The Commotion of Souls

Lisa Zunshine

First, a couple of emotional dilemmas:

I love bringing my six-year-old to the Metropolitan Museum of Art when we are in New York in the summer. On Thursdays, they have a special hour for children. A curator first talks with them about an artwork and then encourages them to draw pictures inspired by it. My son seems to enjoy it. Yet every time I tell him that we are about to go to the MET, he says that he doesn’t want to. Don’t you remember, I plead with him, that you liked it last time? No, he says, he didn’t. I cajole and bribe, and keep hoping that a day will come when he will remember how he felt about it last week.

During a dissertation defense, I ask a question, and, as the candidate begins to answer, I realize that she must have misunderstood me. What she is saying is interesting, though. Should I just go with it, or should I restate my original query in different terms? I wonder, too, if other committee members think that she misunderstood the question or that she did understand it but didn’t know how to answer it and so decided to talk about something else.

A thought within a thought. A feeling within a feeling. A feeling within a thought within a feeling. I hope that my son will remember next week how he feels about the MET this week. I wonder if the other committee members think that the candidate intentionally chose not to answer a difficult question. I am sure that if I ask you to think about your day, you, too, may recall an occasion on which you “embedded” (or “nested”) thoughts and feelings in this recursive fashion. Or: I am sure you will recall an occasion on which you were thinking about other people’s thinking.

It’s difficult to say how often we do this, that is, how often we embed mental states within each other, especially since we don’t usually stop and think about it. On the one hand, many complex social situations seem to depend on this kind of cognitive construction. On the other hand, our days are not always chock-full of complex social situations, which means that we end up thinking about people’s thinking about people’s thinking relatively infrequently.

Fiction is where it gets interesting. Embedded mental states—a thought within a thought within a thought, or a feeling within a thought within a feeling—are everywhere in fiction. That is, they are everywhere in our experience of reading—as opposed to just being there in the text, immanent, intrinsic, unchanged by who opens the book and when. (The cognitive approach, as Ralf Schneider reminds us, “points to the utter variety of the cognitive and emotional activities triggered in readers who encounter beings in fictional worlds.”) For instance, a reader not attuned to Jane Austen’s use of free indirect discourse may accept as a given a particular novel’s unflattering view of a character’s feelings, while a different reader may recognize, with delight and amusement, the implied author’s intention to foreground another character’s unselfconscious bias toward the first character. But while the content of mental states may thus differ from one reader to another, what remains constant is the underlying structure: to make sense of what we read, we embed mental states.

How far can we take this claim? For the time being and until proven demonstrably wrong, I will take it as far as possible and say that without mental states consistently embedded on at least the third level, there is no fiction. That is, no novels, no short stories, no drama, no narrative poetry, and no memoirs concerned with imagination and consciousness, such as Vladimir Nabokov’s Speak, Memory or Maxine Hong Kingston’s The Woman Warrior. (Note that this list does not include storybooks for children under the age of three.)

Pick a book from your shelf and read one paragraph. Think of how you would tell your friend, who has not read it, about what’s going on in this paragraph. There is a good chance that as you do, you’ll find yourself explaining to your friend what a particular character wants other characters to think or feel. Or what an author wants her readers to think or feel: embedded mental states can be associated with characters, narrators, implied readers, and implied authors in an infinite variety of combinations. Although this is not an absolute rule, it seems that more often than not, embedded mental states in fiction start accumulating on the level of paragraphs.

They can also structure entire chapters or acts. For instance, Iago wants Othello to think that Desdemona is in love with Cassio. Romeo doesn’t know that Juliet merely wanted some people to think that she is dead. Odysseus wants the Trojans to think that the Greeks left behind the wooden horse because they hope to propitiate Athena for the desecration of her temple. Tom Sawyer doesn’t want his friends to realize that he hates whitewashing the fence. Sei Shonagon wants other courtiers to admire her poetic dexterity, but she doesn’t want them to think that she is flaunting what they would consider her (unladylike) knowledge of Chinese. Grushnitzki doesn’t know that Pechorin knows about his intention to humiliate him during their duel, by leaving Pechorin’s gun unloaded.

Sometimes individual sentences contain complex embeddings. Here’s one from Stephenie Meyer’s The Twilight Saga: “I tried very hard
not to be aware of him for the rest of the hour, and, since that was impossible, at least not to let him know that I was aware of him” (74). Here’s another, from the Old English poem “The Wanderer,” dated somewhere between the late sixth and the early tenth century:

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.

I wonder why I am not depressed when I think about death. In this example, and elsewhere, observe how difficult it is to separate “emotion” from “thought” when it comes to complex mental states, in poetry or one’s personal life. The term “mental state” encompasses both cognition and affect.

