artist sat on a barrel in the shade close by, dangled his legs, munched his apple, and planned the slaughter of more innocents. There was no lack of material; boys happened along every little while; they came to jeer, but remained to whitewash.”<sup>1</sup>

Ben is sweating in the sun, Tom is sitting in the shade, and Twain is having fun with a biblical reference. His twelve-year-old Herod will soon “slaughter” more “innocents.” With macabre logic, Twain describes those innocents as things inanimate. They merely “happen along,” as a “material” on which “the retired artist” can work at leisure, dangling his legs and munching an apple.

What underlies these ironic twists—that is, the reason that we understand why the boys are described as being massacred and manipulated—is a series of psychological insights developed by Twain’s protagonist. Tom doesn’t want his friends to realize that he hates whitewashing the fence. He discovers that if he makes them think that he enjoys it, they’ll see it as play instead of work and even pay him for the privilege of doing his chore.

Take another look at those insights. Each of them is structured as a mental state within a mental state within yet another mental state: Tom *doesn't want* his friends to *realize* that he *hates* whitewashing the fence; he *wants* them to *think* that he *enjoys* it. Granted, these are my formulations, but if you try to come up with one of your own, you will discover that, if you want to capture the complexity of the social situation conjured up by Twain, simpler descriptions of mental functioning, such as “he *wants* them to do his work for him” or “they *think* that he *likes* painting the fence,” won’t do. In fact, they’ll misrepresent what’s going on, until you find a way to connect them, through another thought or intention. It seems, in other words, that, however you choose to phrase it, you’ll need to recursively embed mental states on at least the third level.

Cognitive psychologists and philosophers of mind talk about “mental states” in conjunction with “theory of mind” and “mindreading,” which are metaphorical terms<sup>2</sup> used to describe our capacity to see behavior as caused by mental states, such as thoughts, desires, feelings, and intentions.<sup>3</sup> Embedment is yet another metaphor, which comes in handy when we want to talk about complex social dynamics that depend on people’s awareness of their own and other people’s states. (Although cognitive scientists have several different terms to talk about this kind of awareness, including, for instance, “recursive intention-reading” and “recursive mind-reading,”<sup>4</sup> I prefer the shorter “embedment.”) To illustrate the way the term “mental states” will be used throughout this study, here are some examples, with mental states italicized:

- “My last name begins with a *Z*” contains no mental states, embedded or otherwise.<sup>5</sup>
- “I’m *glad* that my last name begins with a *Z* because the teacher may not get to the end of the list today” contains just one mental state: my being happy about being at the end of the class list.
- “I am *afraid* that the teacher *will remember* that she hasn’t called on me for a while” contains two embedded mental states: my thinking about what my teacher may be thinking.
- Finally, “I *wonder* if the teacher *realizes* that I’m *hoping* that she won’t call on me today because my last name begins with a *Z* and will thus on purpose start at the end of the list” contains three embedded mental states: my thinking about the teacher’s thinking about my thinking.

Note that we have to rely on this kind of propositional, or representational, language (“I wonder if she realizes that I’m hoping”) to talk about embedded mental states, because it is the tool that we have at our disposal to model the complex intersubjective dynamic of such situations. The actual cognitive processes involved in our experience of those situations may not be structured like embedded representations or may not even “be structured at all.”<sup>6</sup> Moreover, mindreading, especially in face-to-face communication, depends on embodied feedback loops (for instance, there may be something about the expression on my teacher’s face, as she catches me watching her intently, that may strengthen or weaken my hopes), but these important nuances are left out of our crude linear diagrams.

Later, in chapter 3, I will consider in detail this problem of our limited vocabulary. Here, I want us to focus on something else. Ask yourself, How often, in our daily goings-on, do we thus embed mental states on the third and fourth levels? Or, to put it differently, how often do we find ourselves involved in social situations that would require this kind of language to describe them? Although it certainly happens—I am thinking now about faculty meetings, fraught family get-togethers, and love triangles—a majority of our routine social interactions probably don’t require such complex embedments. For instance, I see my neighbor coming out of his house and strolling toward his car, and I assume that he *wants* to go somewhere; or I see my son pulling out a box of pencils, and I assume that he *intends* to draw. (I may not be consciously aware of my assumptions, yet they may influence my subsequent course of action.)

So, on the one hand, yes, “human collaborative activity and cooperative communication both rest on . . . recursive intention-reading.”<sup>7</sup> But, on the other hand, thinking about thinking about thinking (third-level embedment) “occurs in interpersonal cognition in real life less frequently” than, for instance, thinking about thinking (second-level embedment). The former, as the psychologist Patricia Miller et al. put it, “has a lower ecological plausibility.”<sup>8</sup>

Hence an important difference between our daily mindreading and our experience of reading literature. Literature creates intersubjective situations of a kind that can be described as depending on “complex embedments of mental states” at a much greater frequency than it happens in our daily life. Specifically—and this is what I call the secret life of literature—to make sense of what’s going on in novels, plays, and narrative poems, as well as in

memoirs focused on imagination and consciousness, we constantly embed mental states on at least the third level. The key word here is “constantly,” for neither literary critics nor lay readers appreciate the true scale of this phenomenon. To put it starkly, literature, *as we know it today* (this is an important point that I will keep emphasizing) cannot function on lower than the third level of embedment. As such, it differs, for example, from expository nonfiction, such as newspaper articles and textbooks,<sup>9</sup> which may contain occasional forays into the third level but can also subsist, quite happily, on just the first and second levels.

Literature, of course, is a capacious concept, and it encompasses many more genres than I just listed. To give just a few examples, it includes personal essays by writers ranging from Sei Shonagon and Michel de Montaigne to Wole Soyinka and Joan Didion; mirrors for princes, from Augustine’s *The City of God* to Machiavelli’s *The Prince*; and political speeches, from Lincoln’s “The Gettysburg Address” to King’s “I Have a Dream.” While these texts range widely in their frequency of complex embedment (and there are, among them, some pretty spectacular embedders), they mostly do not depend on it to the same high degree as do novels, plays, narrative poems, and memoirs concerned with consciousness.

In the latter, embedded mental states can be found on the level of individual sentences, paragraphs/stanzas, and whole chapters/acts.<sup>10</sup> They can belong to characters, narrators, (implied) authors, and readers, in a vast variety of combinations.<sup>11</sup> In “Tom *wants* his friends to *think* that he *enjoys* his chore,” the third-level embedment involves the novel’s characters. But at the same time, yet another complex embedment arises from an intricate give-and-take between the narrator and his audience.<sup>12</sup> The narrator *expects* that his readers will *appreciate* his mischievous *intention*, as he likens Tom, in the same breath, to King Herod and to a retired artist. Again, this is my formulation, but if you try to explain how this passage achieves its ironic effect, you are likely to find yourself speculating about how the author might have been intuitively anticipating his readers’ thinking.<sup>13</sup>

It would be wrong to assume, however, that we factor mental states of the implied author and reader into any complex embedment. Of course, we can say, “the implied author *wants* us to *know* that Tom *wants* his friends to *think* that he *enjoys* his chore,” and call it a case of fifth-level embedment instead of third, but those extra levels are redundant because they don’t contribute anything to our understanding of the passage. In contrast, the

references to King Herod and to a retired artist are the kind of “communicative event”<sup>14</sup> that necessitates a recognition of a particular intentionality behind it.<sup>15</sup>

When we read, we do not spell it out to ourselves the way I just did. Indeed, in spite of the language that I may use to describe it—such as “we are aware” or “the author wants us to know”—most of it doesn’t rise to the level of conscious awareness. Nevertheless, on some level we must be keeping track of those complex intentionalities (which is a term I will use interchangeably with “mental states” to avoid sounding repetitive), because, otherwise, how would we explain to ourselves, say, Twain’s evocation of the Massacre of Innocents in a scene that had nothing to do with infanticide? To recognize an allusion or to appreciate a metaphor is to acknowledge an intention.<sup>16</sup>

Throughout this book, I use the term “implied mental states” to refer to thoughts and feelings of characters, narrators, authors, and readers that are thus *not* spelled out but are nevertheless integral to our making sense of what we read. But, of course, a work of fiction may also contain complex embedments of mental states that are explicitly spelled out by the author. For instance, think of the time when Tom first encounters Becky Thatcher and starts showing off by engaging in various “dangerous gymnastic performances.” Becky observes him for a while, then throws him a flower and disappears inside her house. Tom keeps up his antics for some time, because he *hopes* that she is still *aware* of his *interest* in her. Or, as Twain puts it, explicitly describing Tom’s embedded thoughts, “Tom comforted himself a little with the *hope* that she had been near some window, meantime, and been *aware* of his *attentions*.”<sup>17</sup>

Here is another explicitly spelled-out complex embedment. When Aunt Polly punishes Tom for breaking a sugar bowl and then finds out that it was Sid who broke it, she can’t bring herself to confess that she has been in the wrong—for “discipline forbade that”—and goes “about her affairs with a troubled heart,” while Tom, perfectly aware of her remorse, is quietly exalting in it. He *knows* that his aunt *is yearning* for his *forgiveness* (third-level embedment), and he *enjoys* knowing that she is yearning for his forgiveness (fourth-level). Or, as Twain puts it, “He knew that in her heart his aunt was on her knees to him, and he was morosely gratified by the consciousness of it.”<sup>18</sup>

What is crucial about these third- and fourth-level embedments is that they do not just occasionally happen along. Instead, any given paragraph contains multiple complex embedments, sometimes implied, sometimes

explicitly spelled out, sometimes a combination of the two. As I am writing this and leafing through *Tom Sawyer*, I reach almost at random for a complex embedment here and a complex embedment there; but in pretty much every case, I can turn to a group of sentences preceding or following any passage that I just quoted for you, and it will contain another implied or explicitly spelled-out complex embedment.

I started this chapter with a picture of a postal stamp, so before moving on, let us briefly circle back to the visual. Do visual representations, such as paintings and movies, also depend on complex embedment of mental states? A short answer to this question is that they do—feature movies more consistently than paintings—and in ways specific to their contexts.

For instance, Bradbury Thompson's portrayal of Tom and Ben is brimming with intentionalities: that of the artist (who apparently decided to portray the boys younger than they are in the story); that of the particular beholder (for I am aware, as I am looking at this stamp, of trying to square the artist's vision with my own perceptions of Tom Sawyer, formed years ago, in a different language, and then layered with the later, "American" impressions); and, of course, that of the characters themselves (i.e., Tom *wants* Ben to *think* that he is too *absorbed* in his task to even notice him). As we take in this stamp, we may toggle between different constellations of complex embedments. This "secret life" of visual images deserves its own study, but for now we return to literature.

## 1.2 A Dime a Dozen

Sometimes, shortly after I'd given a talk about embedded mental states, I would receive emails from members of my audience, something to the effect of "Have you noticed this embedment in such and such work of fiction?" On the one hand, such letters make me happy: they show that the senders are now aware of this phenomenon and want to share their new awareness. On the other hand, a part of me is wondering if it means that I failed to get across one of my key points, which is that third-level embedments in literature are nothing to write home about: they are a dime a dozen. True, their frequency increases dramatically with the advent of certain genres, such as ninth-century Chinese tales of romance, eleventh-century Japanese novels, sixteenth-century Spanish novels, and eighteenth-century English novels. But even the earliest works of literature available to us, such as *The*

*Epic of Gilgamesh* (ca. 2100 BC), already feature some complex embedments. And, generally (although with some fascinating exceptions, which I address later), when it comes to a work of fiction written within the past three hundred years, to discover a third-level embedment in it is roughly as exciting as to discover a noun.

Along the same lines, I suggest to students interested in “cognitive” approaches to literature that merely locating a series of complex embedments in this or that text does not constitute literary analysis. The question is not whether such embedments are there—for they are pretty much guaranteed to be there—but what effect they have on our interaction with the text. For instance, if complex embedments involve mental states of characters *and* are explicitly spelled out, then the text in question is more likely to be considered “popular fiction.” In contrast, “literary fiction” of the kind that may end up on a college syllabus tends to include embedded mental states of narrators and implied authors and readers (in addition to mental states of characters) *and* to imply mental states (in addition to or instead of explicitly spelling them out). So thinking about different types of complex embedment allows us to understand something new about the distinction between “low-brow” and “high-brow” literature.

It also alerts us to cultural contexts that sustain those distinctions. For instance, when students encounter a work of fiction in a college literature course, they tend to work harder on reading implied complex embedments into it and expect to be rewarded for doing so (more about that in chapter 4). In contrast, when readers are faced with a text that they judged a priori as “having lower literary merit”—as, for instance, may be the case with readers prejudiced against science fiction—they may “exert less inference effort”<sup>19</sup> in situations that require supplying mentalistic explanations of characters’ behavior. This is not to say that our intuitions about embedded mental states are solely determined by the context in which we read a given text but that we are sensitive to *both* such contexts and the cues supplied by the text.

Here is another way in which paying attention to complex embedments opens up new venues in literary analysis. It turns out that some fictional characters are consistently portrayed as more capable of embedding mental states on a high (i.e., third and fourth) level than are others. What factors inform the intuitive decision, on the part of the author, to make one character more “sociocognitively complex” than another? More often than not, the decision seems to be influenced by the character’s social status,

which is figured out along the lines of class, gender, or race. A “cognitive” approach thus builds on and complements the rich literary-critical tradition of exploring the role of class, gender, and race considerations in the construction of fictional subjectivity.

Then there is also the issue of the history of complex embedment in literature. What combinations of cognitive/cultural/historical contingencies make it more or less likely that a particular literary tradition would be characterized by a commitment to complex embedments? In some ways, this is the most difficult question one can ask, and we may never come up with a definitive answer. Yet exploring this issue is important, if only because it forces us to become aware of a broader range of historical factors than we usually settle for, in our critical studies.

So the question is not “Are there any complex embedments in this text?”—because, almost always, there are—but “What work do they do?” and “How has it come to be that way?” These are the questions that I encourage my students to ask and that I, myself, ask in the chapters that follow. But, first, I want to give you a range of examples of complex embedments, to show what forms they take in different texts. I hope that, after reading this chapter, you, too, will be struck by this phenomenon: apparently so essential to our interaction with literature yet so invisible, flying, mostly, under the critical radar.

### 1.3 Explicitly Spelled-Out Mental States

The majority of complex embedments in literature are implied rather than explicitly spelled out. In fact, some texts contain next to zero explicitly articulated embedments. Still, many do. In this section, I present examples of such *explicit* embedments, starting with works of fiction published recently and then moving back in time and ending with *The Epic of Gilgamesh*. I do not analyze any of the passages—that will come in later sections. I just pile them up to give you some idea of the range of literary texts that depend on explicit complex embedments. (In what follows, emphasis is mine throughout, unless stated otherwise.)

