

Lisa Zunshine



Theory of Mind and Experimental Representations of Fictional Consciousness

Let me begin with a seemingly nonsensical question. When Peter Walsh unexpectedly comes to see Clarissa Dalloway “at eleven o’clock on the morning of the day she [is] giving a party,” and, “positively trembling,” asks her how she is, “taking both her hands; kissing both her hands,” thinking that “she’s grown older,” and deciding that he “shan’t tell her anything about it . . . for she’s grown older” (40), how do we know that his “trembling” is to be accounted for by his excitement at seeing his Clarissa again after all these years, and not, for instance, by his progressing Parkinson’s disease?

Assuming that you are a particularly good-natured reader of *Mrs. Dalloway*, you could patiently explain to me that if Walsh’s trembling were occasioned by an illness, Woolf would tell us so. She wouldn’t leave us long under the impression that Walsh’s body language betrays his agitation, his joy, and his embarrassment, and that the meeting has instantaneously and miraculously brought back the old days when Clarissa and Peter had “this queer power of communicating without words” because, reflecting Walsh’s “trembling,” Clarissa herself is “so surprised, . . . so glad, so shy, so utterly taken aback to have [him] come to her unexpectedly in the morning!” (40). Too much, you would point out, hinges on our getting the emotional undertones of the scene right for Woolf to withhold from us a crucial piece of information about Walsh’s health.

I then would ask you why it is that were Walsh’s trembling caused by an illness, Woolf would have to explicitly tell us so, but as it is not, she can simply take for granted that we will interpret it as being caused by his emotions. In other words, what allows Woolf to assume that we will automatically read a character’s body language as indicative of his thoughts and feelings?

Lisa Zunshine is Assistant Professor of English at the University of Kentucky, Lexington. She has edited *Nabokov at the Limits*, coedited *Approaches to Teaching the Novels of Samuel Richardson*, and published essays in *Poetics Today*, *Philosophy and Literature*, and *Eighteenth-Century: Theory and Interpretation*.

NARRATIVE, Vol. 11, No. 3 (October 2003)
Copyright 2003 by The Ohio State University

She assumes this because of our collective past history as readers, you perhaps would say. Writers have been using descriptions of their characters' behaviors to inform us about their feelings since time immemorial, and we expect authors to do so when we open the book. We all learn, whether consciously or not, that the default interpretation of behavior reflects the character's state of mind, and every fictional story that we read reinforces our tendency to make that kind of interpretation first.¹

Had this imaginary conversation about readers' automatic assumptions taken place twenty years ago, it would have ended here. Or it would have never happened—not even in this hypothetical form—because the answers to my naïve questions would have seemed so obvious. Today, however, this conversation has to go on because recent research in cognitive psychology and anthropology has shown that not *every* reader can learn that the default meaning of a character's behavior lies with the character's mental state. To understand what enables most of us to constrain the range of possible interpretations, we may have to go beyond the explanation that evokes our personal reading histories and admit some evidence from our evolutionary history.

In what follows, then, I attempt to make a broader case for introducing the recent findings of cognitive scientists into literary studies by showing how their research into our ability to explain behavior in terms of the underlying states of mind—or our *mind-reading* ability—can furnish us with a series of surprising insights into our interaction with literary texts. I begin by discussing the research on autism that alerted cognitive psychologists to the existence of the cognitive capacity that enables us to narrow the range of interpretations of people's behavior down to their mental states, and that makes literature, as we know it, possible. I then consider the potentially controversial issue of the “effortlessness” with which we thus read other people's—including literary characters'—minds. To explore one specific aspect of the role played by such mind-reading in fictional representations of consciousness, I then return to *Mrs. Dalloway*. Here I describe a series of recent experiments exploring our capacity for imagining serially embedded representations of mental states (that is, “representations of representations of representations” of mental states)² and suggest that Woolf's prose pushes this particular capacity beyond its everyday “zone of comfort,” a realization that may account partially for the trepidation that Woolf's writing tends to provoke in some of her readers. I conclude by addressing two issues concerning the interdisciplinary potential of the new field of cognitive approaches to literature. First, I discuss the relationship between cognitive analysis and the more traditional literary-historical analysis of Woolf. Second, I suggest that literary critics should take a more proactive stand toward cognitive scientists' increasing tendency to use literature in their study of human cognition.

I. THEORY OF MIND AND AUTISM

Mind-reading is a term used by cognitive psychologists to describe our ability to explain people's behavior in terms of their thoughts, feelings, beliefs, and desires; for example, “Lucy *reached* for the chocolate because she *wanted* something sweet,”

or “Peter Walsh was *trembling* because he was *excited* to see Clarissa again.” They also call this ability our Theory of Mind (ToM), and I will use the two terms interchangeably throughout this essay.