I quoted from The Twilight Saga and “The Wanderer” to give you examples of explicitly spelled-out mental states. Here is an example in which they are not mentioned at all, and the reader has to deduce implied thoughts and feelings to make sense of what’s going on. The first sentence of Zadie Smith’s On Beauty, “One may as well begin with Jerome’s emails to his father” (3), overflows with recursive intentions. 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 familiar with the opening of E. M. Forster’s Howards End, “One may as well begin with Helen’s letters to her sister” (3), will sense yet another set of intentions in Smith’s first sentence. The author wants her reader to know that the action will be filtered through the consciousness of a reflective narrator—and that she meant 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.

I don’t think we notice it, though. Were I to articulate my feelings upon first opening Smith’s novel, I’d 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 really 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.

In this respect, fictional embeddings are similar to those in our daily life. We are no more aware of nesting thoughts and feelings while reading a novel than we are aware of nesting thoughts and feelings while dealing with an actual social situation. That is, sometimes we articulate our feel-

ings to ourselves and others, as in, “I was wondering if you thought that she didn’t know how to answer that question,” but, quite often, we just act on them without bringing them to our conscious awareness.

So as I sit there, feeling a twinge of worry about the candidate’s performance and wondering if the other committee members are thinking that she has decided to sidestep a difficult question, I may be noticing that one of them just stopped looking at her iPad and is now facing the candidate directly and that another is tapping the desk with her fingers. Without being aware of doing so, I see their body language as indicative of their conjectures about the candidate’s intentions. I could be wrong—in fact, I am almost certainly wrong—but go tell that to my cognitive adaptations that evolved to read people’s bodies in terms of their mental states and that are, moreover, subject to egocentric bias (i.e., when one’s perspective is used as default in figuring other people’s perspectives)! Thus my next question to the candidate may be dictated by my intention to show my colleagues that they don’t need to assume that she didn’t know how to answer the earlier one, even though I’ve never explicitly spelled out to myself my concern about their assumptions.

Just so, when reading Michail Lermontov’s A Hero of Our Time, I am not perplexed when, in the middle of his duel with Grushnitzki, Pechorin asks their seconds to reload his gun. Nor am I bewildered by Pechorin’s cruelty, when after giving Grushnitzki another chance to recant (which he doesn’t take), Pechorin proceeds to kill him. I may have never explicitly articulated to myself the triply-embedded “Grushnitzki doesn’t know that Pechorin knows about his intention to humiliate him during their duel, by leaving Pechorin’s gun unloaded,” but it is this embedment (or one very close to it) that makes my current understanding possible.

I have more examples to offer, along with a discussion of patterns of embedment in fiction, but here is an important question that we must consider first. Do metaphors such as “embedding” and “nesting” actually reflect what’s going on in people’s brains? They are certainly visually compelling—bringing to mind an image of concentric circles, or, perhaps, of matryoshka dolls fit snugly within each other—and the view of the brain/mind that they imply should give us pause. Human social cognition is too messy and distributed to be accurately captured by these neatly spatial images. It is thus worth inquiring into the work of cognitive scientists to see if their research bears out our speculations about “nested” mental states.

The View from the Cognitive Sciences

1. Cognitive, Evolutionary, and Developmental Psychology

A starting point for talking about embedded mental states in fiction is, of course, research in cognitive psychology that deals with “theory of
mind,” a.k.a. “mindreading”—that is, our evolved cognitive tendency to see observable behavior as caused by unobservable mental states, such as thoughts, desires, feelings, and intentions. Mindreading is as fundamental a feature of our social life as it is flawed: we can’t help interpreting behavior in terms of mental states (e.g., “Is he frowning because he doesn’t like what I am saying?”) even though we have had myriad opportunities to learn that we are often wrong in our mindreading attributions (e.g., the person might be frowning because he just realized that he forgot his phone at home).

Our daily mindreadings (and misreadings) happen mostly below the level of conscious awareness. They are endlessly nuanced, fuelled by cultural stereotypes, power dynamics, and personal histories. Moreover, they are so profoundly context-dependent that any essentialist claim that some people are “better” at it than others makes no sense. For instance, studies have shown that people in weaker social positions may engage in more active and perceptive mindreading than people in stronger social positions. Interestingly, “when one is given the role of subordinate in an experimental situation, one becomes better at assessing the feeling of others, and conversely, when the same person is attributed the role of leader, one becomes less good.”

A prime example of essentialist thinking spurred by work on theory of mind is the view, still prevalent in some quarters, that autistics “lack” mindreading abilities. This view reveals more about neurotypicals’ own mind-reading biases—that is, their assumption that theory of mind must manifest itself in a particular, familiar, socially sanctioned way—than it does about the actual cognition of autistics. In other words, precisely because we begin to appreciate the explanatory power of the concept of theory of mind, we must be cautious about the results of our mind-readings. The unreflective speed and readiness with which we attribute mental states are likely to be consequences of the adaptive function of mindreading, but “adaptive” doesn’t translate into “accurate,” and the speed masks deep problems inherent in the process.