In Elena Ferrante’s *The Story of the Lost Child*, the protagonist says to a friend, “I’m laughing out of despair, because I’ve never been so offended, because I *feel humiliated* in a way that I *don’t know* if you can *imagine*.”<sup>20</sup> In Sally Rooney’s *Conversations with Friends*, Frances reports her thoughts as she observes a man raising both eyebrows at another man in response to



something another woman has said: “I *thought* it was cowardly of Philip to look at Andrew, whom I *knew* he didn’t even *like*, and it made me uncomfortable.”<sup>21</sup> In Rachel Cusk’s *Transit*, a woman gets up to leave after a lunch with her friend, while “darting frequent glances” at her: “It was as if she was *trying to intercept my vision* of her before I could *read* anything into what I saw.”<sup>22</sup> In Tsitsi Dangarembga’s *Nervous Conditions*, the narrator doesn’t want to think through the implications of the “treacherous mazes” of her thoughts about the inescapability of female victimization: “I didn’t want to reach the end of those mazes, because there, I *knew*, I would find myself and I was *afraid* I would not *recognize* myself after taking so many confusing directions.”<sup>23</sup> In Jokha Al Harthi’s *Celestial Bodies*, a young woman named Khawla cannot understand why her sister, Asma, does not realize that the religious texts that she is so fond of bore other people to death: “Khawla was *astonished* at how *oblivious* Asma seemed to the *awful boredom* these ancient books induced.”<sup>24</sup> In Ben Lerner’s *Leaving the Atocha Station*, the protagonist thinks that his communication with his Spanish friend, Theresa, is becoming a travesty: “I saw her reflected in my eyes, saw that she *knew*, or was coming to know, that what interest I held for her, all of it, was virtual, that my appeal for her had little to do with my actual writing or speech, and while she was happy to let me believe she believed in my profundity, on some level she was aware that she was merely encountering herself.”<sup>25</sup>

I chose my next example on a lark. The sentence that contains the spelled-out embedment *is* the story in its entirety, Joy Williams’s “The Museum.” Here it is: “We were not *interested* the way we *thought* we would be *interested*.”<sup>26</sup>

With my next example, I want to show you that science fiction writers do not shun complex embedment. This, in response to the assumption that I encounter often enough (and that has served as the impetus for the study quoted earlier),<sup>27</sup> which is that works of science fiction get by without representing complex mental states.<sup>28</sup> In Philip K. Dick’s *Do Androids Dream of Electric Sheep?*, characters can preprogram their feelings on a special “mood organ.” This allows the author both to depict complex emotions arising during a marital quarrel and to comment on the presumably mechanical nature of their emotional life.

So here is Rick Deckard trying to decide how he wants to make himself feel during an unpleasant conversation with his wife: “At his console he hesitated between dialing for a thalamic suppression (which would abolish his mood of rage) or a thalamic stimulant (which would make him irked

enough to win the argument).” Rick *wonders* if he *wants* to quell or to ratchet up his *feeling of rage*. Moreover, his wife is watching him closely, ready to “dial the maximum” on her mood organ if he dials “for greater venom” on his, that is, *intending* to become even *angrier* in response to his *anger*.<sup>29</sup>

To run a bit ahead of myself, this scene also contains some implied embedments, though, perhaps, to appreciate them one has to be rereading the novel. Repeat readers may enjoy the irony of the situation in which the character whose job it is to hunt down and kill androids—those, presumably, not capable of feeling genuine emotions—himself uses a mood organ. As Ralph James Savarese puts it, “here, the technological apparatus is active and animate; the human hero, passive and inanimate. . . . Even his feelings aren’t strictly organic.”<sup>30</sup> Were we to spell out the underlying embedment (which, again, is *not* something we consciously do when we read), we may say that the author wants his readers to be aware of the muddled thinking behind the discrimination between those who “truly” experience emotions and those who “choose” to experience them.

Back to explicit embedments. In Shirley Jackson’s short story “The Beautiful Stranger,” an unhappily married suburban wife has a sudden revelation that her emotionally abusive husband is gone and in his place there is a “beautiful stranger.” This new man, moreover, knows that she is afraid that her husband may return and is thus not surprised when she looks up at him for reassurance that he is not her husband: “She was *aware* from his smile that he had *perceived* her *doubts*, and yet he was so clearly a stranger that, seeing him, she had no need of speaking.”<sup>31</sup>

In E. M. Forster’s *Howards End*, Margaret Schlegel’s fiancé, Henry Wilcox, is revealed to have had an affair, ten years before, with a woman who is now Leonard Bast’s common-law wife, Jacky. Margaret’s sister, Helen, fresh from the ruckus at the Wilcox’s garden party, at which Henry and Jacky have accidentally come face-to-face, flies back to London and forces her brother, Tibby, to consider a baffling dilemma that he’d rather not consider: “Ought Margaret to *know* what Helen *knew* the Basts to *know*?”<sup>32</sup>

These were all examples from relatively recent literary past. Let us now start moving further back in time. In Lev Tolstoy’s *Anna Karenina* (1877), Alexei Karenin is made to listen to Anna’s delirious speech while she, as everybody believes, is dying: “Alexei Alexandrovich’s inner disturbance kept growing, and now reached such a degree that he ceased to struggle with it; he suddenly *felt* that what he *had considered* an inner *disturbance*

was, on the contrary, a blissful state of soul, which suddenly gave him a new, previously unknown happiness."<sup>33</sup>

In Jane Austen's *Pride and Prejudice* (1813), Miss Bingley talks to Elizabeth about Mr. Wickham's regiment in Mr. Darcy's hearing, because she *hopes* that Elizabeth will be *embarrassed imagining* Mr. Darcy *thinking* about the Bennett girls' involvement with that regiment: "[She] had . . . *intended* to discompose Elizabeth, by bringing forward the idea of a man to whom she *believed* her *partial*, to make her betray a *sensibility* which might injure her in Darcy's *opinion*, and perhaps to remind the latter of all the follies and absurdities by which some part of her family were connected with that corps."<sup>34</sup>

In Cao Xueqin's novel *Dream of the Red Chamber*, also known as *The Story of the Stone* (ca. 1750–1760s), Dai-yu explains to Bao-yu why she is angry at him for having earlier tried to prevent their cousin Xiang-yun from making fun of her: "But what about that look you gave Yun? Just what did you mean by that? I *think* I *know* what you *meant*. You meant to warn her that she would cheapen herself by joking with me as an equal."<sup>35</sup>

In Shakespeare's sonnet 42, the speaker is constructing a complicated argument in order to console himself for the heartbreaking discovery that his mistress and his friend are having an affair. The "loving offenders," he proposes, are actually doing it for his sake: they want to prove their devotion to him. The young man wants to love what the speaker loves: "Thou *dost love* her, because thou *know'st* I *love* her." Just so, his mistress allows herself to be loved by the young man because she, too, *wants* to be *loved* by a man whom the speaker *loves*: "And for my sake even so doth she abuse me, / Suffering my friend for my sake to approve her."

The ending of the poem ("But here's the joy; my friend and I are one; / Sweet flattery! then she loves but me alone") is open to two different interpretations. Either the speaker is *happy* that his mistress *wants* to find new ways of expressing her *love* for him, or the speaker *is aware* that he is trying to make himself *feel better* by *thinking* that his mistress *wants* to find new ways of expressing her *love* for him. It is a choice between self-flattery and self-awareness, and it is a complex embedment of mental states either way.<sup>36</sup>

In Nizami Ganjavi's twelfth-century narrative poem *The Story of Layla and Majnun*, Kais doesn't feel jealous of other boys in school who stare "at Layla open-mouthed" or, if the school is closed, "roam the alleyways and the passages between the market-stalls, all in the hope of catching a tiny glimpse of her dimpled face," because he *knows* that they *don't love* her as

much as he *loves* her: “Naturally, Kais knew that the other boys desired [Layla], but he also *knew* that they *could not desire* her as much as he *did*, and so their antics did not perturb him in the least.”<sup>37</sup>

Let us go yet further back, to the ninth century’s *Book of Exeter*. In the Old English poem “The Wanderer,” the speaker *wonders* why he is not more *depressed* when he *thinks* about death:

Indeed I *cannot think*  
 why my spirit  
*does not darken*  
 when I *ponder* on the whole  
 life of men throughout the world,  
 How they suddenly  
 left the floor (hall),  
 the proud thanes.<sup>38</sup>

In Petronius’s *Satyricon* (first century AD), Lichas wants to sleep with Encolpius to make up for Encolpius’s currently sleeping with Lichas’s long-term mistress, Tryphaena. Encolpius is not interested, and Lichas arranges for Tryphaena to fall for Encolpius’s slave and lover, Giton. Lichas hopes that the jealous Encolpius will want to make Tryphaena angry by taking up with him and thus takes “the trouble to draw [his] attention” to Tryphaena’s relationship with Giton. The plan works well. As Encolpius reports, “Therefore I was the more ready to treat him nicely, and he was delighted beyond measure—being of course quite sure that my lady’s ill-treatment of me would kindle my disgust, and that in my anger I should feel more kindly disposed to him.”<sup>39</sup>

In *The Odyssey* (eighth century BC), one of Penelope’s suitors, Eurymachus, wants to assure her that her son, Telemachus, mustn’t be afraid of him: “To this Eurymachus son of Polybus answered: . . . ‘Telemachus is much the dearest friend I have, and has nothing to fear from the hands of us suitors. Of course, if death comes to him from the gods, he cannot escape it.’ He said this to quiet her, but in reality he was plotting against Telemachus.”<sup>40</sup> Eurymachus *wants* Penelope to *stop worrying* about the suitors’ *intentions* vis-à-vis Telemachus. (Homer, of course, hastens to explain to us that Eurymachus is lying, but our conversation about lying and literary history will have to wait until a later chapter.)

In *The Epic of Gilgamesh* (ca. 2100 BC), the king of the city of Shurruk, named Utnapishtim, is told by God Ea to tear down his house and build a boat that would allow him and his family to survive a great flood that

is about to kill everybody else. Utnapishtim then asks Ea, very reasonably, how he should explain his actions to other people in Shurruk: “Then Ea opened his mouth and said to me, his servant, ‘Tell them this: I have learnt that Enlil is wrathful against me, I dare no longer walk in his land nor live in his city; I will go down to the Gulf to dwell with Ea my lord. But on you he will rain down abundance, rare fish and shy wild-fowl, a rich harvest-tide. In the evening the rider of the storm will bring you wheat in torrents.’”<sup>41</sup>

Ea *wants* the people to *believe* that another god, Enlil, is *angry* at Utnapishtim and that by going down to the Gulf, Utnapishtim hopes to escape Enlil’s wrath. There is plenty of cruel irony in the picture of abundance about to rain on the city that Ea expects Utnapishtim to plant in the heads of his doomed compatriots. By the time Utnapishtim is telling this story to Gilgamesh, the giant flood has already taken place, so he must be aware of this irony. That is, he *knows* that Ea *wanted* to make sure that the citizens of Shurruk wouldn’t *get alarmed* at the sight of his boat and try to do something to escape the coming disaster, just as he knows what Ea’s fanciful talk of “rich harvest-tide” and “wheat in torrents” truly portended. But to talk about irony and implicit realizations, we must go to the next section.

#### 1.4 Implied Mental States: Dramatic Irony and Beyond

In ancient Mesopotamia, Ea was associated with wisdom, magic, and mischief—a trickster figure. Indeed, trickster tales, with their plots of deception, may have been the earliest fictional contexts for implied complex embeddings. And so must have been drama, for what is “dramatic irony” but a cultural shortcut for implicitly acknowledging a particular mindreading dynamic? The audience knows that a character doesn’t know. And what is it that the poor character is so fatally unaware of? More often than not, it has something to do with the intentions of another character, of a deity, or, even, with the character’s own intentions, which have been rendered calamitously obscure to them. This state of affairs doesn’t have to be explicitly spelled out, yet the audiences must be aware of it (an awareness that necessitates embedding complex mental states!) in order to make sense of what is going on.

Because dramatic irony is thus a prototypical implied third-level embedding, I start this section with several examples from plays and then segue to narrative poems, short stories, and novels. The trickster tales will have to wait until a later chapter, dealing with the history of complex embeddings.

Note, too, that, when talking about drama, I focus on playscripts and not performance. The latter, of course, brings in more and different complex embeddings than are present in the script. For instance, the social psychologist Tiziano Furlanetto and his colleagues have found that when an actor's "gaze and action [do] not signal the same intention," observers engage in a stronger "spontaneous perspective-taking," which suggests that, "in presence of ambiguous behavioral intention, people are more likely take the other's perspective to try to understand the action." Thus, if we imagine a character onstage who, in the middle of a complex social interaction with someone else, would start reaching for an object without looking at it, that action alone would complicate our perception of their intentions vis-à-vis others.<sup>42</sup> As Furlanetto et al. put it, "observing a person grasping without looking may thus be perceived as ambiguous. What is he planning to do? Why is he not looking at the object he is reaching for?"<sup>43</sup>

This is just one small example of the role of embodiment in modulating and complicating an audience's perception of actors' embedded intentionality.<sup>44</sup> In general, exploration of embedded mental states that emerge when actors widen and explore the space between their characters' words and their body language deserves a separate study. It is not my aim here to undertake such a study, so we return to mental states implied by texts alone.

In Shakespeare's *Othello*, Iago *wants* Othello *to think* that Desdemona *is in love* with Cassio. In *Romeo and Juliette*, Romeo *does not know* that Juliette is not dead but merely *wants* some people *to think* that she is dead. In *Measure for Measure*, Duke Vincentio *wants* Isabella *to think* that he *doesn't believe* her story about Angelo's "intemperate lust."<sup>45</sup> In *Twelfth Night*, Maria, Sir Toby Belch, and Fabian *want* Malvolio *to think* that Olivia *loves* him.

These are all act- and scene-level implied embeddings. For a quick example of a sentence-level implied embedding in drama, consider a scene from *Twelfth Night*, in which Malvolio first courts Olivia and then exits the stage in full anticipation of the "greatness" that will soon be "thrust upon" him. Once he is gone, Shakespeare has another character, Fabian, make a "nod to the audience,"<sup>46</sup> which starts off a complex embedding involving the audience, the author, and the characters. When Fabian says, "If this were played upon a stage now, I could condemn it as an improbable fiction,"<sup>47</sup> the author slyly tells his spectators that he *knows* that they *know* that the characters *don't know* that they are upon a stage now.

More scene-level complex embedments: In George Etherege's *The Man of Mode* (1676), Dorimant *knows* that his new mistress, though believing herself injured by him, will nevertheless help him to deceive his old mistress (who also happens to be her friend) because she is *afraid* that the old mistress will *realize* that the new mistress has been lying to her.<sup>48</sup> In Oliver Goldsmith's *She Stoops to Conquer* (1773), Hastings *doesn't want* his friend Marlow to *know* that he is *mortified* that Marlow gave the jewels that Hastings had earlier entrusted him with, to Mrs. Hardcastle for safekeeping.

In Wang Shifu's play *The Story of the Western Wing* (thirteenth century), Oriole's maid, Crimson, encourages student Zhang to pursue her young mistress, because she *thinks* that she *knows* Oriole's true *feelings* about the attractive young man. As Zhang becomes too importunate, however, and Oriole responds with indignation, Crimson *realizes* that she must have been *wrong* in *assuming* that Oriole *cares less* about her honor than her love for Zhang.