This proliferation of fancy terminology adds extra urgency to the question of why we need this newfangled concept of mind-reading or ToM to explain what appears so obvious. Our ability to interpret the behavior of real-life people—and, by extension, of literary characters³—in terms of their underlying states of mind seems to be such an integral part of being human that we could be understandably reluctant to dignify it with a fancy term and elevate it into a separate object of study. Indeed, the main reason that ToM has received the sustained attention of cognitive psychologists over the last twenty years is that they had come across people whose ability to “see bodies as animated by minds” (Brook and Ross 81) was drastically impaired—people with autism. By studying autism and a related constellation of cognitive deficits (such as Asperger syndrome), cognitive scientists and philosophers of mind began to appreciate our mind-reading ability as a special cognitive endowment, structuring in suggestive ways our everyday communication and cultural representations.

Most scholars working with ToM agree that this adaptation must have developed during the “massive neurocognitive evolution” which took place during the Pleistocene, when our brain increased threefold in size. The determining factor behind the increase in brain size was the social nature of our species (which we share with other primates).⁴ The emergence of a ToM “module” was evolution’s answer to the “staggeringly complex” challenge faced by our ancestors, who needed to make sense of the behavior of other people in their group, which could include up to two hundred individuals. In his influential 1995 study, *Mindblindness: An Essay on Autism and a Theory of Mind*, Simon Baron-Cohen points out that “attributing mental states to a complex system (such as a human being) is by far the easiest way of understanding it,” that is, of “coming up with an explanation of the complex system’s behavior and predicting what it will do next” (21).⁵ Thus our tendency to explain observed behavior in terms of underlying mental states seems to be so effortless and automatic because our evolved cognitive architecture “prods” us toward learning and practicing mind-reading daily, from the beginning of awareness. (This is not to say, however, that our actual interpretations of other people’s mental states are always correct—far from it!)

Baron-Cohen describes autism as the “most severe of all childhood psychiatric conditions,” one that affects between approximately four to fifteen children per ten thousand and that “occurs in every country in which it has been looked for and across social classes” (60). Although “mind-reading is not an all-or-none affair [since]. . . [p]eople with autism lack the ability to a greater or lesser degree” (Origgi and Sperber 163), and although the condition may be somewhat alleviated if the child receives a range of “educational and therapeutic interventions,” autism presently remains “a lifelong disorder” (Baron-Cohen 60). Autism is highly heritable,⁶ and its key symptoms, which manifest themselves in the first years of life, include the profound impairment of social and communicative development and the “lack of the usual flexibility, imagination, and pretence” (Baron-Cohen 60). It is also

characterized—crucially for our present discussion—by a lack of interest in fiction and storytelling, differing in degree, though not in kind, across the wide spectrum of autism cases.

In his book *An Anthropologist on Mars*, Oliver Sacks describes one remarkable case of autism, remarkable because the afflicted woman, Temple Grandin, has been able to overcome her handicap to some degree. She has a doctorate in agricultural science, teaches at the University of Arizona, and can speak about her perceptions, thus giving us a unique insight into what it means not to be able to read other people's minds. Sacks reports Grandin's school experience: "Something was going on between the other kids, something swift, subtle, constantly changing—an exchange of meanings, a negotiation, a swiftness of understanding so remarkable that sometimes she wondered if they were all telepathic. She is now aware of the existence of those social signals. She can infer them, she says, but she herself cannot perceive them, cannot participate in this magical communication directly, or conceive of the many-leveled, kaleidoscopic states of mind behind it" (272).

Predictably, Grandin comments on having a difficult time understanding fictional narratives. She remembers being "bewildered by *Romeo and Juliet*: 'I never knew what they were up to'" (259). Fiction presents a challenge to people with autism because in many ways it calls for the same kind of mind-reading as is necessary in regular human communication—that is, the inference of the mental state from the behavior.

To compensate for her inability to interpret facial expressions, which at first left her a "target of tricks and exploitation," Grandin has built up over the years something resembling a "library of videotapes, which she could play in her mind and inspect at any time—'videos' of how people behaved in different circumstances. She would play these over and over again, and learn, by degrees, to correlate what she saw, so that she could then predict how people in similar circumstances might act" (259–60). This account of Grandin's "library" suggests that we do not just "learn" how to communicate with people and read their emotions (or how to read the minds of fictional characters based on their behavior)—Grandin, after all, has had as many opportunities to "learn" these things as you and me—but that we also have evolved cognitive architecture that makes this particular kind of learning possible. If this architecture is damaged, as in the case of autism, a wealth of experience would never fully make up for the damage.

Whereas the correlation between the impaired ToM and the lack of interest in fiction and storytelling is highly suggestive, the jury is still out on the exact nature of the connection between the two. It could be argued, for example, that the cognitive mechanisms that evolved to process information about human thoughts and feelings are constantly on the alert, checking out their environment for cues that fit their input conditions.⁷ On some level, then, works of fiction manage to "cheat" these mechanisms into "believing" that they are in the presence of material that they were "designed" to process, i.e., that they are in the presence of agents endowed with a potential for a rich array of intentional stances. Literature pervasively capitalizes on and stimulates ToM mechanisms that evolved to deal with real people, even as readers remain aware on some level that fictive characters are not real people at all.⁸

Thus one preliminary implication of applying what we know about ToM to our study of fiction is that ToM makes literature as we know it possible. The very process of making sense of what we read appears to be grounded in our ability to invest the flimsy verbal constructions that we generously call “characters” with a potential for a variety of thoughts, feelings, and desires, and then to look for the “cues” that allow us to guess at their feelings and thus to predict their actions.⁹ (The illusion is complete: like Erich Auerbach, we are convinced that “the people whose story the author is telling experience much more than [the author] can ever hope to tell” [549].)