Given how maddeningly imprecise the terms “theory of mind” and “mindreading” are, as well as certain unattractive baggage that they have already accumulated, some scholars working with cognitive approaches to literature are looking for alternatives. For instance, cognitive narratologist Alan Palmer prefers to talk about “attribution theory” to show “how narrators, characters, and readers attribute states of mind such as emotions, dispositions, and reasons for actions to characters, and, where appropriate, also to themselves.” I admire Palmer’s work no matter what he calls the underlying cognitive process, but, for myself, I don’t find the baggage troublesome enough to shun the original terms.

Instead, when I disagree with some of the cognitive scientists’ claims about theory of mind, I say so in print and move on. “Theory of mind” and “mindreading” are far from perfect, but the phenomenon that they attempt to describe is so complex that I suspect in the long run that any descriptive label will fall short.

Fiction exploits the fact that on some level we don’t attend too closely to the difference between mental states of real and imaginary people: we interpret and misinterpret the behavior of fictional characters using the same cognitive adaptations for mindreading that we do when we interpret and misinterpret the behavior of people around us. Fiction builds on and experiments with our keen interest in social minds: in novels that we read and plays that we watch, mindreading patterns present in our everyday interactions are intensified and enhanced.

(From here on, the discussion of the view from the cognitive sciences will become somewhat technical. Readers not interested in details are advised to skip this section as well as the next one and go directly to the third, “What’s in a Name?”)

When it comes to embedded mental states, we have research of evolutionary psychologists, such as Robin Dunbar, who, in their work with adult subjects, have explored what they call “levels of intentionality” (which is equivalent to levels of embedment, or nesting), suggesting that “fifth-order intentionality,” as in, “I suppose that you believe that I want you to think that I intend,” represents “a real upper limit for most people,” that is, the level after which their understanding of the situation drops drastically.

Then there is also research in developmental psychology, which focuses on “doubly embedded representations” in children, that is, their “awareness not just that people have beliefs (and false beliefs) about the world, but that they also have beliefs about the content of others’ minds (i.e., about others’ beliefs), and similarly, these too may be different or wrong.” This ability arguably matures between five and seven years of age (in contrast, the ability to appreciate false beliefs matures around four years of age, or even earlier), and it is “fundamental to children’s . . . understanding of the epistemic concepts of evidence, inference, and truth” (Aystington et al, 133, 142).

Interestingly, in experiments involving kindergarteners and first-graders writing letters to hypothetical friends who have never experienced some of the things familiar to the authors of letters (such as, for instance, snow), the “recursive understanding of embedded mental states” was shown to be implicated with children’s growing awareness of a reader’s knowledge as distinct from that of the writer. Around seven years of age, children realize that “an effective writer represents how their reader will
interpret their textual meaning (authorial intention) in the light of that reader’s experience” (Peskin et al). Presumably, it is this maturing capacity for knowing that the other person may not know something that you know, which enables a broad range of literary experiences, including, but not limited to dramatic irony, both in books for young children (represented, for instance, by a variety of trickster stories) and for older readers.

2. Cognitive Neuroscience

The neuroscience of mindreading is a large, thriving area, but because I am interested in embedded mental states, I will focus here on research that explores brain regions involved in “thinking about thoughts.” As one recent review of brain connectivity methods has pointed out, “understanding complex social interactions among people who are presumed to be social, interactive, and emotive always involves the processing of self-reflective thoughts and judgments” (Li et al). For instance, handling communicative intentions is “a more complex process than simply thinking about intentions, since we have to recognize that the communicator is also thinking about our mental state. This involves a second-order representation of mental state. We have to represent the communicator’s representation of our mental state” (Frith). Brain areas involved in this kind of “social understanding of others” are associated with the medial prefrontal cortex, its various sub-regions contributing differently to this function (Li et al).

One influential early study, Rebecca Saxe’s “People Thinking About Thinking People,” has identified a region of the temporo-parietal junction “selectively involved in representation of other peoples’ mental states,” that is, a region showing “increased response to tasks/stimuli that invite theory of mind reasoning (about true or false beliefs) compared with logically similar non-social controls” (1836). Since then, other studies have addressed questions ranging from whether the same brain region supports thinking about people’s “appearance, social background, or personality traits” (it seems that it doesn’t!14 to what neural populations may underwrite the “representations underlying human emotion inference.”15

Then there is also the puzzling divergence between “recent advances in developmental psychology [which] suggest that children have some understanding of false beliefs much earlier than age 3 years, and initial neuroimaging studies of children’s brains [which] suggest that key maturation changes in the [right temporo-parietal junction] occur much later than age 5 years.”16

To account for this divergence, some scientists now propose a two-systems model of mindreading.17 It would consist of “‘low-level’ processes that are cognitively efficient but inflexible, and ‘high-level’ processes that are highly flexible but cognitively demanding” (Apperly 143). In this view, when we “make explicit judgments about what others think or want” (or want us to think), we rely on the “high-level” processes, but what “gets us through our social day” is a “combination of low-level mindreading processes and the rich endowment of social knowledge that we gain through development” (Apperly 155). The “two-system” proposal is likely to be vigorously debated in the coming years, but then so is everything else that we currently know about the attribution of “cognitive and affective mental states,” both to real people and to fictional characters.