In *Layla and Majnun*, when the main protagonists, still children, are basking "in the glow of each other's love," the narrator asks us—that is, *wants* us—to *imagine* what other people around them *may be thinking* about their *feelings*: "Did others realize what had happened between Kais and his Layla? Did they see the stolen looks, the furtive glances that passed between them? Could they read the signs and crack the codes of secret love that bound their hearts together? Who knew about them and how much was known? Until one day, in the market, a voice was heard to say, 'Kais and Layla are in love. Have you not heard?'"<sup>49</sup>

In Austen's *Emma* (1815), Frank Churchill *wants* onlookers to *think* that he is *interested* in Emma in order to conceal his engagement with Jane Fairfax. In Forster's *Howards End*, all throughout the novel, that is, "throughout Margaret's various conversation with the Wilcoxes, her marriage to Henry Wilcox, and her sister's involvement with the Basts," readers *know* that Margaret *doesn't know* (while the Wilcoxes do know) that the late Mrs. Wilcox had *wanted* her to inherit Howards End.<sup>50</sup>

In the opening of Anton Chekhov's short story "Rothschild's Fiddle" (1894), we learn that the "town was small, worse than a village, and populated almost only by old people, who died so rarely that it was quite annoying."<sup>51</sup> One can't help wondering what kind of person would consider it so patently obvious that old people should hurry up and die. To map out our implicit reaction explicitly, the narrator *wants* the reader to *wonder* who would *want* old people to die and why.

In Lu Xun's "A Madman's Diary" (1918), the protagonist finds it infinitely amusing that a doctor, whom his older brother brought in to consult, says that he'll be "better" if he rests "quietly for a few days." Because he thinks that the doctor is "the executioner in disguise" and that what he and the brother really want is to eat him, resting quietly for a few days will only fatten him up and thus give them "more to eat." So he laughs uproariously and watches them turn pale, "awed" by his "courage and integrity": "I could not help roaring with laughter, I was so amused. I knew that in this laughter were courage and integrity. Both the old man and my brother turned pale, awed by my courage and integrity."<sup>52</sup> The reader *knows*, however, that the protagonist *doesn't realize* that the reason that the two men turn pale is that they *think* that his laughter is a sure sign of his insanity. Or, to put it differently, the implied author *wants* the readers to *realize* that the mad protagonist *misinterprets* the body language of his visitors.

When Maggie, the protagonist of Hannah Pittard's novel *Listen to Me* (2016), finds an empty bottle of champagne in the recycling bin, her heart sinks. Her husband, Mark, has apparently tossed "without ceremony" the bottle left over from their last anniversary, which she has been saving. She wonders, next, if this is a test and if Mark is "measuring her steadiness"—for she has been going through a rough time lately—"by relieving her of an ultimately trivial trinket." If so, she decides, she "would pass his test with flying colors." That is, she *wants* him to *think* (were he to see the bottle, now placed "at the very top" of the bin) that she *is not overly sentimental* about it. Except that (dramatic irony!) Maggie has just torn off and saved as a keepsake "a sliver of the pink foil," which means that she is actually failing the test that she imagines Mark has set up for her.<sup>53</sup> Or, to spell out this embedment explicitly, the implied author *wants* the reader to *realize* that Maggie *is fooling herself*. (And so is, for that matter, Mark, but the implied embedments involved in his self-deception are constructed on other occasions.)

The first sentence of Zadie Smith's *On Beauty* (2005), "One may as well begin with Jerome's emails to his father," overflows with embedded intentionality.<sup>54</sup> The implied author *wants* her readers to *know* that the action will be filtered through the *consciousness* of a reflective narrator. And there is more, of course. Those who are familiar with the opening of *Howards End*, "One may as well begin with Helen's letters to her sister," will sense yet another set of intentions in Smith's first sentence.<sup>55</sup> The author wants her readers to know that the action will be filtered through the consciousness



of a reflective narrator—and that she means her novel to be a meditation on Forster’s novel. There are no direct references to mental states in the sentence about Jerome’s emails to his father, yet its impact on the reader is directly bound to its embedded intentionality.<sup>56</sup>

I don’t think we notice it, though. Were I to articulate my feelings upon first opening Smith’s novel, I would say that I experienced a pleasing jolt of recognition and something that could be expressed in words as, “Oh, so it’s that kind of book!” It is when I try to slow down and figure out what kind of mental work goes into “Oh, so it’s that kind of book!” that I end up considering the embedded intentions of the author.

Let us now revisit Williams’s one-sentence story “The Museum”: “We were not interested the way we thought we would be interested.”<sup>57</sup> I used it in the previous section as an example of explicitly spelled-out embedments in literature, but its affective punch may actually reside with its implied embedments. “The Museum’s” protagonists watch closely their emotional responses, especially when they find themselves in a cultural context that is expected to elicit a particular kind of response. The story thus draws the reader’s attention to a specific sensibility: one predicated on self-awareness yet not always happy about the burden of such awareness. This may be the reason why at least one reviewer characterized “The Museum” as “rueful.”<sup>58</sup>

As with explicit embedments, discussed earlier, I want to see here how far back into literary history I can reach to find examples of implied complex embedment, especially those involving implied authors and readers. Let’s start with a novel written in the second century AD, Apuleius’s *The Golden Ass*. When one of its characters, goddess Venus, learns that her son, Cupid, has ignored her order to humiliate and destroy Psyche (of whose beauty Venus was jealous) and instead married Psyche and that they are now expecting a baby, she rushes into the bedroom where Cupid lies and begins “roaring with all the strength in her”:

Pretty classy goings-on, huh? A nice way to make your family look good! . . . I was in a fight to the finish with a girl, and now I have to put up with her as my daughter-in-law? And what’s more, you worthless, disgusting hound, you assume that you’re the only one fit to breed, as if I’m too old to have a baby. This is just to let you know: I am going to have another son, much better than you, and to humiliate you even more I’m going to adopt one of the slaves born in my house, sign everything over to him: those wings and that torch, and that bow, and your actual arrows—all the tools of my trade, which I didn’t give you to use like this.

It's totally up to me, because there was no money set aside from your father's estate to buy you this equipment.<sup>59</sup>

Venus wants Cupid to know that she is extremely angry. What Venus doesn't know, however, is that, just now, Cupid has abandoned Psyche for not trusting him and following the advice of her envious sisters and that Psyche is desperate to win back Cupid's love. (Were Venus to know all this, she might try attacking Psyche while the girl is lonely and vulnerable, instead of simply venting her anger at her son.)

Those are straightforward enough embedments, but they are not what makes the passage hilarious. What makes it hilarious is the interplay of mental states of the implied author and the reader. As the novel's recent translator Sarah Ruden puts it, Apuleius "exquisitely [manages] the tension between the high and low, the inside and outside points of view."<sup>60</sup> The goddess of love, beauty, fertility, and prosperity comes across as garrulous, jealous, feeling her age, and penny-pinching. Apuleius knows that we don't expect Venus to sound like this, and we *know* that he *knows* that we *didn't expect* this.

The comic effect of Venus's speech—if, that is, we find it funny!—stems from this embedded awareness. This point is worth emphasizing because to phenomenologically "get" the joke, readers must swiftly process embedded mental states. My map here thus seeks to capture something that readers actually interpretively do, "rather than being an analytical account of the semantics of the text, divorced from the reader."<sup>61</sup>

As always with complex passages, there are often several ways to map them out. I just suggested one—"we know that Apuleius knows that we didn't expect Venus to sound like this"—but a different mapping is also possible. Readers may or may not remember that, within the novel, the story of Cupid and Psyche is narrated by an old crone who keeps house for pirates and who wants to soothe and entertain a young woman kidnapped by those pirates. So if we do remember it, we can say that "Apuleius uses the old crone as his framing device because he *wants* a narrator *incapable of imagining* a Venus who would *feel* differently from herself under these circumstances."

Let us revisit *Gilgamesh*, which is considered to be one the earliest surviving works of literature. When we learn that Ea wants Utnapishtim to tell the people of Shurruk that the same god who is angry at Utnapishtim will bless them with good fortune ("But on you he will rain down abundance, rare fish and shy wild-fowl, a rich harvest-tide. In the evening the rider of

the storm will bring you wheat in torrents”), it is difficult not to think that Ea is having a joke at the expense of the Shurrukians.<sup>62</sup> Not only does he *want* those people to *think* that they are *beloved* by a god (when exactly the opposite is the case), but he also chooses a very particular vocabulary to convey his lie. How we read the complex give-and-take of intentions implied by his evocation of rain, tide, and torrents depends on whether we perceive Ea as being sadistic, philosophical, or just true to his trickster “nature.” But whichever way we view him, we seem to assume that the narrator of *Gilgamesh* *wants* to draw the audience’s *attention* to Ea’s *desire* to comment on what is to come. This is to say that to judge the ethics of the situation, we have to be intuitively aware of the underlying mental states and hence the irony implied by the disjunction between what is stated and what is intended.

### 1.5 Who Are “We”? Historical Speculations and Empirical Studies

Do we have any evidence that *Gilgamesh*’s early audiences were also aware of this ironic disjunction? While we may never be able to know for sure, thinking about this question raises important issues. One such issue is the identity of the “we” whose reaction to *Gilgamesh* I seem to be quietly presenting as normative and then comparing with that of its early listeners/readers. To put it broadly, are some readers more aware of implied mental states in literature than are others? Does it take a particular training in reading and interpreting works of literature to become a part of this enlightened “we” community?

I can tell you right away that I do not have definitive answers to these questions. What I do have is a series of considerations that bear upon them, directly or indirectly: some based on historical analysis of patterns of reading, others on ethnographical studies of indigenous performance genres, yet others on lab experiments conducted by interdisciplinary teams of literary critics and cognitive scientists. I will share these considerations with you, and then we can see if they add up to any kind of provisional answer.

Let us start with studies that suggest that expert readers of literature become sensitized to certain features of literary texts, including various types of implied intentionality. The social psychologists David Comer Kidd and Emanuele Castano, who study effects of reading fiction on theory of

mind, have shown that long-term exposure to literary fiction makes readers less willing to settle for unambiguous interpretation of mental states and more eager to look for cues of intentionality.<sup>63</sup> This may mean that the more literature one reads, the more implied mental states one is prepared to see in what one is reading.

Moreover, the literary critic and neuroscientist Natalie Phillips has found that professors of literature did not make good subjects in fMRI experiments that focused on reading for pleasure, because they found it hard to stop close reading, that is, analyzing what they read. (At least that was what happened when the text in question was deemed worthy of close reading; one wonders if those professors would have a similarly hard time refraining from analyzing a work of science fiction.) In contrast, graduate students qualified for participation in such experiments because they still seemed to be able to read classic literature for pleasure, even though they were en route to becoming professional close readers. (Note that Phillips does not explicitly equate close reading with uncovering new embedded mental states,<sup>64</sup> while I believe that close reading typically involves such uncovering.<sup>65</sup>) So one take-home message from her experiments is that if you spend a good portion of your life not just reading literature but also thinking about it, your experience of reading (which, as I have argued elsewhere, necessarily involves mindreading)<sup>66</sup> becomes different from that of people who either don't read literature or read it but don't think about it professionally.

What we should not do, however, based on these studies, is to overstate that difference and treat it as a *constant*. Consider the work of the cognitive literary critic Andrew Elfenbein, who brings together empirical studies of what readers do today and historical reconstructions of the eighteenth- and nineteenth-century practices of reading. Elfenbein emphasizes continuity between different kinds of reading, reminding us that "many readers . . . are neither novices nor experts but somewhere in between" and that even expert readers are routinely faced with pressures and distractions that lead them to engage in merely "good enough processing, which occurs when [they] process what they have read just enough to make sense of it."<sup>67</sup> What this means is that we should not assume that experienced readers of literature (even, perhaps, the professors from Phillips's experiment) would *always* see more implied embedments in the text than would less experienced readers.

This may be a good time to bring up the difference between reading and interpretation. Although I would dearly love to believe that my maps

of mental states reflect something obvious and hence do not require any special interpretive effort, the truth is that some of them do, and I am not always the best judge of the extent to which such maps depend on my having taken extra time to think about the passage under consideration. To put it bluntly, if I am giving you the results of my interpretive effort while claiming that this is really an effortless reading of the kind that just about any expert reader would produce under the circumstances, then I am inflating the difference between expert reading and novice reading. To quote Elfenbein again, “full comprehension and reading do not co-occur, which is why literary scholars should hesitate more than they do to make ‘reading’ synonymous with ‘interpretation.’”<sup>68</sup>

That said, there is some room between “full comprehension” and “good enough” comprehension. While neither expert nor novice readers may immediately and fully comprehend a rich variety of complex embeddings structuring a given passage, they may grasp enough of some of those embeddings’ meaning to carry them through. In fact, Elfenbein’s discussion of automatic processes involved in reading literature provides a useful framework for thinking about fictional mindreading. If we adopt his model (which itself is based on the work of the psychologist Agnes Moors), we can characterize embedding of complex mental states as “top-down automatic processing: processes that have become automatic as a result of training and repetitive practice.” As Elfenbein explains, “[Such] processes are usually unconscious, but they are not inaccessible to consciousness. They can become conscious when attention is directed to them. . . . Some of these processes include comprehending (understanding the meaning of what is read) and situation model building (integrating what has been read with general world knowledge, cognitive and emotional inferences, predictions, and evaluations).”<sup>69</sup>

More often than not, such “understanding the meaning of what is read” and “situation model building” involve attributing mental states to fictional characters, narrators, and the author. And, just as in real life, much of this fictional mindreading does not rise to the level of awareness, except when we consciously direct our attention to it.

Expert readers of literature may, indeed, develop, “through long practice, a set of strategies for understanding imaginative literature,”<sup>70</sup> which means that sometimes they would indeed be more attuned to intentionality cues in the text than would be less experienced readers. That said,

automatic ascription of mental states to observed behavior is something that, arguably, all readers do. Literary scholars may thus regularly shift “between strategies common to many readers” and those that they have developed as professionals.<sup>71</sup> While in the latter mode, they are more likely to be aware of a richer set of counterintuitive implied embeddings—and thus begin to interpret the text—than they are when in the former mode, when, for instance, they may be distracted or in a hurry or uninterested in what they are reading or having decided that this particular text does not deserve much attention.

So where does all this leave us in respect to the initial question of whether *Gilgamesh's* earliest audiences might have been as aware of its ironic interplay of intentionalities as “we” can be today. I suspect that, *even then*, some members of the audience—those whose experience with imaginary intentionalities was, for whatever reason, more extensive than that of other people—were particularly eager, on some occasions, to intuit more implied mental states in the text. To *them*, Ea’s promise to the people of Shurruk, that soon they will be happily drowning in the torrents of fish and fowl and wheat, might have, sometimes, felt more “dramatically ironic” than it did to others.