II. “EFFORTLESS” MIND-READING

As we discuss mind-reading as an evolved cognitive capacity enabling both our interaction with each other and our ability to make sense of fiction, we have to be aware of the definitional differences between the terminology used by cognitive scientists and literary critics. Cognitive psychologists and philosophers of mind investigating our ToM ask such questions as: what is the evolutionary history of this adaptation, i.e., in response to what environmental challenges did it evolve? At what age and in what forms does it begin to manifest itself? What are its neurological foundations? They focus on the ways “in which mind-reading [plays] an essential part in *successful* communication” (Baron-Cohen 29 emphasis mine). When cognitive scientists turn to literary (or, as in the case below, cinematic) examples to illustrate our ability for investing fictional characters with minds of their own and reading those minds, they stress the “effortlessness” with which we do so. As Dennett observes, “watching a film with a highly original and unsteretyped plot, we see the hero smile at the villain and we all swiftly and effortlessly arrive at the same complex theoretical diagnosis: ‘Aha!’ we conclude (but perhaps not consciously), ‘He wants her to think he doesn’t know she intends to defraud her brother!’” (48).

Readers outside the cognitive science community may find this emphasis on “effortlessness” and “success” unhelpful. Literary critics, in particular, know that the process of attributing thoughts, beliefs, and desires to other people may lead to *misinterpreting* those thoughts, beliefs, and desires. Thus, they would rightly resist any notion that we could effortlessly—that is, correctly and unambiguously, nearly telepathically—figure out what the person whose behavior we are trying to explain is thinking. It is important to underscore here that cognitive scientists and lay readers (here including literary critics) bring very different frames of reference to measuring the relative “success” of mind-reading. For the lay reader, the example of a glaring failure in mind-reading and communication might be a person’s interpreting her friend’s tears of joy as tears of grief and reacting accordingly. For a cognitive psychologist, a glaring failure in mind-reading would be a person’s not even knowing that the water coursing down her friend’s face is supposed to be somehow indicative of his feelings at that moment. If you find the latter possibility absurd, recall that this is how (many) people with autism experience the world, perhaps because of neurological deficits that prevent their cognitive architecture from narrowing the range of

interpretive possibilities and restricting them, in this particular case, to the domain of emotions.

Consequently, one of the crucial insights offered by cognitive psychologists is that by thus parsing the world and narrowing the scope of relevant interpretations of a given phenomenon, our cognitive adaptations enable us to contemplate an infinitely rich array of interpretations *within* that scope. As Nancy Easterlin puts it, “without the inborn tendency to organize information in specific ways, we would not be able to experience choice in our responses” (“Making Knowledge” 137).¹⁰ “Constraints,” N. Katherine Hayles observes in a different context, “operate constructively by restricting the sphere of possibilities” (145).¹¹ In other words, our ToM allows us to connect Peter Walsh’s trembling to his emotional state (in the absence of any additional information that could account for his body language in a different way), thus usefully constraining our interpretive domain and enabling us to start considering endlessly nuanced choices *within that domain*. The context of the episode would then constrain our interpretation even further; we could decide, for instance, that it is unlikely that Peter is trembling because of a barely concealed hatred and begin to explore the complicated gamut of his bittersweet feelings. Any additional information that we would bring to bear upon our reading of the passage—biographical, sociohistorical, literary-historical—would alert us to new shades in its meaning, and could, in principle, lead us to some startling conjectures about Walsh’s state of mind. Note too, that the description of Walsh’s “trembling” may connect to something in my personal experience that will induce me to give significantly more weight to one detail of the text and to ignore others, which means that you and I may wind up with wildly different readings of Peter’s and Clarissa’s emotions “at eleven o’clock on the morning of the day she [is] giving a party.” None of this can happen, however, before we have first eliminated a whole range of other explanations, such as explanations evoking various physical forces (for instance, a disease) acting upon the body, and have focused instead solely on the mind of the character.