3. What’s in a Name?

To sum up, what can we say about the cognitive underpinnings of recursively embedded mental states? According to the research from cognitive, evolutionary, and developmental psychology, as well as from cognitive neuroscience, the phenomenon under consideration matures in development, presents enough of a cognitive burden to have something resembling an upper limit set to it, and is supported by specific brain regions.

To return to my earlier question, which started the present excursion into the cognitive sciences—namely, whether the terms “embedment” and “nesting” capture any of the actual structures in the mind/brain—it seems that the answer should be no. In fact, the proliferation of terms—such as recursive embedment, perspective embedment, nesting, level-two perspective taking, second-order theory of mind, second-order false-belief understanding, levels of intentionality, multiple-order intentionality, and so forth—suggests that there is no particular need to put too much ontological stock into any of them.

It is worth noting that all these names converge on the image of layered, or leveled, structure, with some implication of hierarchical relationship among those layers. But that image may reflect the long cultural history of visualization18 as well as cognitive biases that shape our thinking, not any inherent properties of the brain’s structure. And so with embedment and nesting—my personal terms of choice—one has to remember that while the phenomenon that they attempt to describe is likely to be real, the descriptions themselves are likely to remain metaphorical.

But let’s say we keep firmly in mind Johanna Drucker’s warning that “visualizations are always interpretations” (7) and resist the interpretive potential of the image of layered hierarchical structure. We are still faced with another difficult question. To show that works of fiction rely on nested mental states for their meaning, I map out those mental states, and to do that, I reduce and amplify a sentence, a paragraph, or a scene under consideration. How, then, can I claim to observe and report
a certain underlying cognitive dynamic of our engagement with fictional texts, when my observation is clearly an act of interpretation, and thus must impact that dynamic?

I do not have a fully satisfying answer to this question. I can say that cognitive literary studies is not the worst nor only offender in this respect; that cognitive scientists who research “thinking about thoughts” are subject to similar pressures; that any attempt to explore a complex system with the tools supplied solely by that system will impact the results; that cognition itself is interpretation, and that our best hope is to be aware of it and regard what seems to be an “objective” observation with a healthy dose of skepticism. For, it seems that for cognitive literary analysis (no less so than for other, more traditional forms of literary criticism), the pleasures of interpretation are also its dangers.

Case Studies and Close Readings

I now offer you a series of embeds culled from novels spanning two thousand years: from ancient Rome and Greece and eleventh-century Japan, to eighteenth-century England, nineteenth-century France, and twentieth-century U.S. I chose each excerpt to demonstrate a particular feature of fictional embedment. Here is the brief rundown of these features:

- We can’t reduce a high-level embedment to a low-level one and still get an accurate sense of the meaning of the passage.
- The number of embedded mental states does not have to correspond to the number of characters.
- Although a story may seem to focus on “flesh-and-blood” characters, the mental states that we embed to make sense of what we read may belong to “disembodied” entities such as its narrator, fate, God, the implied author, and the implied readers, as well as the characters.
- A work of fiction may be conspicuously presented as not dealing with thoughts and feelings. The characters may be shown to lack “psychology,” “interiority,” “depth,” etc., or live in a society that eschews any discussion of emotional life. This does not change the fact that the only way in which readers can make sense of what’s going on is by embedding mental states.
- Mappings of embedded mental states aren’t pretty.

As you follow my case studies, you will recognize some of them as instances of close reading: a foundational technique of literary analysis and teaching. The reason that an inquiry into embedded mental states may end up as a close reading is that any close reading is an explication of mental states—those of characters, implied readers, and/or implied authors.

1. We can’t reduce a high-level embedment to a low-level one and still get an accurate sense of the meaning of the passage

In Apuleius’s The Golden Ass (second century A.D.), 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 presssing 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” (167), thus leaving himself vulnerable to her gory revenge.

Note that you can’t reduce third-level embedment to first- or second-level and still get the full meaning of the situation described by Apuleius. “The widow is eager to sleep with the man who killed her husband” is plain wrong. “The man thinks that the widow is eager to sleep with him” 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” begin to get there.