But, one may argue, surely, given the variety of artifacts that experiment with nuances of intentionality today (all those novels and movies!), surely, our culture, as a whole, must be more attuned to implied embedded mental states than would be a culture not exposed to such an abundance. I agree with this argument on the condition that we humbly acknowledge that we often have no clue what performative and literary genres may be thriving (or had thrived, thinking back to *Gilgamesh*) in a culture different from ours.<sup>72</sup> So, as long as we keep our potential ignorance in mind, I would say that, yes, a community with a long and rich tradition of representing mental states might be more open to intuiting complex mental states in a given cultural artifact, though, even within that community, some people would still be more eager to look for cues of intentionality than would be others.

So here is one way to think about a hypothetical community of readers—those “we” and “us” and “ours”—which I regularly evoke in this book when I describe embedded intentions along the lines of, “the narrator of *Gilgamesh* wants to draw *our* attention to Ea’s intentions, as Ea refers to the flood that will soon destroy the unsuspecting Shurrukians, as a downpour of abundance.” My hope is that such first-person plural pronouns designate readers who are paying “good enough” attention to what they

are reading, which is to say that they are neither terribly distracted nor engaged in the act of professional literary interpretation. Perhaps I don't always manage to hit that sweet spot (i.e., that "good enough" state of the reading mind) in my mapping of mental states, but this is to what I aspire.

Moreover, given the pragmatics of when and where I wrote and you are reading this book, it is reasonable to assume that these "good-enough" readers have had some exposure to cultural artifacts—such as novels, plays, television series, and movies—that call for attribution of mental states to a broad variety of actual and imagined entities. It remains open to debate whether such exposure makes them (us) better mindreaders in their daily life.<sup>73</sup> Still, it provides them with some training in teasing out hidden intentionalities of characters, authors, and implied audiences, a training that comes in handy in a culture that (mostly) values thinking and talking about one's own and others' mental states. (I have more to say about this in chapter 4, in which I talk about cultures that may not encourage such conversations.)

## 1.6 Studying "Us" in the Lab

Now I will tell you about a different kind of attempt to figure out if there is such a thing as a collective of readers when it comes to the processing of complex embedment of mental states. To see if there is any evidence that different readers are likely to agree on their estimate of the level of embedment in a given passage, a team of cognitive scientists and literary scholars, headed by Douglas H. Whalen, (late) Michael Holquist, and myself, ran a series of experiments at Yale University, Haskins Laboratories in New Haven, and the CUNY Graduate Center. The first set of experiments presented participants with short vignettes, crafted specifically for the occasion and featuring different levels of embedment; the second, with an excerpt from an actual work of literature, Harper Lee's novel *To Kill a Mockingbird*.

To give some idea of what our artificial vignettes looked like, here are two of them, one in which each sentence contains a second-level embedment and another in which each sentence contains a fourth-level embedment:

- [second level] I am not even sure why Stephanie wants to go the movies with Alice and me. She hates the kinds of movies that we like. I remember the last time Alice wanted us to see this retrospective of silent films. We both thought that Stephanie wouldn't enjoy it at all. She went along

and sat through the whole four-hour thing, but we could tell that she was bored. I think I need to figure out how to talk to the two of them about this problem.

- [fourth level] I think my daughter begins to find it a bit irksome that when we visit my aunt she has to be very careful about choosing topics of conversation that won't offend. She knows, for example, that my aunt can't stand it if we suggest that it's not a good idea for her to live alone. We also have to keep in mind not to argue with her about her conviction that she can remember her doctor appointments without ever writing anything down.<sup>74</sup>

Altogether, we had eighty-four vignettes (387 sentences) ranging in their level of embedment from zero to five. When it comes to results, the "great majority of responses were within one level (94.2%), but differences did account for 25.54% of the judgments."<sup>75</sup> This is to say that the participants' judgments were in perfect agreement with the experimenters' judgments in 74.5 percent of cases, while the greatest difference involved one level of embedment in either direction. (That is, in 10.5 percent of cases, participants judged the vignettes to have one more mental state than did the experimenters, and in 9 percent of cases, they judged the vignettes to have one fewer mental states.) In contrast, the participants judged the vignettes to have two more mental states than did the experimenters only in 2 percent of cases, and they judged them to have two fewer mental states only in 0.78 percent of cases; and the numbers went even further down with the difference of three levels, to 0.17 percent and 0.21 percent, respectively.

What we found in the second set of experiments was that an excerpt from *To Kill a Mockingbird*, which featured twelve sentences (three consecutive paragraphs), yielded a lower rate of agreement. Specifically, in half the cases, the participants' judgments were essentially the same as the experimenters', while in three cases, they were above the experimenters' judgments, and in three cases, below.<sup>76</sup> I will not discuss here the setup of our studies, because we already have done it, extensively, elsewhere.<sup>77</sup> I will focus only on three take-home lessons that are relevant for us now.

First, as we realized in the process of devising our vignettes, elements of style such as metaphors, alliterations, and allusions bring in mental states.<sup>78</sup> A single metaphor, even a subdued one, can inadvertently change the tone of a whole vignette, evoking a mental state in a reader, even if that mental



state may be too subtle to describe in a propositional format.<sup>79</sup> To adapt Patrick Colm Hogan's argument from a related context, such a mental state may not be "strongly activated." It will be, "rather, 'primed' in the cognitive sense of the term. Thus [it will be] partially activated in such a way as to affect the orientation of thought and feeling without entailing precise, reasoned consequences."<sup>80</sup> While we did our best to control for this "priming" factor in our synthetic vignettes by draining them of anything that could be seen as a sign of style, we were forcefully reminded that it is style rather than straightforward propositional statements (such as "I know that she knows that I know") that may generate complex embeddings in literary texts.

Second, during one of the sessions in which we introduced our subjects to the concept of counting embedded mental states on the level of an individual sentence (i.e., the unit level on which we eventually settled), we discovered something similar to what Natalie Phillips later observed in her fMRI studies of pleasure reading. It became clear to us that when our subjects happened to be expert readers—such as graduate students in English and comparative literature—they sometimes saw more mental states in a given sentence/paragraph than did lay readers.<sup>81</sup> This observation made a lot of sense if you would consider that people who apply to graduate programs in literary studies may already have higher-than-average interest in intricate social situations that call for attribution of complex mental states<sup>82</sup> and that they may become even more so after spending years dissecting and interpreting mental states of literary characters, authors, and other scholars.<sup>83</sup>

Still, even with those complicating factors (that is, mental states introduced by various elusive elements of style and the difference in our subjects' expertise), "the broad agreement about the levels of embedment in individual sentences demonstrated by our experiment [showed] that sentence-level embedment of mental states is a real phenomenon that can be reliably assessed in a laboratory setting."<sup>84</sup> Studying "us" in the lab is, thus, a legitimate endeavor, especially if one clearly differentiates between one's expectations, in the case of artificial vignettes, and excerpts from literature.

For here is the third take-home lesson from our experiments. It was very encouraging to learn that both researchers and subjects could be trained to judge levels of embedment quickly. It was also heartening to see that their subsequent judgments—in the case of vignettes—displayed a sizable agreement. That said, disagreements—which were especially pronounced in the

case of *To Kill a Mockingbird*—turned out to be illuminating in their own right. In fact, one of the conclusions of our last study was that, particularly when it comes to individual sentences in works of literature, high agreement rates on their levels of embedment should not be expected. While disagreements may have multiple causes, including flaws in the design of experiments, one clear cause must be the “necessary complicating role of large-scale (i.e., paragraph, chapter, and cross-chapter) embedments of mental states in the perception of the sentences,” while another may have to do with the role of personal responses to literature.<sup>85</sup>

To illustrate how such complications work, I will now turn to a novel that our last study mentioned only briefly: E. M. Forster’s *Howards End*. Specifically, I will show that a seemingly clear-cut sentence carries the potential for expanding, contracting, and otherwise changing its levels of embedment. This happens because, far from being a one-shot game, the sentence is part of the dynamic mindreading ecology of the novel—what the philosophers of mind Hanne De Jaegher and Ezequiel Di Paolo would call “the ongoing engagement” between the text and its readers.<sup>86</sup>

### 1.7 Are Embedments *in* the Text?

“Ought Margaret to know what Helen knew the Bastis to know?”

What could be more straightforward than this example of explicitly spelled-out embedded mental states from Forster’s *Howards End*? Yet this straightforwardness is treacherous. The sentence is a Trojan horse harboring implied mental states that rush at us as soon as we move in for a closer look.

Until now, I avoided providing much context for my examples of spelled-out embedments. I did so because I wanted to first clearly lay out the terms of my discussion: “here are the explicit, and here are the implied.” But reality is messier than this neat division may imply: explicitly spelled-out embedments are often integrated with implied ones. Sometimes the relationship between the two is complementary, but, just as often, the implied embedments subvert the explicit ones.

For instance, taken on its own, “Ought Margaret to know what Helen knew the Bastis to know?” seems to present a social dilemma and invite a discussion of what to do next. However, if we look at its context, we realize that this sentence actually does something very different. It mocks drawn-out conversations about relationships and *refuses* to discuss the social dilemma

in question. It thus makes it possible for us to pass over the question about what Margaret ought to know, filing it away, as it were, as a bit of a tedious joke.

To see how it works, let us expand the quote:

[Tibby Schlegel] had never been interested in human beings, for which one must blame him, but he had had rather too much of them at Wickham Place. Just as some people cease to attend when books are mentioned, so Tibby's attention wandered when "personal relations" came under discussion. Ought Margaret to know what Helen knew the Basts to know? Similar questions had vexed him from infancy, and at Oxford he had learned to say that the importance of human beings has been vastly overrated by specialists. The epigram, with its faint whiff of the eighties, meant nothing. But he might have let it off now if his sister had not been ceaselessly beautiful.<sup>87</sup>

This is a very complex passage, and, as is often the case with such, I expect, as I map out its implied embedments, that my understanding of what is going on may not coincide with yours. Still, different as your understanding may be, I encourage you to take note of the mental states involved. For my argument depends not on the unique perceptiveness of my interpretation but on the *complexity of embedments* expected from a reader to make sense of this passage.

The narrator *anticipates* that readers will *dislike* Tibby for not being *interested* in personal relations. The narrator *wants* his readers to *imagine* what it might have *felt* like for Tibby to grow up in a household where such relations were constantly discussed. By doing so, the narrator *wants* us to *recognize* Tibby's *aversion* to such discussions as a self-defense mechanism, even as he lets us *suspect* that he himself may still *consider* Tibby's *supercilious thinking* unsympathetic.

Planted in the middle of these implied embedments, "Ought Margaret to know what Helen knew the Basts to know?" acquires rather unflattering overtones. Instead of signaling social complexity, it signals impatience with overthinking "personal relations." To put it differently, instead of taking the content of the phrase at its face value and engaging earnestly with the question of whether Margaret ought to know and so on, we may now dismiss this content, because, as Tibby has shown us, one way of dealing with this dilemma is to say, "Who gives a hoot?" The implied embedments thus undercut the explicit one.

"Ought Margaret to know what Helen knew the Basts to know?" can be said to be a case of "free indirect discourse," which is yet another term that,

similar to “dramatic irony,” functions as a cultural shortcut designating a specific mindreading dynamic. To use our present vocabulary (instead of a more traditional literary-critical one)<sup>88</sup> to describe this dynamic, we can say that free indirect discourse occurs when the implied author wants the reader to distrust information that the text *seems* to be treating as true. Thus, while it may *seem* that whether Margaret knows what Helen knows is a real concern, on closer inspection, it turns out to be just another example of the type of question that Tibby doesn’t like thinking about.

But guess what? The question whether Margaret knows and so on may be a red herring, but it won’t be put to rest. Far from being confined to the immediate environs of one paragraph, some of its implied complex embedments continue to unspool throughout the novel. For instance, we learned, in Forster’s previous chapter, that Margaret already knows that her fiancé, Henry Wilcox, had had an affair, ten years ago, with the woman who was to become Leonard Bast’s common-law wife and that Margaret has already forgiven Henry. This means that Helen’s present worries are misplaced. That is, Helen *doesn’t know* that Margaret *already knows* “what Helen knew the Basts to know.”

But wait, there is more: Margaret had written a note to Helen—before she realized that Helen may already know about the affair—and that note was driven by Margaret’s *wish* (anticipating and mirroring Helen’s present wish!) to preserve Helen from *knowing* what Margaret *knew*. Over that note, Forster explains, Margaret “took less trouble than she might have done; but her head was aching, and she could not stop to pick her words.”<sup>89</sup> Ironically, upon receiving that less-than-carefully-worded letter, Helen can’t make any sense of it and thus assumes that it is the doing of Henry Wilcox (as she puts it to Tibby, “He makes Meg write”)<sup>90</sup> and that Henry wants to prevent Margaret from learning the truth about his past. For readers who remember Henry’s role in suppressing his late wife’s will (which would have Margaret inherit the Wilcoxes’ country house, Howards End), Helen’s assumption may ring less mistaken than it is.

In other words, every time you change the context for the original straightforward embedment, “Ought Margaret to know what Helen knew the Basts to know,” your interpretation shifts,<sup>91</sup> and every one of these interpretive shifts (i.e., on the level of individual sentences, paragraphs, or chapters) unfolds as yet another complex embedment of mental states involving characters, readers, and the implied author.

Let me pause here, before you start feeling like Tibby and ask exasperatedly, “Ought readers to know what Zunshine considers the implied author to intend?”—a question that may really mean, as we have seen, “Who gives a hoot?”—and decide how many hoots we should give about any of this. There are two points I want to make here. The first is a simple assertion: to read *Howards End* is to embed complex mental states incessantly, whether you are aware of it or not. Yours may not be the same as mine, but that difference in content is less important than the structural similarity: the fact that neither of us can make sense of the text without constantly embedding *some* complex mental states.

The second point is that complex embedments are not merely *in* the text, ready to affect the same way whoever opens the book.<sup>92</sup> Instead, they emerge as a specific reader *acts* on what they are reading by intuitively choosing a context in which to make sense of a potential embedment. Thus, while one reader may indeed focus on the question of whether Margaret ought to know what Helen knew the Basts to know, another reader (or the same reader on a different occasion) may adopt some of Tibby’s indifferent perspective on personal relations and pass over that question; while yet another may particularly respond to the implied author’s attitude toward Tibby’s superciliousness regarding Helen’s concern, and so forth. I suspect that were we to bring *Howards End* to the lab and ask our subjects to count the levels of embedment involved in this paragraph, the numbers that they would report would remain generally high—that is, between three and five—but there would be quite a bit of fluctuation within that range, given that the content and configuration of embedments would differ from one reader to another.

Here is, then, one way to describe the experience of reading literature. As we read, we construct contexts in which to make sense of potential embedments, and then we use the information that we derived from those embedments to construct contexts for subsequent embedments. We can think of this process of continuous and contingent construction as “participatory sense-making”—to borrow the term that De Jaegher and Di Paolo use to characterize mindreading involved in daily social interactions.<sup>93</sup> This means that by the time readers arrive to “Ought Margaret to know what Helen knew the Basts to know,” they already “have a history of interaction” with the novel,<sup>94</sup> that is, they have already been primed, by preceding embedments, to treat some contexts as more relevant than

others. Some of this priming has been planted (so to speak) by the author, but some has not.<sup>95</sup>

Let us take a closer look at the aspects of priming that are less predictable and thus fall outside the range of responses that I seek to capture with my hopeful “we.” So far I have described the construction of contexts for complex embedments as a forward-oriented process: with past embedments influencing embedments-to-come. But the “participatory sense-making” can move backward as well as forward. Something that we just read may trigger a complex embedment that hails from a preceding chapter, an embedment that has been lying dormant until now.