This elimination of irrelevant interpretations can happen so fast as to be practically imperceptible. Consider an example from Stanley Fish’s famous essay, “How to Recognize a Poem.” To demonstrate his point that our mental operations are “limited by institutions in which we are already embedded,” Fish reports the following classroom experiment:

While I was in the course of vigorously making a point, one of my students, William Newlin by name, was just as vigorously waving his hand. When I asked the other members of the class what it was that [he] was doing, they all answered that he was seeking permission to speak. I then asked them how they knew that. The immediate reply was that it was obvious; what else could he be thought of doing? The meaning of his gesture, in other words, was right there on its surface, available for reading by anyone who had the eyes to see. That meaning, however, would not have been available to someone without any knowledge of what was involved in being a student. Such a person might have thought that Mr. Newlin was pointing to the fluorescent lights hanging from the

ceiling, or calling our attention to some object that was about to fall (“the sky is falling,” “the sky is falling”). And if the someone in question were a child of elementary or middle-school age, Mr. Newlin might well have been seen as seeking permission not to speak but to go to the bathroom, an interpretation or reading that would never have occurred to a student at Johns Hopkins or any other institution of “higher learning.” (110–11)

Fish’s point that “it is only by inhabiting . . . the institutions [that] precede us [here, the college setting] that we have access to the public and conventional senses they make [here, the raised hand means the person seeks permission to speak]” (110) is well taken. Yet note that all of his patently “wrong” explanations (e.g., Mr. Newlin thought that the sky was falling; he wanted to go to the bathroom, etc.) are “correct” in the sense that they call on a ToM; that is, they explain the student’s behavior in terms of his underlying thoughts, beliefs, and desires. As Fish puts it, “what else could he be *thought* of doing?” (emphasis mine). Nobody ventured to suggest, for example, that there was a thin, practically invisible string threaded through the loop in the classroom’s ceiling, one end of which was attached to Mr. Newlin’s sleeve and another held by a person sitting behind him who could pull the string any time and produce the corresponding movement of Mr. Newlin’s hand. Absurd, we should say, especially since nobody could observe any string hovering over Mr. Newlin’s head. Is it not equally absurd, however, to explain a behavior in terms of a mental state that is completely unobservable? Yet we do it automatically, and the only reason that no “normal” (i.e., non-autistic) person would think of a “mechanistic” explanation (such as the string pulling on the sleeve) is that we have cognitive adaptations that prompt us to “see bodies as animated by minds.”

But then, by the very logic of Fish’s essay, which urges us not to take for granted our complex *institutional* embedment that allows us to make sense of the world, shouldn’t we inquire with equal vigor into our *cognitive* embedment that—as I hope I have demonstrated in the example above—profoundly informs the institutional one? Given the suggestively constrained range of the “wrong” interpretations offered by Fish (that is, all his interpretations connect the behavior to a mental state), shouldn’t we qualify his assertion that unless we read Mr. Newlin’s raised hand in the context of his being a student, “there is nothing *in the form* of [his] gesture that tells his fellow students how to determine its significance” (112)? Surely the *form* of the gesture—staying with the word that Fish himself has emphasized—is quite informative because its very deliberateness seems to delimit the range of possible “wrong” interpretations. That is, had Mr. Newlin unexpectedly jerked his hand instead of “waving” it “vigorously,” some mechanical explanation such as a physiological spasm or someone pushing his elbow, perhaps even a wire attached to his sleeve, would seem far less absurd.

To return, then, to the potentially problematic issue of the effortlessness with which we “read” minds: a flagrantly “wrong,” from lay readers’ perspective, interpretation, such as taking tears of grief for tears of joy or thinking that Mr. Newlin raises his hand to point out that the sky is falling, is still “effortless” from the point of view of cognitive psychologists because of the ease with which we correlate tears

with an emotional state or the raised hand with a certain underlying desire/intention. Mind-reading is thus effortless in the sense that we “intuitively” connect people’s behavior to their mental states—as in the example involving Walsh’s “trembling”—although our subsequent description of their mental states could run a broad gamut of mistaken or disputed meanings. For any description is, as Fish reminds us on a different occasion, “always and already interpretation,” a “text,” a story reflecting the personal history, biases, and desires of the reader.¹²

III. CAN COGNITIVE SCIENCE TELL US WHY WE ARE AFRAID OF *MRS. DALLOWAY*?

How much prompting do we need to begin to attribute a mind of her own to a fictional character? Very little, it seems, since any indication that we are dealing with a self-propelled entity (e.g., “Peter Walsh has come back”) leads us to assume that this entity possesses thoughts, feelings, and desires, at least some of which we could intuit, interpret, and, frequently, misinterpret. Writers exploit our constant readiness to posit a mind whenever we observe behavior when they experiment with the amount and kind of interpretation of the characters’ mental states that they supply themselves and that they expect their readers to supply. When Woolf shows Clarissa observing Peter’s body language (Clarissa notices that he is “positively trembling”), she has an option of providing us with a representation of either Clarissa’s mind that would make sense of Peter’s physical action (something to the effect of “how excited must he be to see her again!”) or of Peter’s own mind (as in “so excited was he to see his Clarissa again!”). Instead she tells us, first, that Peter is thinking that Clarissa has grown older and, second, that Clarissa is thinking that Peter looks “exactly the same; . . . the same queer look; the same check suit” (40). Peter’s “trembling” still feels like an integral part of this scene, but make no mistake: we, the readers, are called on to supply the missing bit of information (such as “he must be excited to see her again”) that makes the narrative emotionally cohesive.