To follow this revenge plot, readers have to embed mental states of its protagonists. But The Golden Ass also contains plenty of situations in which, in addition to the mental states of characters, we also have to embed mental states of the implied author and the implied reader. For instance, when the goddess Venus learns that her son, Cupid, has ignored her order to humiliate and destroy Psyche (of whose beauty Venus was jealous), and has 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
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 third-level embedments, but they are not what make the passage hilarious.

What makes it hilarious is the interplay of mental states of the implied author and the implied 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” (xv). 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’ speech stems from this nested awareness.

When embedment is driven by style (here, parody) as opposed to content (see the earlier example from The Twilight Saga), there are often several ways to map it out. I just suggested, “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.”

Either way, your interpretation and hence mapping may differ drastically from mine. But to be plausible and non-reductive—that is, to reflect as accurately as possible what you perceive as this passage’s meaning—it has to embed mental states on at least the third level. First- or second-level embedments simply will not do justice to the complexity of Apuleius’s writing.

In Heliodorus’ An Ethiopian Romance (third century A.D.), 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 first meeting of the protagonists, Chariclea and Theagenes. During a public celebration at the altar of Apollo, Theagenes is supposed to receive a torch first from a priestess (i.e., 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 . . . (73)

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, the plot makes sense to us because it hinges on Calasiris’s knowing that Charicles doesn’t know that Chariclea loves Theagenes.

An Ethiopian Romance is full of stratagems aimed at deceiving people who are not aware of the true motives of others. A stepmother wants to punish a stepson who rejected her amorous advances. She tells his father that the young man has attacked her and that, prior to this, she has long admonished him about his intemperate behavior, without telling his father, because she didn’t want her husband to think that she disliked his stepson, as stepmothers are assumed to do: “I knew what his behavior was, but would not tell you, lest I be suspected of talking like a stepmother” (10). Although the lecherous woman succeeds in implicating her stepson, she herself is later set up by her servant who wants to “procure her own safety by ensnaring her mistress” (15). And so it goes, subplot after subplot involving characters manipulating other characters into believing that they know others’ true intentions.

In Murasaki Shikibu’s The Tale of Genji (1008), shortly after Genji’s mother’s death, the Emperor sends a messenger, a gentlewoman named Yugei no Myobu, to the boy’s grandmother, inviting her and Genji to the palace. Upon receiving the grieving Emperor’s letter, the grandmother talks to Myobu 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.’ (8)

There are plenty of explicit third-level embedments in this passage, but for me, the most interesting one is the one that is simultaneously explicit and implied. While declining the Emperor’s invitation, Genji’s grandmother quotes from Kokin rokujo 3057, in which, as the translator, Royall Tyler, explains, “the poet laments feeling even older than the pine of Takasago, a common poetic 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’”(8). The bereaved mother knows that the Emperor will be pained by her refusal to visit him, and she wants him to understand that her feelings of depression and hopelessness make it impossible for her to grant his request.

These are implied embedded mental states, but the poem that she evokes to convey them also contains an explicit embedment, “I am ashamed to imagine what that pine must think of me.” As in other works of fiction that integrate references to poetry with characters’ motivations, such as Sei Shonagon’s The Pillow Book and Cao Xueqin’s The Story of the Stone, the effect is cascading. As they process these passages, some readers may end up constructing additional embedments, which may involve the intentions of the implied author who speaks to a more exclusive group of readers who can appreciate the nuances of classic poetry and the sensibility of characters evoking it.

2. The number of embedded mental states does not have to correspond to the number of characters

There is, perhaps, no better text than Daniel Defoe’s Robinson Crusoe (1719) to illustrate the claim that a single character can nest enough mental states to sustain a three-hundred-page novel. Crusoe spends twenty-three years on a desert island with nobody to talk to (Friday joins him 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?’ (97)

Crusoe is shocked that he would pretend to be grateful for the condition that he would, in fact, prefer to escape. This passage is quite typical for Defoe’s novel, which demonstrates on every page ample narrative possibilities of the embedded consciousness of a solitary protagonist.

But if a single character can be a source of mental states nested on the third and fourth levels, the opposite is also true. A large group of characters can share a single mental state (thus forming what Palmer calls an “intermental unit”), which then can be embedded the same way as the mental state of just one character. Here is another, typically self-reflexive sentiment of Crusoe, who begins by contemplating his own feelings and then turns to the thoughts of “all considering men”:

Upon these and many like reflections I afterwards made it a certain rule with me, that whenever I found those secret hints or pressings of mind to do or not doing anything that presented, or going this way or that way, I never failed to obey the secret dictate; though I knew no other reason for it than such a pressure or such a hint hung upon my mind. I could give many examples of the success of this conduct in the course of my life, but more especially in the latter part of my inhabiting this unhappy island; besides many occasions which it is very likely I might have taken notice of, if I had seen with the same eyes then that I see with now. 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. (115).