For instance, perhaps I did not pay much attention, the first time around, to the nuances of the narrator’s view of Tibby’s attitude toward his sister’s dilemma, yet, later, as I come across another social situation involving Tibby’s perspective of other people’s emotions, I may find myself realizing that the narrator has been feeling less than charitable toward this character for a while and thus retroactively revise the meaning of “Ought Margaret to know . . .” The cognitive literary scholar Anezka Kuzmičová describes this reverse sense-making in terms of “probes” that illuminate this or that aspect of our past reading experience. As she explains, “in reading long-form narrative . . . the number of verbal probes that can guide one’s grasp of the preceding text . . . is endlessly [high]. In light of this insight, it seems a mystery how any two people can ever come close to converging in their subjective experience of a story or novel.”<sup>96</sup>

Here is something to deepen this mystery yet further. Kuzmičová observes that “for many leisure readers” (a group that comes closest to my ideal of “good-enough” readers), “the added value of narrative lies . . . in momentarily becoming conscious of one’s self and one’s problems in specific ways that may be less readily available otherwise.” These “personal realizations inform consciousness” in a variety of ways. “Often enough, they may come in the form of propositional thought (‘Oh my, this character is acting just like me’). Just as often, however, they may assume the form of mental imagery,” feeding on “personal memories triggered by the narrative.” To psychologists, such associations are known to be “much more common in literary narrative compared to other types of reading materials.” Their frequency “directly affects the pleasure taken in reading. . . . It is in this sense that literature affords a unique form of self-consciousness, in

which you focus on yourself and yet you do not, because the story you are reading is really about others."<sup>97</sup>

To give you an example of this kind of unique (no *we* here—only *me*!) experience of embedding, imagine that I am reading the “Ought Margaret to know” passage at a particular juncture in my life at which I may feel a sharp pang of recognition by thinking of myself as a beleaguered Tibby surrounded by overbearing Helens. I would thus be more likely to construct an embedment in which I would give more weight to the nuances of that paragraph that portray Tibby with sympathy and compassion. The problem is that were that process to take place in a lab and were someone to ask me to map out the embedments of mental states present in this paragraph, I would not include any of my personal reflections into my report and instead would come up with something along the lines of, “The narrator *wants* his readers to *imagine* what it might have *felt* like for Tibby to grow up in a household where such relations were constantly discussed. By doing so, the narrator *wants* us to *recognize* Tibby’s *aversion* to such discussions as a healthy boundary-setting reaction.”

What does it mean for the experiment—for its accuracy, reliability, replicability, and so on—that my map would effectively bury that very important aspect of self-consciousness, in which I focus on myself and yet I do not, because the story I am reading is really about others?<sup>98</sup> At the very least, it means that we have to remember that, as any other literary-critical tool, our maps of embedment may conceal as much as they reveal about readers’ interactions with the text.

## 1.8 Enactive Embedments

Each instance of reading thus has its own unfolding history dependent on a uniquely situated reader—a particular person at this exact point in their life.<sup>99</sup> There is no predicting how a literary text will meet a reader at a given time; in what direction the probes “of consciousness will be thrust”,<sup>100</sup> which contexts will have more traction and which will have less; and, ultimately, what specific sequence of complex embedments the text and the reader will *jointly* create.<sup>101</sup>

This perspective on reading is congenial with the so-called enactive school of cognitive science, which emphasizes that mind is always constituted by

“organism-environment interactions.”<sup>102</sup> In particular, cognitive scientists committed to the enactive paradigm caution against thinking of mindreading as a form of problem-solving: one “detached individual trying to figure out the other.”<sup>103</sup> To be fair, cognitive literary critics, such as myself, have never approached mindreading as problem-solving. As I have emphasized on numerous occasions, mindreading takes place away from conscious access: it is too fast and intuitive and enmeshed with body language to be thought of along the formal lines of “figuring out the other.” Still, there *are* some occasions on which we would do well to heed that warning, and one such occasion is studying embeddings in the lab.

For think again about the first part of our experiment, in which we presented our subjects with context-free synthetic constructs such as “I am not even sure why Stephanie wants to go the movies with Alice and me.” Our expectations certainly conformed to the model of “detached individuals trying to figure out the other.” There was a “correct” answer associated with each sentence; the vignettes were to be decoded, and our subjects were the decoders.

In contrast, reading a work of literature does not reduce the text and the reader to those roles. Instead, reading can be described as a form of social interaction between the two autonomous agents (i.e., the text and its reader) that unfolds as they settle on a particular sequence of contexts for complex embeddings.<sup>104</sup> Unlike decoding, this enactive process of cocreation is less predictable, less likely to yield high agreement rates, and harder to study in the lab.<sup>105</sup>

Still, harder does not mean impossible, and there are some unexpected bonus points along the way. For instance, when my colleagues and I were working on tallying the data collected from our subjects, I noticed that cognitive scientists began to sound like literary critics, that, in fact, we *all* began to sound like participants in a literature seminar, avidly discussing motivations of characters, narrators, and readers in order to figure out why this or that sentence was assigned this or that level of embedding. As far as interdisciplinary projects go, a study of embedded mental states may thus be a particularly gratifying experience for a literary scholar, because it builds on the strengths of each participating discipline (e.g., I had to defer to my colleagues from the Haskins Laboratories for their expertise in brain-imaging techniques and statistical analysis) without losing sight of the complexity of the issues under consideration.<sup>106</sup>



Elfenbein has pointed out that scholars of literature have long been prone to “quick condemnation” of each other’s work as “reductive,” so it is “not surprising” that, given what they *think* psychologists are doing, they now accuse them “of the same perceived sin.”<sup>107</sup> What my experience suggests is that one way to put that stale prejudice to rest is to develop a collaborative project with one’s colleagues from cognitive science: to hear them talk your language and to attempt to understand theirs.<sup>108</sup> Then even those aspects of the project that would seem to point toward shortcomings of studying literature in the lab (such as a failure to obtain high agreement rates on the level of embedment in passages from a novel) may yield important insights about the participatory nature of literary mindreading.

### 1.9 Sitting Ducks

It seems that in literature certain types of explicitly spelled-out complex embedment—specifically those involving characters’ assertions about their own and other characters’ mental states—may function as sitting ducks. Just like “Ought Margaret to know what Helen knew the Bastis to know?” they may be set up to be upended by their contexts. So when we come across a sentence that reads suspiciously like one of my awkward mindreading maps—for example, “he thinks that she knows that he knows”—we may want to be on the lookout for implied embedments that would subvert these explicit ones.

Here are some examples of such subversion.

Recall Shirley Jackson’s “The Beautiful Stranger.” While its protagonist is happily interpreting the man’s smile as an indication that he knows that she has been worried that he may be her husband, after all (“She was *aware* from his smile that he had *perceived* her *doubts*”),<sup>109</sup> readers may have a reason to doubt her insight. Although it is possible that the husband has been (say) abducted by aliens and somewhat imperfectly replicated, which means that the smiling man *is* a beautiful stranger,<sup>110</sup> another explanation is that the protagonist has gone insane. If this is what we think is happening, then, even while we’re following the wife’s train of (embedded) thoughts, we are also aware that the implied author *wants* us to *consider* that she is *misinterpreting* the meaning of her husband’s smile.

My second example comes from Lara Vapnyar’s *Still Here* (2016), which tells several interlocking stories of Russian immigrants in the United States

during the dot-com bubble of the late 1990s. At one point in the novel, a recently-separated-from-her-husband woman named Vika meets an attractive stranger. He tries to start a conversation with her, but she rejects his overture outright because he strikes her as a variation on “type of Husband.” And, as we learn through a spelled-out embedment, she thinks that, right now, she needs something different: “A Husband knew her the way she didn’t want to be known, at her worst, her ugliest, her most embarrassing. . . . What she needed was a Lover.”<sup>111</sup>

Though not quite reaching, in its bluntness, the parodic level of Forster’s explicit embedment, Vapnyar’s “a Husband *knew* her the way she *didn’t want* to be *known*” still stands out in the sea of implied embedments surrounding it. These embedments include Vika reflecting with anguish on having disappointed a terminally ill patient in her care, who had hoped for an emotionally honest response from her; Vika deciding to visit the Metropolitan Museum of Art because she “truly enjoys” it and not because she cares whether other people think that she likes art; and readers beginning to suspect that Vika doesn’t realize how much she misses her estranged husband.

And so, perhaps, the real reason that Vika rejects the stranger is not that he is a “type of Husband” but that he is not *her* husband.<sup>112</sup> Of course, it will take many chapters before she becomes aware of that. Right now she is denied that intuition. All she has at her disposal is an explicit embedment, which is compact and expressive and almost aphoristic (“a Husband knew her the way she didn’t want to be known”) and, as such, provides Vika with a convincing (perhaps too convincing!) explanation of her own behavior.

Muriel Spark’s *The Girls of Slender Means* (1963) takes place in bombed-out London in the early summer of 1945. The novel centers on a group of young women living in the dormitory-style “May of Teck Club” and on their male admirers. In the following passage, the explicit embedments describe one of these men’s awareness of his thoughts, while the implied embedments make us wonder whether this self-awareness truly differentiates him from another, much less sympathetic character: “The Colonel seemed to be in love with the entire club, Selina being the centre and practical focus of his feelings in this respect. This was a common effect of the May of Teck Club on its male visitors, and Nicholas was enamoured of the entity in only one exceptional way, that it stirred his poetic sense to a point of exasperation, for at the same time he discerned with irony the process of his own thoughts, how he was imposing upon this society an image incomprehensible to itself.”<sup>113</sup>

Nicholas thinks of the girls from the May of Teck Club as beautiful and pathetic in their communal poverty and, moreover, glorying in their economic hardship. In their heroic penury, they are emblematic of war-torn England at its best. Of course, the girls themselves experience their poverty as a temporary evil that they can't wait to overcome. Not altogether blind to their perspective, Nicholas is *aware* that the girls *wouldn't recognize* themselves in his *vision* of them, or, as Spark puts it, Nicholas "*discerned with irony*" that "*he was imposing upon this society an image incomprehensible to itself.*"

Those are the spelled-out mental states. But then Spark also seems to *want us to suspect* a certain affinity between the *feelings* of Nicholas and an American colonel (especially since they both sleep with the same girl, Selina). This is not a pleasant comparison, for, the colonel is obtuse and philistine, while Nicholas is sensitive and sophisticated. Still, the way I see it, Spark won't let her readers off this hook. We wonder uncomfortably—that is, she *wants us to wonder*—if Nicholas's *self-awareness* is enough to mark him as different from other men who are "*in love with the entire club*" (as opposed to being attracted to one particular person) and thus displace onto it their sexual or political fantasies.

Hence a word of caution to a fictional character: thinking that you know well your own, or someone else's, mind—especially if you are spelling it out as a complex embedment—may not bode well for you.

### 1.10 Why Maps of Mental States Are Ugly

Bitter is the fate of the literary critic who has selflessly dedicated herself to pursuing the secret life of literature. Droning on that "*the implied author wants the implied reader to understand that this character doesn't know . . .*" does not endear her to her readers. And who can blame them? There is nothing appealing about such mindreading "maps." They are boring, repetitive, almost grotesque, and often hard to follow. They look pathetic next to the texts that they claim to represent. There is even a vague feeling of violence being inflicted on the elegant originals. The originals recover well (they have seen it all), but the critic may be stuck with the reputation of a plodding pedant.

Recent work by Max Van Duijn, Ineke Sluiter, and Arie Verhagen may help to explain why mindreading maps look so off-putting next to original texts. As these scholars suggest, by the end of the second act of *Othello*,

“the audience has to understand that Iago *intends* that Cassio *believes* that Desdemona *intends* that Othello *believes* that Cassio *did not intend* to disturb the peace.”<sup>114</sup> This looks, to me, like a very complex embedment (in fact, I would make it simpler, by scaling this map at least one level down), but imagine Shakespeare actually making Iago step forward and regale his audience with an aside in which he would say something along the lines of, “I want Cassio to believe that Desdemona intends that Othello . . .” and so forth.

Better to not imagine it. Not only would it sound unbearably stilted, but also, after a while, it would become “hard or even impossible for a reader or hearer to make the right inferences” about the characters’ intentions.<sup>115</sup> For, as the cognitive evolutionary psychologist Robin Dunbar and his colleagues have shown, “fifth-order intentionality” (fifth-level embedment of mental states) represents “a real upper limit for most people,” that is, the level at which their understanding of the situation worsens dramatically.<sup>116</sup> (Works of literature, I should add, do not often go the fifth level and higher. Extremely intricate social nuances can be conveyed on the third and fourth level of embedment. Even for such authors as Henry James—who, one may assume, would soar freely in the fifth-level empyrean—there is plenty to do on the third level. Shakespeare, Jane Austen, Muriel Spark, and Penelope Fitzgerald; Pushkin, Dostoevsky, Tolstoy, and Tatiana Tolstaya; Apuleius and Heliodorus, Cao Xueqin and Murasaki Shikibu, ply most of their unhumble trade on the seemingly humble third level.)

And so Shakespeare does not make Iago step forward and spell out his intentions as a sequence of embedded mental states. Instead, as Van Duijn and his coauthors explain, “narrative takes over”; that is, readers and viewers have at their disposal a number of “strategies characteristic of (literary) narrative discourse that support [their] ability to keep track of the [mental states] of characters.” These strategies supply “support and scaffolding for readers’ abilities to process [embedded mental states] by providing cues that prompt them to construct a fictional social network using mainly the same socio-cognitive skills as in real-life interaction.”<sup>117</sup>

To construct a map, we strip off this vital scaffolding. While embedded mental states in their natural environment are often implied, distributed over a paragraph or a scene, and embodied, we spell them out and force them into sentence-like propositions. “He thinks that she thinks that he

wants X”; “she remembers that she used to think that were X to happen, she would feel Y.” But who in their right mind would enjoy reading that kind of stuff? If a work of fiction is a living, breathing body, then a map of embedded mental states is a skeleton, with all the appeal and charm of a skeleton.

There is, thus, a good reason why writers themselves don’t let those bones stick out. “He thinks that she thinks that he wants X” may be what’s going on, but they do not put it that way. If they do, then, as we have seen with the “Ought Margaret to know what Helen knew the Bastis to know” example, it may be a joke, a parody, or a comment on someone’s lack of interest in social subtleties.

In the section that follows, I consider a fascinating case of the difference between the skeleton and the body. It shows that thinking on at least the third level of embedment is essential to the writing process, even if writers do not articulate it consciously to themselves. It so happened that this author (i.e., Patricia Highsmith) articulated it, but one can hardly hope to find many such examples in print. (Although, as I show immediately after, there *are* ways of finding other cultural contexts in which writers can be observed working through these issues.)