Hemingway famously made it his trademark to underrepresent his protagonists’ feelings by forcing the majority of his characters’ physical actions to stand in for mental states (for example, as in the ending of *A Farewell to Arms*: “After a while I went out and left the hospital and walked back to the hotel in the rain” [314]). Hemingway could afford such a deliberate, and in its own way highly elaborate, under-telling for the same reason that Woolf could afford to let Peter’s trembling “speak for itself”: our evolved cognitive tendency to assume that there *must be* a mental stance behind each physical action and our striving to represent to ourselves that possible mental stance even when the author has left us with the absolute minimum of necessary cues for constructing such a representation.

It is thus when we start to inquire into how writers of fiction *experiment* with our mind-reading ability, and perhaps even push it further, that the insights offered by cognitive scientists become particularly pertinent. Although cognitive scientists’ investigation of ToM is very much a project-in-progress, literary scholars have enough carefully documented research already available to them to begin asking

such questions as: is it possible that literary narrative trains our capacity for mind-reading and also tests its limits? How do different cultural-historical milieus encourage different literary explorations of this capacity? How do different genres? Speculative and tentative as the answers to these questions could only be at this point, they mark the possibility of a genuine interaction between cognitive psychology and literary studies, with both fields having much to offer to each other.

This section's tongue-in-cheek title refers to my attempt to apply a series of recent experiments conducted by cognitive psychologists studying ToM to *Mrs. Dalloway*. I find the results of such an application both exciting and unnerving. On the one hand, I can argue now with a reasonable degree of confidence that certain aspects of Woolf's prose do place extraordinarily high demands on our mind-reading ability and that this could account, *at least in part*, for the fact that many readers feel challenged by that novel. On the other hand, I have come to be "afraid" of *Mrs. Dalloway*—and, indeed, other novels—in a different fashion, realizing that any initial inquiry into the ways fiction teases our ToM immediately raises more questions about ToM and fiction than we are currently able to answer. My ambivalence, in other words, stems from the realization that ToM underlies our interaction with literary texts in such profound and complex ways that any endeavor to isolate one particular aspect of such an interaction feels like carving the text at joints that are fundamentally, paradigmatically absent.

This proviso should be kept in mind as we turn to the experiments investigating one particular aspect of ToM, namely, our ability to navigate multiple levels of intentionality present in a narrative. Although ToM is formally defined as a second-order intentionality, as in the statements "I believe that you desire X" or "Peter Walsh thinks that Clarissa 'would think [him] a failure'" (43), the levels of intentionality can "recurse" further back, for example, to the fourth level, as in a statement like "I believe that you think that she believes that he thinks that X." Dennett, who first discussed this recursiveness of the levels of intentionality in 1983, thought it could be, in principle, infinite. A recent series of striking experiments reported by Robin Dunbar and his colleagues have suggested, however, that our cognitive architecture may discourage the proliferation of cultural narratives that involve "infinite" levels of intentionality.

In those experiments, subjects were given two types of stories—one that involved a "simple account of a sequence of events in which 'A gave rise to B, which resulted in C, which in turn caused D, etc.'" and another that introduced "short vignettes on everyday experiences (someone wanting to date another person, someone wanting to persuade her boss to award a pay rise), . . . [all of which] contained between three and five levels of embedded intentionality." Subjects were then asked to complete a "series of questions graded by the levels of intentionality present in the story," including some factual questions "designed to check that any failures of intentionality questions were not simply due to failure to remember the material facts of the story." The results of the study were revealing: "Subjects had little problem with the factual causal reasoning story: error rates were approximately 5% across six levels of causal sequencing. Error rates on the mind-reading tasks were similar (5–10%) up to and including fourth-level intentionality, but rose dramatically to

nearly 60% on fifth-order tasks.” Cognitive scientists knew that this “failure on the mind-reading tasks [was] not simply a consequence of forgetting what happened, because subjects performed well on the memory-for-facts tasks embedded into the mind-reading questions” (Dunbar 241). The results thus suggest that people have marked difficulties processing stories that involve mind-reading above the fourth level.

An important point that should not be lost in the discussion of these experiments is that it is the *content* of the information in question that makes the navigation of multiply-embedded data either relatively easy or difficult. Cognitive evolutionary psychologists suggest the following reason for the relative ease with which we can process long sequences such as “A gave rise to B, which resulted in C, which in turn caused D, which led to E, which made possible F, which eventually brought about G, etc.,” as opposed to similarly long sequences that require attribution of states of mind, such as “A wants B to believe that C thinks that D wanted E to consider F’s feelings about G.” It is likely that cognitive adaptations that underwrite the attribution of states of mind differ in functionally important ways from the adaptations that underwrite reasoning that does not involve such an attribution, a difference possibly predicated on the respective evolutionary histories of both types of adaptations.¹³ A representation of a mind as represented by a mind as represented by yet another mind will thus be supported by cognitive processes distinct (to a degree which remains a subject of debate) from cognitive processes supporting a mental representation, for example, of events related to each other as a series of causes and effects or of a representation of a Russian doll nested within another doll nested within another doll. The cognitive process of representing depends crucially on *what* is being represented.