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, whatever their perception of the source of those intentions may be. This is as large a group of people as they come—a massive “intermental unit”—all sharing one (embedded) mental state, which is embedded, in its turn, by our protagonist.

3. Although a story may seem to focus on “flesh-and-blood” characters, some of the mental states that we nest to make sense of what we read may belong to “disembodied” entities such as fate, Providence, God, and karma

Crusoe is not alone among fictional characters in pondering the thinking processes of the “invisible intelligence.” Other characters have grappled with the “secret intimations” of such “intelligences,” ranging in form from the karmic destiny of Cao’s The Story of the Stone 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 toathom, with varying degrees of success, all the while generating nested mental states.

Here is a brief example from Edith Wharton’s short story “Xingu” (1916), in which a well-heeled provincial lady, Mrs. Plinth, feels keenly...
that the heavenly power which 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 dearly 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]. (25)

Having explored nested mental states in “Xingu” elsewhere, I offer here only one brief possible mapping of this passage: Mrs. Plinth resents that Mrs. Ballinger refuses to acknowledge the intention of Providence, who wanted Mrs. Plinth to host distinguished visitors. “Providence,” apparently, is as invested in Mrs. Plinth’s social success as the invisible intelligence of Defoe’s novel was invested in teaching Crusoe a lesson. We may have come a long way from 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, hoping, and picking favorites among mortals.

4. A work of fiction may seem not to feature any thoughts and feelings

Some novels are conventionally thought not to contain any mental states—all the more so the embedded ones—novels whose characters are considered to lack “psychology,” “interiority,” “depth,” etc., or living in a society that eschews any discussion of emotional life. One such novel is Yevgeny Zamyatin’s novel We (1921). It has apparently fooled enough readers in several languages, because when I give talks, it is almost inevitably brought up during the question-and-answer period as an example of a work of fiction that contains no nested mental states. Yet as I have demonstrated elsewhere, We prompts us to construct embedded mental states to make sense of what is going on as much as any other novel. The fact that we are not aware of those mental states testifies, once again, to the unreflective speed with which we attribute thoughts and feelings when we encounter behavior.

Just as Zamyatin’s novel, Cormac McCarthy’s Blood Meridian: Or the Evening Redness in the West (1985) was offered up to me as a text that does not nest mental states. The colleague who brought it up felt that in eighteenth- and nineteenth-century England, writers had to rely on nesting—what with all those thick courtship novels focused on characters’ feelings!—but, surely, the later-day authors, such as McCarthy, bred on modernism and postmodernism, have outgrown such a reliance. In a mo-

ment, we will take a look at Blood Meridian, but, first, a short digression on suggestions.

I am always happy when other scholars come up with works of fiction that, they think, don’t embed thoughts and feelings. Apart from being grateful to them for providing me with good case studies (out of nine or so novels discussed in this essay, four were suggested to me in this fashion), I also think that theirs is an appropriate reaction to the grand claim that I make, which is that without mental states embedded on at least the third level, implied or explicit, there is no fiction. Of course, literary scholars should be skeptical and reach out for titles that might falsify this claim. What I find fascinating, and will merely report here without any comment, is the pattern underlying their examples.

There seems to be an unspoken consensus that modernist writers, such as Henry James, Marcel Proust, Virginia Woolf, and E. M. Forster, luxuriate in their nested mental states, but when you move far enough to the East or back enough in time, third-level nesting evaporates. A corollary to this view is that contemporary writers have outlived the need for embedment. To sum it up, a lot of fiction is either not sophisticated enough or too sophisticated to embed mental states. Writers from far removed times and places are apparently not sophisticated enough, while contemporary writers (such as McCarthy) are too sophisticated. I will leave to you to decide what to make of this unspoken consensus. If while reading this, you have already come up with some examples of your own, see if they fit the pattern.

Back to McCarthy’s 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. Here is the opening of the novel:

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 unashed. 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. (3)
Let’s look at these three paragraphs the way a first-time reader would. (To approximate this experience, I only read the first two pages of the novel before starting to analyze its opening.) On the one hand, you can see why this novel may strike some as not nesting any thoughts and feelings. This is a far cry from, say, Proust’s *Remembrance of Things Passed*, in which a typical sentence nests 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 . . .” (34).

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 nesting mental states of the mysterious narrator and the implied author.

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. 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 “incubate in her own bosom the creature who would carry her off.”

The puzzling opening sentence now makes sense, too. “See the child,” and McCarthy’s prose achieves its uncanny effect by nesting mental states of the mysterious narrator and the implied author.

He has known all along how it would turn out—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 riff-raff 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, very appropriately, to support his point: “the child the father of the man.”

Thus already in the first paragraphs of his 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 nesting the implied author’s intentions.