### 1.11 Do Writers Themselves Make Such Maps?

When the idea for a novel about a passionate love affair, between a gorgeous older woman and a young woman struggling to make it on her own in New York, occurred to Highsmith, she jotted in her diary the following description of the first meeting between the protagonists: “I see her the same instant she sees me, and instantly, I love her. Instantly, I am terrified, because I know she knows I am terrified and that I love her. Though there are seven girls between us, I know, she knows, she will come to me and have me wait on her.”<sup>118</sup> *I know she knows I am terrified. I know she knows I love her.* This is good enough for a map, so that the writer herself knows what’s going on in the scene, but it won’t do for an actual novel. Here is how this scene looks in Highsmith’s *The Price of Salt* (1952):

Their eyes met at the same instant, Therese glancing up from a box she was opening, and the woman just turning her head so she looked directly at Therese. She was tall and fair, her long figure graceful in the loose fur coat that she held open

with a hand on her waist. Her eyes were gray, colorless, yet dominant as light or fire, and caught by them, Therese could not look away. She heard the customer in front of her repeat a question, and Therese stood there, mute. The woman was looking at Therese, too, with a preoccupied expression as if half her mind were on whatever it was she meant to buy here, and though there were a number of salesgirls between them, Therese felt sure the woman would come to her. Then Therese saw her walk slowly toward the counter, heard her heart stumble to catch up with the moment it had let pass, and felt her face grow hot as the woman came nearer and nearer.<sup>119</sup>

If we map out this paragraph, we may come up with several third-level embedments. Some of them may even be similar to “I know she knows I love her” from Highsmith’s diary. But unlike those explicit embedments, the ones in *The Price of Salt* are implied. That is, they may still supply the underlying bone structure for the first encounter between Carol and Therese, but they are not anymore visible to the naked eye.

No wonder my own maps of embedded mental states—structured as strings of mentalizing verbs, such as “think” or “believe”—are destined to be clunky and off-putting. Although reading literature means reading mental states,<sup>120</sup> it seems we can only enjoy those mental states in context. Just as we, apparently, cannot absorb vitamins when we take them in the form of pills, “pure” mental states do nothing for us, except, after a very short while, irritate us. Highsmith’s desiccated embedments, “I know she knows I am terrified. I know she knows I love her,” may have a poetic ring to them. Still, they accrue a certain interest and cultural value (as, when a literary critic, such as myself, is thrilled to discover them in the writer’s diary) only because she has already seduced us with the text in which these mental states are *implied*.

Moreover, something else happened in the process of building up from the bare bones of “I know she knows I love her.” Other embedments came into being, those involving not just the main characters but the implied author and the implied reader and arising from the style of the narrative and its historical contexts. Observe, for instance, that, while Therese feels helplessly “caught” by the “light or fire” of Carol’s eyes, Carol, too, is powerfully compelled to come “nearer and nearer.” If one remembers that this dance of fatally attracted butterflies is taking place in 1952, one wonders if Highsmith wanted her audience to fear that her story would fall into the predictable 1950s pattern of depicting a lesbian love relationship as doomed.<sup>121</sup>

It is an interesting question at what point in the second part of the twentieth century the expectation of that particular doom faded. Or, to put it in

terms of our present discussion, at what point has it become possible *not to think* that the author *expects* that the reader would *assume* that a story about a love affair between two women cannot end well?

Our awareness of historical contexts is thus yet another factor to consider when we ask if implied mental states are already “in” a given text or are intuited into it by some readers but not others. I suggested earlier that complex embeddings arise as social situations built by a text are filtered through the unique consciousness of a particular reader. But, as the case of *The Price of Salt* shows, specific historical circumstances and their attendant ideologies also influence what kinds of implied mental states would be read into a text. A given reader’s awareness of the author’s stylistic choices—here, reference to “fire” that “catches” the hapless prospective lover—may alert them to intentionality behind the scene. However, their construction of the meaning of that intentionality—Is this a common poetic trope or a sign of danger? Is the protagonists’ relationship doomed because of their sexual orientation?—would reflect, among other things, their position in a particular historical moment.

### 1.12 What Do Writers Actually Say When They Talk about the Secret Life of Literature?

Several years ago, I enrolled in a graduate seminar in my university’s MFA program. My goal was to see if writers are aware of the “secret life of literature,” that is, if they are aware of the extent to which their texts depend on the constant embedment of complex mental states. That meant paying close attention both to our workshop discussions and to my own writing process, for, like other students in that class, I had to come up with two original short stories and have others comment on them.

Here is what I found, in brief. It is impossible to write fiction while thinking about embedding mental states, because the state of mind in which one puts oneself as a creative writer is different from that of a literary critic. But here are two important caveats.

First, even though I do not think *consciously* of embedding complex mental states when I am writing fiction, I, nevertheless, keep coming up with social situations that call for such embeddings. So one way to rephrase what I said earlier is to say that a creative writer puts oneself into a state of mind in which one produces complex embeddings without being aware of doing so.

Second, after the first draft is done and I start revising it, thinking consciously of ways to add yet another mental state to this or that social situation becomes helpful, to some degree. It seems that, in the process of revision, a writer begins to think like a critic or, at least, *more* like a critic than they did before.

As to whether writers talk about embedding complex mental states during their workshop discussions, the dynamic is similar. They are not familiar with this vocabulary and thus do not use these terms. Nevertheless, when they comment on each other's drafts, their suggestions for improvement tend toward making social situations present in the original more emotionally complex, which, of course, depends on cultivating complex embeddings of mental states.

While some of those suggestions center on characters, many involve various states of awareness between the implied reader and the implied author. Again, the "implied reader" and the "implied author" are not the terms writers use. They talk instead about texts, protagonists, narrators, authors, and readers. Thus, they may say, "The protagonist doesn't know it, but are we supposed to think that the text knows it?"<sup>122</sup> or "Even if the narrator is unsure what the story is about, the reader must sense that the author knows what the story is about, what it's doing."<sup>123</sup> Or, to quote from a workshop participant's written response to one of my stories, "The simplest way I can think of for this would be to utilize a third-person perspective so that the narrator could give us insight that the current narrator wasn't willing to. Or you could leave it in first person and just use the asides of the narrator to also give us possible suspicions that might be fleeting in her mind that she refuses to give much thought to."<sup>124</sup>

In other words, when writers are writing and revising/talking about their craft, they operate on a high level of embedding, even if they are not aware of it. Indeed, if my own experience is to be trusted, *consciously* focusing on embedding mental states is detrimental to all of these processes, although it is significantly more detrimental during the initial writing stages. The "secret life of literature" must remain secret even to the people who make literature happen.

Let me now show you how an author may use a feedback received from their peers to make a given social situation more emotionally complex, by bringing in more embeddings. Here are two excerpts—an original and a



revision—from one of my stories written for the workshop. The story features a middle-aged protagonist thinking back to a time when she was nineteen and she and her best friend, “Julia,” were in love with the same man, “Zhenia.” The man eventually chose Julia, and the protagonist remembers asking Julia about what the two of them did together: “I don’t wish to know where Julia and Zhenia go together and what they do. But, of course, I keep asking, and she tells me.”

There is already one complex embedment here. The protagonist is aware that she can’t stop herself from doing something that, she knows, will make her feel bad. But look what happened after I followed the advice given to me by the workshop participant who suggested highlighting the difference between the past and the present narrator, so that the asides of the present narrator can “give us possible suspicions that might be fleeting in her mind that she refuses to give much thought to”:

I didn’t wish to know where Julia and Zhenia went together or what they did. But, of course, I kept asking, and she kept telling. Today, I think it is odd that Julia didn’t seem to realize that it was painful for me to listen to those stories. But, perhaps, she did realize it, which was yet another sign that she had already given up on our friendship. I can say that now, knowing how quickly we were about to grow apart, in spite of my desperate attempts to hold on to her. At the time, however, I interpreted her behavior differently. It made sense to me that she would not think that I might be hurt by Zhenia’s choice. After all, I didn’t consider myself lovable either.

To map some of the new complex embedments structuring this passage, the older narrator *thinks* it is odd that Julia *did not think* that her friend would *feel bad* hearing her stories; the older narrator *wonders* if Julia *did know* that her friend would *feel bad* hearing her stories; the older narrator is *aware* that her younger self *was not willing* to *consider* that Julia *did not care* about their friendship anymore; the older narrator is *aware* that her younger self *believed* that no one could *love* her; and so forth.

I chose an excerpt from the revised version of my story that contains mostly spelled-out mental states in order to make this discussion more manageable. Initially I had wanted to give you a passage that contained no explicit mental states—only implied ones (which are often more interesting)—but then I realized that doing so would require supplying much more information about the story’s plot. Because explicit embedments often present on the level of individual sentences, while implied

ones may function on the level of paragraphs, chapters, and plots, explicit embedments are easier to demonstrate.

Here, then, are two key takeaway messages from my experience of taking an MFA course. First, the process of generating complex embedments without being aware of it, while writing, provides a useful insight into our reading practices. For there, too, mentalizing takes place mostly away from conscious access. The “felt experience of reading,” the cognitive literary scholar Elaine Auyong reminds us, is “distinct from the mental acts underlying it.”<sup>125</sup> To make sense of what we read, we constantly process complex embedments, yet if we pause and take a stock of doing so, the pleasure of reading may evaporate.

Second, we can now come back to the main claim of this book—which is that literature as we know it today cannot exist without embedding mental states on at least the third level—and add the following. Readers for whom this secret life of literature is *most fully present* (even if they do not think about it in those terms) are writers in the process of writing and revising. I dedicated several preceding sections of this chapter to figuring out if some readers are more immediately attuned to complex embedments in literature than others are. While that question mostly remains open, we can confidently say that there is at least one group highly attuned to such embedments, and these are writers when they are writing.

This view finds support in the work of Robin Dunbar, who has suggested that the reason that “good writers [are] so rare” is that they have to constantly keep in mind a higher-order intentionality than do readers.<sup>126</sup> Dunbar and I differ in one respect: I think that, both in our daily life and while reading literature, we operate on a somewhat lower level of embedment that he thinks we do. Thus, he writes that “in everyday social life, we probably don’t work at much beyond the third order most of the time,”<sup>127</sup> while I would say (along with Patricia H. Miller et al.) that we don’t work at much beyond the *second* level and rise to the third level only occasionally.<sup>128</sup> Similarly, Dunbar observes that writers “are among the very small proportion of individuals who can successfully cope with sixth and seventh order intentionality,”<sup>129</sup> while I think that literature can get plenty complex on the third level and expect that writers do not have to reach to such highs as sixth and seventh very often.

But those nuances notwithstanding, I find Dunbar’s argument that writers have to process more higher-level embedments than do readers

congenial to my argument that people who are most appreciative of high levels of embedment in literature are those in the process of creating those embedments. As Dunbar puts it,

When the audience ponders Shakespeare's Othello, for example, they are obliged to work at fourth order intentional levels. . . . [But whatever level of intentionality they are working on], Shakespeare himself is being forced to work at one level of intentionality higher, because he must intend that we (the audience) believe that Iago intends . . . , etc. . . . In effect, a successful story-teller has to be able to work at the very limits of normal adult competence in social cognition. The significance of this is perhaps best reflected in the contrast with the fact that, in everyday social life, we probably don't work at much beyond third order most of the time. . . . The need to be able to work at one or more orders . . . higher than the reader means that the story-teller has to be a rather unusual individual: they are among the very small proportion of individuals who can successfully cope with sixth and seventh order intentionality.<sup>130</sup>

Dunbar and I thus focus on different manifestations of the same phenomenon. He says that literature *sometimes* operates on the sixth and seventh level of intentionality; I say that it *constantly* operates on at least the third. As you can see, these claims are complementary rather than mutually exclusive. The bottom line is that we both think that literature turns up the volume on something fundamental to our everyday social functioning (i.e., mindreading) and that people who operate the dial are the ones who immediately feel the difference.

### 1.13 Bodies without Minds

"Constant" sounds a whole lot like "universal," which has a bad rap in literary studies, so let us face this issue squarely here. When I say that literature as we know it today cannot function without constantly embedding mental states on at least the third level, do I claim that the secret life of literature is, in effect, "universal"? And, if so, do I also claim that there are no exceptions to this unspoken "rule"?

To start with the second question first: of course, there are exceptions. (For instance, in the next chapter, I will show that some socialist-realist novels published in the Soviet Union and East Germany operated on a lower-than-third level of embedment.) This said, before we pronounce a particular text an exception, we'd better take a good look to make sure that

it actually is. In my experience, works of literature that leap to people's minds when they start searching their mental databases for exceptions, typically do not turn out to be such, upon closer inspection.

Patrick Colm Hogan provides a useful framework for thinking about exceptions in his work on "literary universals."<sup>131</sup> Perhaps, one day, what I call the secret life of literature will indeed be considered a literary universal, on the terms that he outlines, but I don't think we are there yet. At this point, it is still an empirical issue. This is to say that we'd do well to keep our mind open and continue checking for this pattern as we study literature from different cultural traditions and historical periods. I expect that social contexts of complex embedment would differ from one author, text, genre, and culture to another, and I think that sensitivity to those contexts is a more interesting and immediate research challenge than the adjudication of the question of universality.

Meanwhile, let us look at some texts that often figure as candidates for exception. What happens, for instance, when writers craft stories that contain no explicit references to mental states, for instance, when their characters seem to come across as lacking "psychology," "interiority," and "depth" or else live in a dystopian society that eschews any discussion of emotional life? More often than not, such stories still contain numerous complex embedments of mental states, but they are all implied. This is to say that readers have to do all of the heavy lifting associated with reading intentions into the behavior of characters and/or into stylistic choices of the author.

And readers do step up to that plate—for otherwise they wouldn't be able to make sense of what is happening in the story *or* appreciate its tone. Yet, ironically, even while they do that, they may continue to take at face value the text's claims (so to speak) to "mindlessness." Consider Evgeny Zamyatin's novel *We* (1921), set in a dystopian future where feelings are jettisoned for mathematical formulas. *We* has apparently fooled enough readers in several languages, because, when I give talks about complex embedment, it is one of the two novels (the other one being Alain Robbe-Grillet's *Jealousy*, which I will discuss later) almost inevitably brought up during the question-and-answer period as an example of a work of fiction that contains no mental states, much less any embedded ones.

Yet *We* constantly prompts us to construct embedded mental states to make sense of what is going on. Look at the first meeting of its protagonists, D-503 and I-330, narrated by D-503: "All this without smiling, I'd even say

with certain reverence (perhaps she knows that I'm a builder of the "Integral"). But I'm not sure—in her eyes or eyebrows—there is some strange irritating X, and I can't quite catch it, can't assign it a numerical expression."<sup>132</sup>

There is a whole constellation of complex embedments here. For instance, D-503 *wonders* if I-330 is *impressed* because she *knows* what he does. Also, he is *irritated* that he *can't fathom* her exact *attitude*. Moreover, the implied reader *understands* that D-503 *doesn't realize* that he's *falling in love* with I-330. The fact that we don't notice any of these and even may end up thinking of Zamyatin's novel as devoid of mental states testifies to the unreflective speed with which we attribute thoughts and feelings when we encounter behavior (more about this in chapter 5, on the history of complex embedment in literature).