Consider now a randomly selected passage roughly halfway into Woolf’s *Mrs. Dalloway*, in which Richard Dalloway and Hugh Whitbread come to Lady Bruton to write a letter to the *Times*, and in which, to understand what is going on, we have to confront a series of multiply embedded states of mind:

And Miss Brush went out, came back; laid papers on the table; and Hugh produced his fountain pen; his silver fountain pen, which had done twenty years’ service, he said, unscrewing the cap. It was still in perfect order; he had shown it to the makers; there was no reason, they said, why it should ever wear out; which was somehow to Hugh’s credit, and to the credit of the sentiments which his pen expressed (so Richard Dalloway felt) as Hugh began carefully writing capital letters with rings round them in the margin, and thus marvelously reduced Lady Bruton’s tangles to sense, to grammar such as the editor of the *Times*, Lady Bruton felt, watching the marvelous transformation, must respect. (110)

What is going on in this passage? We are seemingly invited to deduce the excellence of Millicent Bruton’s civic ideas—put on paper by Hugh—first from the resilience of the pen that he uses, and then from the beauty of his “capital letters with rings around them on the margins.” Of course, this reduction of lofty sentiments and superior analytic skills to mere artifacts, such as writing utensils and calligraphy, achieves just the opposite effect. By the end of the paragraph, we are ready to accept

Richard Dalloway's view of the resulting epistle as "all stuffing and bunkum," but a harmless bunkum at that. Its inoffensiveness and futility are underscored by the tongue-in-cheek phallic description of the silver pen (should "silver" bring to our mind "gray"?) that has served Hugh for twenty years but that is still "in perfect order"—or so Hugh thinks—once he's done "unscrewing the cap."

There are several ways to map this passage out in terms of the nested levels of intentionality. I will start by listing the smallest irreducible units of embedded intentionality and gradually move up to those that capture as much of the whole narrative gestalt of the described scene as possible:

1. The makers of the pen *think* that it will never wear out. (First level)
2. Hugh *says* that the makers of the pen *think* it will never wear out. (Second level)
3. Lady Bruton *wants* the editor of the *Times* to *respect* and publish her ideas. (Second level)
4. Hugh *wants* Lady Bruton and Richard to *believe* that because the makers of the pen *think* that it will never wear out, the editor of the *Times* will *respect* and publish the ideas recorded by this pen. (Fourth level)
5. Richard *is aware* that Hugh *wants* Lady Bruton and Richard Dalloway to *believe* that because the makers of the pen *think* that it will never wear out, the editor of the *Times* will *respect* and publish the ideas recorded by this pen. (Fifth level)
6. Richard *suspects* that Lady Bruton indeed *believes* that because, as Hugh *says*, the makers of the pen *think* that it will never wear out, the editor of the *Times* will *respect* and publish the ideas recorded by this pen. (Fifth level)
7. By inserting a parenthetical observation ("so Richard Dalloway felt"), Woolf *intends us to recognize* that Richard *is aware* that Hugh *wants* Lady Bruton and Richard to *think* that because the makers of the pen *believe* that it will never wear out, the editor of the *Times* will *respect* and publish the ideas recorded by this pen. (Sixth level)

It could be argued, of course, that in the process of reading we automatically cut through Woolf's stylistic pyrotechnics to come up with a series of more comprehensible, first-, second-, and third-level attributions of states of mind, such as "Richard does not particularly like Hugh"; "Lady Bruton thinks that Hugh is writing a marvelous letter"; "Richard feels that Lady Bruton thinks that Hugh is writing a marvelous letter, but he is skeptical about the whole enterprise"; and so on. Such abbreviated attributions may seem destructive since the effect that they have on Woolf's prose is equivalent to the effect of paraphrasing on poetry, but they do, in fact, convey some general sense of what is going on in the paragraph. The main problem with them, however, is that to arrive at such simplified descriptions of Richard's and Lady Bruton's states of mind, we have to grasp the full meaning of this passage, and to do that, we first have to process several sequences that embed at least five levels of intentionality. Moreover, we have to do it on the spot, unaided by pen and paper and not forewarned that the number of levels of intentionality that we

are about to encounter is considered by cognitive scientists to create “a very significant load on most people’s cognitive abilities” (Dunbar 240).

Note that in this particular passage, Woolf not only “demands” that we process a string of fifth- and sixth-level intentionalities but she also introduces such embedded intentionalities through descriptions of body language that in some ways approach those of Hemingway in their emotional blandness. No more telling “trembling,” as in the earlier scene featuring Peter and Clarissa. Instead, we get Richard watching Lady Bruton watching Hugh producing his pen, unscrewing the cap, and beginning to write. True, Woolf offers us two emotionally colored words (“carefully” and “marvelously”), but what they signal is that Hugh cares a great deal about his writing and that Lady Bruton admires the letter that he produces—two snapshots of the states of mind that only skim the surface of the complex affective undertow of this episode.