What it all adds up to is that *Blood Meridian* embeds mental states as much as *Remembrance of Things Passed* does, even if, in direct contrast to it, *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.

5. Mappings aren’t pretty

Here’s something odd. I mapped the first sentence of Zadie Smith’s *On Beauty* as follows: “The author wants her readers to know that the action will be filtered through the consciousness of a reflective narrator.” I can map the three first paragraphs of *Blood Meridian* in almost the same way: “The author wants his readers to know that the action will be filtered through the consciousness of a very peculiarly minded narrator.” What does it mean that the openings of novels as drastically different as Smith’s and McCarthy’s seem to share the same underlying structure when it comes to their embedded mental states?

It means that the map is not the territory and should not be treated as such. Mapping nested mental states is an important critical exercise because it alerts us to the underlying structure of fiction and opens a productive conversation about similarities and differences between the real-life and fictional attribution of mental states. But there is nothing appealing about the mappings themselves. They are boring, repetitive, almost grotesque, and sometimes hard to follow.

For instance, as Max Van Duijn, Ineke Sluiter, and Arie Verhagen have shown, by the end of Act II of Shakespeare’s *Othello*, “the audience has to understand that Iago intends that Cassio believes that Desdemona believes that Cassio did not intend to disturb the peace” (148; italics in the original). However, if this representation of embedded mental states relied on such sentences, it would soon become “hard or even impossible for a reader or hearer to make the right inferences” about the characters’ intentions (151).

Instead, in Van Duijn, Sluiter, and Verhagen’s elegant formulation, “narrative takes over,” that is, readers 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” (149). 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” (153).

In our analysis, we strip off this vital scaffolding. While in their natural environment nested mental states are often implied and distributed over the text, 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 her 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. “She knows that he knows that she knows” may be what’s going on, but they do not put it that way. If they do, then, more often than not, it’s a joke, a parody, or a comment on someone’s lack of interest in social subtleties. For instance, Forster’s *Howards End* contains the following bare-bones sentiment: “Ought Margaret to know what Helen knew the Basts to know?”, but the character who is spouting this crude nesting is Tibby, a young man, who, the narrator hastens to inform us, is bored by “personal relations” (254).

Here is an interesting case of the difference between the skeleton and the body. 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 Patricia 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. (quoted in Schenkar 270).

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. (27)

If we map out this paragraph, we may come up with several third-level embeddings. Some of them may even be similar to “I know she knows I love her” from Highsmith’s diary. But unlike those explicit embeddings, 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.

Meanwhile, something else happened in the process of building up from the bare bones of “I know she knows I love her.” Other embeddings came into being, those involving not just the main characters, but the implied author and the implied reader. For instance, does Highsmith want her readers to think 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”? Moreover, I catch myself wondering whether, to someone reading in 1952, this dance of fatally attracted butterflies might have indicated that Highsmith wanted her audience to fear that her story would fall into the predictable 1950s pattern of depicting a lesbian love relationship as doomed.

The gossamer thread of such thoughts reminds us again why embedded mental states in fiction are emergent rather than immanent. Because fictional embeddings are generated by style, genre, and ideology, but also by the history of reading and the individual perspective, the tension between the text and its map will never be resolved.

**Conclusion: What Is Truly Exciting?**

When I talk about embedded mental states with my colleagues in literary studies, they sometimes wonder about my focus on the third level, as opposed to higher and thus more challenging and presumably more exciting levels. As one anonymous reader has put it, “even if one agrees that triply-nested mental states are pervasive in fiction, pervasiveness and importance are very different things. It is perhaps what fiction does rarely, intermittently, and unexpectedly that defines its cognitive potential—not what it does all the time.” This is a fair objection, so let’s consider it in some detail.

First of all, one can certainly make a fruitful study of higher levels of embedment in fiction. I have done so myself on several occasions, looking at the fifth level of embedment in the prose of Virginia Woolf as well as at even more spectacular—perhaps sixth!—level of embedment in *Restoration Comedy*. The latter case is particularly interesting because it appears that on stage, to adapt Van Duijin, Sluiter, and Verhagen’s phrasing, bodies “take over.” That is, the body language of actors may contribute to the narrative scaffolding, facilitating viewers’ comprehension of high-level nestings.