Here is another example. It was suggested to me by a colleague sympathetic to the idea that, in eighteenth- and nineteenth-century Europe, writers heavily relied on complex embedment—what with all those thick courtship novels focused on characters' feelings! Modernists, too: just think of Proust's and Woolf's obsession with the multiply storied consciousness. But surely (so my sympathetic colleague thought), latter-day postmodernist authors have outgrown all that preoccupation with psychology. Take Cormac McCarthy's *Blood Meridian: Or the Evening Redness in the West* (1985). Its characters are notorious for their lack of interiority, which means we do not need to embed mental states as we follow their actions.

To see if this supposition is true, consider the opening of the novel. *Blood Meridian* tells the story of a nameless teenager, "the Kid," who joins a gang of scalp hunters terrorizing the border between the United States and Mexico in 1849–1850. We start by learning about the birth and upbringing of "the Kid":

See the child. He is pale and thin, he wears a thin and ragged linen shirt. He stokes the scullery fire. Outside lie dark turned fields with rags of snow and darker woods beyond that harbor yet a few last wolves. His folks are known for hewers of wood and drawers of water but in truth his father has been a schoolmaster. He lies in drink, he quotes from poets whose names are now lost. The boy crouches by the fire and watches him.

Night of your birth. Thirty-three. The Leonids they were called. God how the stars did fall. I looked for blackness, holes in the heavens. The Dipper stove.

The mother dead these fourteen years did incubate in her own bosom the creature who would carry her off. The father never speaks her name, the child does not know it. He has a sister in the world that he will not see again. He watches,

pale and unwashed. He can neither read nor write and in him broods already a taste for mindless violence. All history present in that visage, the child the father of the man.<sup>133</sup>

Looking at these three paragraphs, you can see why this novel may strike some readers as not featuring any thoughts and feelings. This is a far cry from, say, Marcel Proust's *Remembrance of Things Past*, in which a typical sentence embeds explicit mental states, as in, "Sometimes when, after kissing me, she opened the door to go, I longed to call her back and say to her 'Kiss me just once more,' but I knew that then she would at once look displeased, for the concession which she made to my wretchedness and agitation in coming up to give me this kiss of peace always annoyed my father, who thought such rituals absurd."<sup>134</sup> On the other hand, even though McCarthy's "Kid" doesn't seem to be able—in stark contrast to the little boy in Proust—to consider other people's feelings, McCarthy's prose achieves its uncanny effect by embedding mental states of the mysterious narrator, the implied author, and the reader.

For there is a very peculiar narratorial consciousness at work in these early paragraphs. McCarthy's narrator inserts himself in the story ("I looked for blackness, holes in the heaven") and starts making the case, as it were, against the Kid. First, by being born, the Kid murdered his own mother, though, admittedly, she was complicit in the crime. She "did," after all, "incubate in her own bosom the creature who would carry her off." There is another victim, too. The mother's death destroyed her husband, a former schoolteacher, a weak soul, who now "lies in drink," quoting from poets "whose names are now lost." The child "watches" his father—the word "watches" is repeated twice. He even "crouches" as he "watches": a little predator, in whom there "broods already a taste for mindless violence." The puzzling opening sentence now makes sense, too. "See the child," ladies and gentlemen of the jury, see the defendant on the stand.

*He* has known all along how it would turn out—the "I" of the second paragraph—the narrator who watched the heaven on the night the Kid was born. God-like he is, but also accomplished, in ways that only certain sophisticated readers would appreciate. He wants those readers to know that, unlike other riffraff populating the story, *he* recognizes the unintelligible sounds issuing from the drunk father as bits of forgotten poems. He also can cite from the poet whose name has not been forgotten—Wordsworth—and he does so, appropriately, to support his point: "the child the father of the man."

Thus, already in the first paragraphs of the novel, McCarthy *wants* his readers to *know* that the story will be told by a narrator who *is determined* to aggrandize himself and to condemn the Kid. Of course, we don't put it this way to ourselves, but to the extent to which we are aware of the strange tone of the opening, starting with "See the child," we are embedding the implied author's intentions. (To quote again one of the MFA workshop's participants, "fiction is a cohesive intentional work.")<sup>135</sup>

What it all adds up to is that *Blood Meridian* embeds complex mental states just as *Remembrance of Things Past* does, even if, in direct contrast to Proust's novel, *Blood Meridian* contains almost no explicit references to mental states. We embed implied intentions of the narrator and the author to make sense of the novel's *tone*—the crucial component of McCarthy's poetic prose.

#### 1.14 Minds without Bodies

But if some stories pretend to be "mindless" and thus make us work harder at reading mental states into their characters' body language, the opposite—that is, stories in which mental states are spelled out but there are no bodies behind them—is also possible. Consider Daniel Defoe's *Robinson Crusoe* (1719), whose protagonist regularly ponders intentions of "Providence," an entity that has landed him on a desert island: "These reflections made me very sensible of the goodness of Providence to me, and very thankful for my present condition, with all its hardships and misfortunes; and this part also I cannot but recommend to the reflection of those who are apt, in their misery, to say, 'Is any affliction like mine?' Let them consider how much worse the cases of some people are, and their case might have been, if Providence had thought fit."<sup>136</sup> As Crusoe imagines people who complain about their affliction, he *wants* them to *consider* that had Providence *thought* fit to land them in an even worse situation than they are currently in, it could have easily done so.

Crusoe is not alone thinking about various "secret intimations" of the "invisible intelligence."<sup>137</sup> Other fictional instances of such "intelligences" range in form from the karmic destiny of Cao's *Dream of the Red Chamber* to "Aubrey McFate" of Nabokov's *Lolita*. What such nebulous entities have in common is their apparent capacity for intentions and attitudes, which characters and readers try to fathom, all the while generating embedded mental states.

Here, for instance, is Mrs. Plinth, a well-heeled provincial lady from Edith Wharton's short story "Xingu" (1916). Mrs. Plinth can't help feeling keenly that the heavenly power that has made her rich intended for her the honor of hosting distinguished visitors, an honor currently usurped by another, less worthy lady, Mrs. Ballinger: "An all-round sense of duty, roughly adaptable to various ends, was, in her opinion, all that Providence exacted of the more humbly stationed; but the power which had predestined Mrs. Plinth to keep footmen clearly intended her to maintain an equally specialized staff of responsibilities. It was the more to be regretted that Mrs. Ballinger, whose obligations to society were bounded by the narrow scope of two parlour-maids, should have been so tenacious of the right to entertain [the current special guest]."<sup>138</sup>

Mrs. Plinth resents that Mrs. Ballinger refuses to acknowledge the intention of Providence, which wanted Mrs. Plinth to host distinguished visitors. Providence, apparently, is as invested in Mrs. Plinth's social success as it is willing to let some people, including Robinson Crusoe, to get away relatively scot-free, while smiting others. We may have come a long way from Apuleius's Venus and Cupid: divine entities that guide fictional characters have, nowadays, shed their bodies. But their social minds are as keen and active as ever, plotting and picking favorites among mortals.

### 1.15 One Body, Many Mental States

Let us stay with Robinson Crusoe a bit longer. If you want to know how many fictional characters one needs to start generating complex embeddings, the answer seems to be just one. A single character can embed enough mental states to sustain a three-hundred-page novel, as does Crusoe, who spends twenty-three out of his twenty-eight years on a desert island with nobody to talk to. (Friday joins him only at the tail end of his confinement.) His loneliness does not prevent him, however, from engaging in introspective musings such as this one:

From this moment I began to conclude in my mind that it was possible for me to be more happy in this forsaken, solitary condition than it was probable I should ever have been in any other particular state in the world; and with this thought I was going to give thanks to God for bringing me to this place.

I know not what it was, but something shocked my mind at that thought, and I durst not speak the words. "How canst thou become such a hypocrite," said I, even audibly, "to pretend to be thankful for a condition which, however thou mayest endeavour to be contented with, thou wouldst rather pray heartily to be delivered from?"<sup>139</sup>



This passage is typical for Defoe's novel, which demonstrates on every page ample narrative possibilities of the embedded consciousness of a solitary protagonist.<sup>140</sup> Crusoe *imagines* that he can be *grateful* to God for bringing him to a place where he can be *happier* than anywhere else in the world. But then he is *shocked* that he *would pretend* to be *grateful* for a condition that he would, in fact, *prefer* to escape. He accuses himself of becoming a hypocrite—"hypocrisy" being yet another cultural shorthand for a complex embedment, for a hypocrite *wants* to make others *think* that he or she has *beliefs* and moral standards that he or she, in fact, *does not have*.

Unlike Crusoe, the speaker of William Wordsworth's poem "Lines Composed a Few Miles above Tintern Abbey, On Revisiting the Banks of the Wye during a Tour" (1798) is not alone: accompanying him on his "tour" is his sister, Dorothy. Still, for most of the poem, he is thinking about the relationship among his various selves situated at different points in time, watching himself, for instance, to form impressions that, he knows, will influence him for years to come:

And now, with gleams of half-extinguished thought,  
With many recognitions dim and faint,  
And somewhat of a sad perplexity,  
The picture of the mind revives again:  
While here I stand, not only with the sense  
Of present pleasure, but with pleasing thoughts  
That in this moment there is life and food  
For future years.<sup>141</sup>

The speaker *imagines* his future self being made *happy* by *remembering* how *happy* he was here (by remembering, that is, his "present pleasure"). David Herman has described this literary dynamic as "distributed temporality,"<sup>142</sup> and we can see this interplay among the mental states of past, present, and future selves throughout "Tintern Abbey." Embedments arising out of a temporally distributed self can be encountered in any work of literature, but they may be particularly common in memoirs (be they prose or poetry, such as Nabokov's *Speak, Memory* or Wordsworth's *Prelude*) concerned with imagination and consciousness,

### 1.16 Many Bodies, One Shared Mental State

But if a single character can be a source of mental states embedded on the third and fourth level, the opposite is also true. A large group of characters

can share a single mental state—thus forming what the cognitive narratologist Alan Palmer calls an “intermental unit.”<sup>143</sup> Such an intermental unit can then be embedded within other mental states the same way as a mental state of just one character can be embedded within other mental states.

To illustrate this, here is another, typically self-reflexive sentiment of Crusoe, who begins by contemplating his own feelings and then turns to the thoughts of an intermental unit comprising, perhaps, millions of people: “But it is never too late to be wise; and I cannot but advise all considering men, whose lives are attended with such extraordinary incidents as mine, or even though not so extraordinary, not to slight such secret intimations of Providence, let them come from what invisible intelligence they will.”<sup>144</sup> Crusoe is thinking about the thoughts of, if not the whole of humankind, then a large part of it. He *wants* “all considering men” to *pay attention* to the *intentions* of Providence. This is as large a group of people as they come—a massive “intermental unit”—all sharing one mental state, which is embedded, in its turn, within the thoughts of the protagonist.

We also may want to take a look into the novelistic construction of crowds and ask how writers get around the problem of representing a large number of minds—fifty, a hundred, a thousand—numbers that would instantaneously take us outside our zone of cognitive comfort were we to try to imagine the mental states of those people one by one. It seems that authors can deal with this challenge in several ways. They may portray a crowd through two or three distinct personalities—the spokespeople who capture various points of view held by the multitude. Or they may depict a crowd as being of “one mind,” shouting or grumbling in unison. This, in turn, makes it possible for this unified “mob mind” to interact with two or three other distinct individuals, who respond to the mob’s concerns, so that the cumulative number of embedded mental states still stays within the comfortable range of four.<sup>145</sup>

Think, for instance, about the preelection scene in George Eliot’s *Middlemarch* (1871), which starts with Mr. Brooke, who is running for Parliament, giving a short speech in front of a large crowd of potential voters. In response to his claptrap, first one heckler and then another (“the invisible Punch”) make fun of him. Mr. Brooke, however, misunderstands their reactions, thinking that the second heckler intends to ridicule the first, until “a hail of eggs” directed at him and his effigy makes the crowd’s feelings abundantly clear:

“That reminds me,” [Mr. Brooke] went on, thrusting a hand into his side-pocket, with an easy air, “if I wanted a precedent, you know—but we never want a precedent for the right thing—but there is Chatham, now; I can’t say I should have supported Chatham, or Pitt, the younger Pitt—he was not a man of ideas, and we want ideas, you know.”

“Blast your ideas! we want the Bill,” said a loud rough voice from the crowd below.

Immediately the invisible Punch, who had hitherto followed Mr. Brooke, repeated, “Blast your ideas! we want the Bill.” The laugh was louder than ever, and for the first time Mr. Brooke being himself silent, heard distinctly the mocking echo. But it seemed to ridicule his interrupter, and in that light was encouraging; so he replied with amenity—

“There is something in what you say, my good friend” . . . here an unpleasant egg broke on Mr. Brooke’s shoulder.<sup>146</sup>

Readers may walk away with an impression that a multitude of “weavers and tanners of Middlemarch” have expressed their opinions about Mr. Brooke’s candidature,<sup>147</sup> when all we really have here are two distinct (if invisible) spokespersons and Mr. Brooke’s initially mistaken view of their attitude toward each other. Once the crowd’s minds have thus been compressed to a manageable number, we are ready to process the scene’s complex embedments of mental states and consider its meaning. For instance, as Eliot’s biographer Nancy Henry explains, a “crowd of [Middlemarch voters] detects and mocks the insincerity of Mr. Brooke’s commitment to reform.”<sup>148</sup> Or, to put it in terms of our present discussion, this crowd *knows* that Mr. Brooke only *wants* them to *think* that he *cares* about reform.<sup>149</sup>

### 1.17 Downgrade This!

As we are nearing the end of this chapter, let us revisit the issue of “simplifying” our descriptions of mental functioning, first brought up in the section on *Tom Sawyer*. For, I can still imagine a reader who thinks that it just *may* be possible to make sense of scenes that, as I claim, embed mental states on at least the third level while staying on the first or second level. To see what that would look like—that is, what downgrading the levels of embedment does to a story—let us take a look at three examples from classical Roman, Greek, and Japanese literature.

In Apuleius’s *The Golden Ass* (second century AD), a young widow learns that her beloved husband was treacherously murdered during a boar hunt

by the man who had long wanted her himself. Unaware that she knows about his perfidy, that man is now pressing the widow for marriage. She “pretend[s] to be won over” and suggests that they have a clandestine affair, “just until the year travels the full length of its remaining days,” at which point they would wed. She wants him to believe that she is eager to sleep with him yet is ashamed that people would think it unseemly for a new widow. So he agrees to come to her house late at night, muffled “from head to foot and bereft of [his] escort,” thus leaving himself vulnerable to her gory revenge.<sup>150</sup>

Let us see how much of this episode’s meaning is retained if we insist on scaling down its levels of embedment:

- “The widow is *eager* to sleep with the man who killed her husband.” This is one mental state, and you can decide for yourself how accurately it describes what is going on.
- “The man *thinks* that the widow is *eager* to sleep with him.” That’s two embedded mental states, and this configuration is still wrong, because it reflects only the limited perspective of the doomed character.
- “The widow *wants* the man *to think* that she *wants* to sleep with him” or “The widow *wants* the man *to think* that she *is afraid* of what people will say if she becomes his mistress so early into her bereavement.” Once we start operating on the third level, we, finally, begin to capture the complexity of the situation.