Because Woolf has depicted physical actions relatively lacking in immediate emotional content, here, in striking contrast to the scene in Clarissa’s drawing-room, she hastens to provide an authoritative interpretation of each character’s mental state. We are told what Lady Bruton feels as she watches Hugh (she feels that the editor of the *Times* will respect so beautifully written a letter); we are told what Hugh thinks as he unscrews the cap (he thinks that the pen will never wear out and that its longevity contributes to the worth of the sentiments it produces); we are told what Richard feels as he watches Hugh, his capital letters, and Lady Bruton (he is amused both by Hugh’s exalted view of himself and by Lady Bruton’s readiness to take Hugh’s self-importance at its face value). The apparently unswerving linear hierarchy of the scene—Richard can represent the minds of both Hugh and Lady Bruton, but Hugh and Lady Bruton cannot represent Richard’s representations of their minds—seems to enforce the impression that each mind is represented fully and correctly.

Of course, Woolf is able to imply that her representations of Hugh’s, Lady Bruton’s, and Richard’s minds are exhaustive and correct because, creatures with a ToM that we are, we *just know* that there *must be* mental states behind the emotionally opaque body language of the protagonists. The paucity of textual cues that could allow us to imagine those mental states ourselves leaves us no choice but to accept the representations provided by the author. We have to work hard for them, of course, for sifting through all those levels of embedded intentionality tends to push the boundaries of our mind-reading ability to its furthest limits.

When we try to articulate our perception of the cognitive challenge induced by this task of processing fifth- and sixth-level intentionality, we may say that Woolf’s writing is difficult or even refuse to continue reading her novels. The personal aesthetics of individual readers thus could be grounded *at least in part* in the nuances of their individual mind-reading capacities. By saying this I do not mean to imply that if somebody “loves” or “hates” Woolf, it should tell us something about that person’s general mind-reading “sophistication”—a cognitive literary analysis does not support such misguided value judgments. The nuances of each person’s mind-reading profile are unique to that person, just as, for example, we all have the capacity for developing memories (unless that capacity has been clinically impaired), but each indi-

vidual's actual memories are unique. My combination of memories serves me, and it would be meaningless to claim that it somehow serves me "better" than my friend's combination of memories serves her. At the same time, I see no particular value in celebrating the person's dislike of Woolf as the manifestation of his or her individual cognitive make-up. My teaching experience has shown that if we alert our students to the fact that Woolf tends to play this particular kind of cognitive "mind game" with her readers, it significantly eases their anxiety about "not getting" her prose and actually helps them to start enjoying her style.¹⁴

IV. COGNITIVE LITERARY ANALYSIS OF *MRS. DALLOWAY*

It is now time to return to the imaginary conversation that opened my essay. Some versions of that exchange did take place at several scholarly forums where I have presented my research on ToM and literature. Once, for instance, after I described the immediate pedagogical payoffs of counting the levels of intentionality in *Mrs. Dalloway* with my undergraduates, I was asked if I could foresee the time when such a cognitive reading would supersede and render redundant the majority of other, more traditional approaches to Woolf.¹⁵ My immediate answer was, and still remains, an unqualified no, but since then I have had the opportunity to consider several of that question's implications that are important for those of us wishing cognitive approaches to literature to thrive.

First of all, counting the levels of intentionality in *Mrs. Dalloway* does not constitute *the* cognitive approach to Woolf. It merely begins to explore one particular way—among numerous others—in which Woolf builds on and experiments with our ToM, and—to cast the net broader—in which fiction builds on and experiments with our cognitive propensities.¹⁶ Many of these propensities, I feel safe saying in spite of remarkable advances in the cognitive sciences during the last two decades, still remain unknown to us.

However, the current state of cognitive approaches to literature already testifies to the spectacular diversity of venues offered by the parent fields of cognitive neuroscience, artificial intelligence, philosophy of mind, cognitive linguistics, evolutionary biology, cognitive psychology, and cognitive anthropology. Literary critics have begun to investigate the ways in which recent research in these areas opens new avenues in gender studies (F. Elizabeth Hart); feminism (Elizabeth Grosz); cultural materialism (Mary Thomas Crane, Alan Richardson); deconstruction (Ellen Spolsky); literary aesthetics (Elaine Scarry, Gabrielle Starr); history of moral philosophy (Blakey Vermeule); ecocriticism (Nancy Easterlin); and narrative theory (Porter Abbott, David Herman, Paul Hernadi). What these scholars' publications show is that far from displacing the traditional approaches or rendering them redundant, a cognitive approach ensures their viability as it builds on, strengthens, and develops their insights.