Second, I don’t want to be misunderstood as saying that fiction relies only on third-level embeddings. In fact, many of my examples, here and
defines” fiction’s “cognitive potential” is that, immersed in it, we can
work of fiction will not lodge itself into our consciousness unless we spend
simulated otherwise. It’s really quite incredible if you think about it. A
requiring “pervasive” third-level embedment, which would not be as
experiences, which rely on low-level nesting, “social scripts and schemas,
ing of this essay rather stand out amidst the majority of my daily social
The triply embedded emotional dilemmas that I described in the begin
knowledge that we gain through development” (155; emphasis added).
quote Apperly again, what “gets us through our social day” is a “combi
don’t construct it at the drop of a hat if a first-level nesting will suffice. To
new things about it?

isn’t the visual system an amazing feat of evolutionary engineering, end
features is an important step toward understanding it.
embedment of mental states on at least the third level as one of its key
mindreading is thus its own unique phenomenon, and recognizing the
stylistic tools for embedding mental states. We don’t just push aside those
these adaptations, having amassed a repertoire of extremely nuanced
cognitive adaptations for mindreading, fiction has also run away with
benefits,
effect, it may improve one’s vocabulary, which has its own cascading
of reading fiction is that it makes one a better reader of fiction. As a side
study literature. To put it starkly, it’s quite possible that the main effect
desire to read more novels or the decision to go to a graduate program to
them at a greater frequency?

And, moreover, is it even true that “pervasiveness and importance are very different things”? It’s certainly not true in the case of many other
cognitive endowments. For instance, our ability to see is pervasive; but
isn’t the visual system an amazing feat of evolutionary engineering, end-
lessly important, and appearing more impressive as we keep learning new
things about it?

Just so, a third-level nesting is not a cognitive commonplace. We
don’t construct it at the drop of a hat if a first-level nesting will suffice. To
quote Apperly again, what “gets us through our social day” is a “combi-
nation of low-level mindreading processes and the rich endowment of social
knowledge that we gain through development” (155; emphasis added).
The triply embedded emotional dilemmas that I described in the begin-
ning of this essay rather stand out amidst the majority of my daily social
experiences, which rely on low-level nesting, “social scripts and schemas,
and the normative principles” (Apperly 129).

And yet we have tremendous cultural repositories of information
requiring “pervasiveness” third-level embedment, which would not be as-
similated otherwise. It’s really quite incredible if you think about it. A
work of fiction will not lodge itself into our consciousness unless we spend
hours awash in implied and explicit third-level embedments. To me, what
“defines” fiction’s “cognitive potential” is that, immersed in it, we can
spend hours on end embedding mental states on the third level and enjoy
it, and not that once in a while we also process a sixth-level embedment.
Just because the “rare” is possible, there is no reason to take what fiction
“does all the time” for granted.

One can’t help wondering whether sustained exposure to the intensi-
sified mindreading offered by fiction has any long-term impact on our
social life. For instance, if we read a lot of novels, do we become more
attuned to triply embedded mental states in our immediate social envi-
rонment and begin to seek situations that would allow us to experience
them at a greater frequency?

It seems to me that we don’t, unless you count as such an impact the
desire to read more novels or the decision to go to a graduate program to
study literature. To put it starkly, it’s quite possible that the main effect
of reading fiction is that it makes one a better reader of fiction. As a side
effect, it may improve one’s vocabulary, which has its own cascading benefits,
but it does not translate into superior mindreading skills in
daily social interactions.

Nor should it, if you think about it. While building on our evolved
cognitive adaptations for mindreading, fiction has also run away with
these adaptations, having amassed a repertoire of extremely nuanced
stylistic tools for embedding mental states. We don’t just push aside those
stylistic trimmings to get to the real meat of social mindreading. Literary
mindreading is thus its own unique phenomenon, and recognizing the
embedding of mental states on at least the third level as one of its key
features is an important step toward understanding it.

University of Kentucky

Notes
1. “The Cognitive Theory,” 130. See also Schneider’s “Toward a Cognitive Theory” for a
discussion of differences in perception between “expert and nonexpert readers” (612-13).
2. For a discussion of what fiction may look like when it doesn’t embed mental states above
the second level, see Zunshine, “For I am a Bear.”
3. For a discussion of narrative empathy as “an aspect of literary cognitivism’s project,”
see Keen, 347. See also Hogan, Affective Narratology.
4. Snodgrass, 49; quoted in de Vignemont.
5. See Savarase and Zunshine.
7. For an example of such a disagreement, see Savarase and Zunshine, 21-26.
8. See Zunshine, Why We Read Fiction and Getting Inside Your Head.
9. How many, 180. See also Whalen et al, “Increases” and Whalen et al, “Validating,”
11. For a review, see Apperly, 11-34. See also Miligan et al.
15. Skerry and Saxe, 7.
18. See Drucker, 64-137.
20. Compare to other translations, e.g., Seidensticker’s: “Ashamed before the Takasago pines, I would not have it known that I still live” (9), or Waley’s: “Though I know that long life means only bitterness, I have stayed so long in the world that even before the Pine Tree of Takasago I should hide my head in shame” (9).

21. See Zunshine, “From the Social.”
25. Note that Van Duijn, Sluiter, and Verhagen use the term “multiple-order intentionality” (149) rather than “embedded mental states.”
27. See Zunshine, Why We Read Fiction and “Why Jane Austen,” 287-89.

Works Cited