In Heliodorus’s *An Ethiopian Romance* (third century AD), an Egyptian priest, Calasiris, tells to his acquaintance Cnemon the story of the first meeting of the protagonists, Chariclea and Theagenes. During a public celebration at the altar of Apollo, Theagenes is supposed to receive a torch from a priestess (Chariclea) with which to light the altar piled with animal sacrifices. The surrounding crowd includes Chariclea’s adopted father, Charicles, who is, however, too busy right now to observe his daughter closely:

At first they stood in silent amazement, and then, very slowly, she handed him the torch. He received it, and they fixed each other with a rigid gaze, as if they had sometime known one another or had seen each other before and were now calling each other to mind. Then they gave each other a slight, and furtive smile, marked only by the spreading of the eyes. Then, as if ashamed of what they had done, they blushed, and again, when the passion, as I think, suffused their hearts, they turned pale. In a single moment . . . their countenances betrayed a thousand shades of feeling; their various changes of color and expression revealed the

commotion of their souls. These emotions escaped the crowd, as was natural, for each was preoccupied with his own duties; they escaped Charicles also, who was busy reciting the traditional prayer and invocation. But I occupied myself with nothing else than observing these young people.<sup>151</sup>

Calasiris *knows* that Charicles *doesn't know* that Chariclea and Theagenes are *falling in love* with each other. We may not articulate this to ourselves as we read the novel. But later, when Calasiris hatches a plot to help the young people elope together, it makes sense to us because it hinges on Calasiris's *knowing* that Charicles *doesn't know* that Chariclea *loves* Theagenes. Get rid of one of those levels of embedment and the elopement plot falls apart.

In Murasaki Shikibu's *The Tale of Genji* (eleventh century AD), shortly after Genji's mother's death, the emperor sends a messenger to the boy's grandmother, inviting her and Genji to the palace. Upon receiving the grieving emperor's letter, the grandmother talks to the messenger about what it means for her to have outlived her only daughter:

"Now that I know how painful it is to live long," she said, "I am ashamed to imagine what that pine must think of me, and for that reason especially I would not dare to frequent his Majesty's Seat. It's very good indeed of him to favor me with these repeated invitations, but I am afraid that I could not possibly bring myself to go. His son, on the other hand, seems eager to do so, although I am not sure just how much he understands, and while it saddens me that he should feel that way, I cannot blame him. Please let his Majesty know these, my inmost thoughts."<sup>152</sup>

Observe that, while declining the emperor's invitation, Genji's grandmother quotes from a poem, *Kokin rokujo* 3057, in which, as the translator, Royall Tyler, explains, "the poet laments feeling even older than the pine of Takasago, a common [lyrical] exemplar of longevity: 'No, I shall let no one know that I live on: I am ashamed to imagine what the Takasago pine must think of me.'"<sup>153</sup> The bereaved mother knows that the emperor will be pained by her refusal to visit him, and she *wants* him to *understand how she feels*. By evoking the poem (which is itself a third-level embedment of mental states: "I am *ashamed* to *imagine* what that pine must *think* of me"), she makes him aware of a somewhat unexpected nuance of her grief: shame. If the emperor considers that even a tree would reproach her for outliving her child, he would surely understand that she doesn't want to be seen by others, especially in a place to which people go with the purpose of being seen, such as the emperor's palace.

Try conveying any of this through lower-level embedments. “Genji’s grandmother *is thinking* about a poem” (one mental state) or “Genji’s grandmother *wants* the emperor to *recall* a famous poem” (second-level embedment) or “Genji’s grandmother *wants* the emperor to *pity* her” (also second-level embedment) all distort the meaning of what is going on. Until we start thinking on at least the third level—for instance, “Genji’s grandmother *wants* the emperor to *understand* that she is too *depressed* to make an effort to be seen by others”—our reading of the passage remains tone-deaf.

### 1.18 Can a Computer Program Tell the Difference between “Popular Fiction” and “Literature”?

Can one design a computer program that will count levels of embedment in a given sentence, paragraph, or chapter? The possibility of such a program has been mentioned to me on several occasions, with cautious enthusiasm by computer scientists and with dread by my colleagues from literary studies. I would be excited to see software for counting mental states in fiction because I suspect that it will fail and that its failure will be as illuminating as was the failure of various artificial intelligence projects in the 1950s–1970s.

The latter, as you may remember, alerted scientists to the unprecedented complexity of evolved human cognition. The machines could not replicate cognitive processes that came so easily to people that they hadn’t even been aware of them. Just so, by failing to register embedded mental states in literature, a computer program would illuminate cognitive processes that make reading literature possible and that we take completely for granted, such as a constant attribution of embedded mental states to characters, implied authors and readers, and narrators.

It will be particularly instructive if, in this case, the failure turns out to be selective. For I believe that a computer may be able to count embedments in some texts but not in others. That is, it may succeed with works of fiction that embed mental states of their characters *and* describe these mental states explicitly but not with those that embed implied mental states of characters, narrators, implied authors, and readers.

Consider this passage from John Irving’s novel *The 158-Pound Marriage* (1974): “‘I am going to get a lover,’ she said, ‘and I’m going to let you know about it. I want you to be embarrassed when you make love to me wondering if I am bored, if *he* does it better. I want you to imagine what I say that

I can't say to you, and what *he* has to say that you don't know."<sup>154</sup> I believe that one can indeed design a computer program that will do well with this novel. Make it pick such words as "want," "embarrassed," "wonder," "bored," and "imagine," and you will have a fairly accurate map of a given passage's embedment. "I *want* you to be *embarrassed* because you *wonder* if I am *bored*"—that's fourth-level embedment, and a computer may just be able to perform this calculation.

Now picture software faced with a sentence from Cao Xueqin's *Dream of the Red Chamber*, in which its female protagonist, Lin Dai-yu, reflects on her winsome cousin, Xue Bao-Chai: "And now suddenly this Xue Bao-chai had appeared on the scene—a young lady who, though very little older than Dai-yu, possessed a grown-up beauty and aplomb in which all agreed Dai-yu was her inferior."<sup>155</sup> What's going on in this sentence? Here is one way to spell out the mental states that we infer as we make sense of it: the narrator *wants* his readers to *realize* that Dai-yu *feels distressed* because she *is certain* that everyone around her *considers* her inferior to Bao-chai. That's at least four embedded mental states, but to articulate them, we have to take in subtle cues, such as the unhappy tone with which Dai-yu refers to her cousin. She calls her "a Xue Bao-Chai" (一個薛寶釵) or "this Xue Bao-chai" in David Hawkes's translation. The use of the pronoun "this" or "a" (yīgè) before a personal name is particularly important here, because it reflects Dai-yu's anguished sense of propriety. She can't say anything harsh or vulgar, so a vaguely dismissive "this" becomes an expression of her irritation and jealousy.

If we look for explicit references to mental states that this sentence contains, we notice the word rendered by the translator as "agreed" (wèi, 謂).<sup>156</sup> This word may describe an attitude of some people around Dai-yu, but the meaning of the passage does not reside with it. Instead, as we've seen, that meaning is expressed through embedded mental states implied but not stated by the text.

What will a computer do in this case? It may pick up on the word "agreed," but, as we have already seen, that word contributes little to the complex embedment present in the sentence. The problem is that a computer program cannot register *implied* mental states, much less figure out context-specific relationships that organize these mental states into embedments. Because in *Dream of the Red Chamber*, any word—including "a" and "this"—can create an implied embedded mental state, only a human mind, with its infinite sensitivity to contexts, can follow it.

But what about passages from *Dream* that spell out embedded mental states of its characters? After all, Dai-yu's diatribe about the look that Bao-yu gave to Xiang-yun, which I quoted in section 1.3 ("But what about that look you gave Yun? Just what did you mean by that? I *think* I *know* what you *meant*. You meant to warn her that she would cheapen herself by joking with me as an equal"), is not terribly different from Irving's "I want you to be embarrassed when you make love to me wondering if I am bored."

The difference between the two is that Cao's novel (as, indeed, other texts that we tend to put on our course syllabi) does this only occasionally. In contrast, *The 158-Pound Marriage* or, for that matter, Dan Brown's *Da Vinci Code*, Stephenie Meyer's *Twilight Saga*, or Danielle Steel's *Against All Odds* do it constantly. Computers will have a ball counting mental states in the fly-by-night favorites that spell out mental states of their characters and do not demand that their readers process implied mental states of narrators and implied authors.

Several strains of research in social and developmental psychology may bear on these issues. For instance, the social psychologist Emanuele Castano and his colleagues, working with theory of mind and fiction, suggest that "life-time exposure to literary fiction positively predicts attributional complexity, while exposure to popular fiction negatively predicts it."<sup>157</sup> (Psychologists use the term "attributional complexity" to describe motivation to seek complex explanations for human behavior, explanations that include, though are not limited to, mental states.)<sup>158</sup> Although Castano et al. are careful to observe that "literary and popular fiction foster different socio-cognitive processes and cognitive styles, all of which are important,"<sup>159</sup> the distinction between the two has been central to their research projects for a while.<sup>160</sup> Moreover, following up on Kidd and Castano's earlier studies, the cognitive neuroscientist Iris van Kуйjk and her colleagues have suggested that, compared "to popular fiction, reading literary fiction might encourage participants to process the meaning of words, sentences and their relationships more deeply and that might produce [theory of mind] differences."<sup>161</sup>

Literary critics may take issue with the term "popular," on several counts. For instance, they may object to the cognitive scientists' identification of popular fiction with "character-based" stories and the consequent exclusion of science fiction from the domain of the literary.<sup>162</sup> They are also aware of the slipperiness of the term, because historically, it is known



to have covered a broad range of texts, some of them straddling “the categories of literary, genre, and popular.”<sup>163</sup> Nevertheless, when we compare patterns of embedment in Irving, Meyer, Brown, and Steel with patterns of embedment that we encounter, say, in Cao, Tatiana Tolstaya, and Zadie Smith, the difference seems to be quite obvious. However you choose to call them—popular, genre fiction, mainstream, lowbrow—novels by Meyer, Brown, and Steel spoon-feed complex embedments to their readers, which must have an effect on those readers’ theory of mind that is different from the texts that require them to work at constructing them.

For instance, there are intriguing studies by developmental psychologists who have found that adding explicit references to thoughts, feelings, and intentions of characters in stories for young children does not promote their understanding of mental states.<sup>164</sup> I will discuss those studies in chapter 6 (i.e., on children’s literature), but, for now, I just want you to note that, even at a young age, *actively figuring out* implied mental states based on context seems to result in different sociocognitive outcomes than merely *being told* what this or that character thinks.

### 1.19 Conclusion: Close (Mind)Reading

If you are a teacher of literature, you may have noticed by now, particularly with the Cormac McCarthy example but also with the excerpts from Mark Twain, Cao Xueqin, E. M. Forster, and Patricia Highsmith, that the process of identifying embedded mental states in literature looks a lot like close reading—a “fundamental practice” of literary analysis, which consists of “examining closely the language of a literary work or a section of it.”<sup>165</sup> The reason that an inquiry into embedded mental states may end up as a close reading is that close reading is often an explication of mental states, those of characters, narrators, authors, readers, and other critics.

We do not think about it in these terms, but it is worth paying attention to. Next time you are developing a close reading with your students, pause and take a closer look at the embedment of mental states that you perform along the way. Conversely, think about passages that you tend to select for this kind of exercise. See if they tend to “promise” (something that experienced instructors learn to perceive at a glance) a discussion that is likely to embed complex mental states.

Of course, as Jonathan Culler observes, “there are all sorts of ways of achieving closeness in reading.”<sup>166</sup> These range from memorialization, translation into a foreign language, and inquiry into how culture shapes the meaning of the text to looking for “conflicts or tensions” within the text, which can be manifested by “ambiguous words, undecidable syntax, incompatibilities between what a text says and what it does, incompatibilities between the literal and the figurative, . . . and so on.”<sup>167</sup> Note, however, how integral attribution of complex mental states is to nearly all of those endeavors. Consider, for instance, translation as a (somewhat less popular, today) form of close reading.<sup>168</sup> Central to translating is figuring out what the author *meant* by this or that choice of word in the source language—and hence, which word would convey the author’s *intention* most accurately in a target language.<sup>169</sup>

Here is one way to think about the sociocognitive role played by all those various practices of achieving “closeness in reading.” It is as if it were not enough, for some of us, to merely process texts that continuously embed complex mental states. If we happen to live in what I have dubbed elsewhere a “culture of greedy mindreaders,” we may also join special communities for doing so.<sup>170</sup> Those communities reward their members (e.g., students, critics) who are adept at discerning complex intentionalities present in literature, prizing, in particular, intentionalities that are unexpected and yet plausible.

Far from being an isolated phenomenon, the omnipresence of complex embedment in literature is thus supported by a variety of cultural practices. Those practices seem to emerge in response to specific historical circumstances (as did the close scrutiny of textual “conflicts or tensions” in literary studies) and thus are not typically thought of as bound with the intricacies of our social cognition. In the chapters that follow, I will bring the two together. That is, I will show that the cognitive and the historical are inextricably connected in the case of complex embedment and that to understand why a particular work of fiction embeds mental states the way it does, we have to inquire into the political and cultural history of its creation and reception.

This is a section of [doi:10.7551/mitpress/13964.001.0001](https://doi.org/10.7551/mitpress/13964.001.0001)

# The Secret Life of Literature

By: Lisa Zunshine

## Citation:

*The Secret Life of Literature*

By: Lisa Zunshine

DOI: 10.7551/mitpress/13964.001.0001

ISBN (electronic): 9780262367653

Publisher: The MIT Press

Published: 2022

The open access edition of this book was made possible by generous funding and support from The MIT Press Frank Urbanowski Memorial Fund



The MIT Press

© 2022 The Massachusetts Institute of Technology

This work is licensed to the public under a CC-BY\_NC\_SA license.  
Subject to such license, all rights are reserved.



The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Zunshine, Lisa, author.

Title: The secret life of literature / Lisa Zunshine.

Description: Cambridge, Massachusetts : The MIT Press, [2022] | Includes bibliographical references and index.

Identifiers: LCCN 2021031221 | ISBN 9780262046336 (hardcover)

Subjects: LCSH: Literature—Psychological aspects. | Cognition in literature. | Psychology and literature. | Narration (Rhetoric)—Psychological aspects. | Discourse analysis, Literary—Psychological aspects.

Classification: LCC PN56.P93 Z86 2022 | DDC 801/.9—dc23

LC record available at <https://lccn.loc.gov/2021031221>