Second, the ongoing dialogue with, for instance, cultural historicism or feminism is not simply a matter of choice for scholars of literature interested in cognitive approaches. There is no such thing as a cognitive ability, such as ToM, free-floating

“out there” in isolation from its human embodiment and its historically and culturally concrete expression. Evolved cognitive predispositions, to borrow Patrick Colm Hogan’s characterization of literary universals, “are instantiated variously, particularized in specific circumstances” (226).¹⁷ *Everything* that we learn about Woolf’s life and about the literary, cultural, and sociohistorical contexts of *Mrs. Dalloway* is thus potentially crucial for understanding why this particular woman, at this particular historical juncture, seeing herself as working both within and against a particular set of literary traditions, began to push beyond the boundaries of her readers’ cognitive “zone of comfort” (that is, beyond the fourth level of intentionality).

At the same time, to paraphrase David Herman (“Regrounding”), the particular combination of these personal, literary, and historical contexts, in all their untold complexity, is a “necessary though not a sufficient condition” for understanding why Woolf wrote the way she did. No matter how much we learn about the writer herself and her multiple environments, and no matter how much we find out about the cognitive endowments of our species that, “particularized in specific circumstances,” make fictional narratives possible, we can only go so far in our cause-and-effect analysis. As George Butte puts it, “accounts of material circumstances can describe changes in gender systems and economic privileges, but they cannot explain why *this* bankrupt merchant wrote *Moll Flanders*, or why *this* genteely-impooverished clergyman’s daughter wrote *Jane Eyre*.” There will always remain a gap between our ever-increasing store of knowledge and the phenomenon of Woolf’s prose—or, for that matter, Defoe’s, Austen’s, Brontë’s, and Hemingway’s prose.

Yet to consider just one example of how crucial our “other” knowledges are for our cognitive inquiry into *Mrs. Dalloway*, let us situate Woolf’s experimentation with multiple levels of intentionality within the history of the evolution of the means of textual reproduction. It appears that a written culture is, on the whole, more able than an oral culture to support elaborately nested intentionality simply because a paragraph with six levels of intentional embedment does not yield itself easily to memorization and subsequent oral transmission. It is thus highly unlikely that we would find many (or any) passages that require us to go beyond the fourth level of intentionality in oral epics such as *Gilgamesh* or *The Iliad*. Walter Benjamin captures the broad point of this difference when he observes that the “listener’s naïve relationship to the storyteller is controlled by his interest in retaining what he is told. The cardinal point for the unaffected listener is to assure himself of the possibility of reproducing the story” (97). The availability of the means of written transmission, such as print, enables the writer “to carry the incommensurable to extremes in representations of human life,”¹⁸ and by so doing, to explore (or shall we actually say “develop,” thus drawing upon Paul Hernadi’s recent argument about the evolutionary origins of literature?)¹⁹ the hitherto quiescent cognitive spaces.

Of course, for a variety of aesthetic, personal, and financial reasons, not every author writing under the conditions of print will venture into such cognitive unknown. Even a cursory look through the best-selling mainstream fiction, from Belva Plain to Danielle Steel, confirms the continuous broad popular appeal of narratives dwelling under the fourth level of intentional embedment. It is, then, the personal histories of individuals (here, individual writers and their audiences) that insure that,

as Alan Richardson and Francis Steen observe, the history of cognitive structures “is neither identical to nor separate from the culture they make possible” (3).

In the case of Woolf, scholars agree that severing ties with the Duckworth—the press that had brought forth her first two novels and was geared toward an audience that was “Victorian, conventional, anti-experimentation” (*Diary* 1:261)—“liberated [her] experimentalism” (Whitworth 150). Having her own publishing house, the Hogarth Press, meant that she was “able to do what” she “like[d]—no editors, or publishers, and only people to read who more or less like that sort of thing” (*Letters* 167). Another factor possibly informing the cognitive extremes of *Mrs. Dalloway* was Woolf’s acute awareness of the passing of time: “my theory is that at 40 one either increases the pace or slows down” (*Diary* 2:259). Woolf wanted to *increase* the pace of her explorations, to be able to “embody, at last,” as she would write several years later, “the exact shapes my brain holds” (*Diary* 4:53). Having struggled in her previous novels with the narrator “choked with observations” (*Jacob’s Room* 67), she discovered in the process of working on *Mrs. Dalloway* how to “dig out beautiful caves behind [her] characters; . . . The idea is that the caves shall connect, and each comes to daylight at the present moment” (*Diary* 2:263). Embodying the “exact shapes” of Woolf’s brain thus meant, among other things, shifting “the focus from the mind of the narrator to the minds of the characters” and “from the external world to the minds of the characters perceiving it” (Dick 51, 52), a technique that would eventually prompt Auerbach to inquire in exasperation, “Who is speaking in this paragraph?” (531).²⁰

Woolf’s meditations on her writing remind us of yet another reason that simply counting levels of intentionality in *Mrs. Dalloway* will never supersede other forms of critical inquiry into the novel. When Woolf explains that she wants to construct a “present moment” as a delicate “connection” among the “caves” dug behind each character, the emerging image overlaps suggestively with Dennett’s image of the infinitely recursive levels of intentionality. (“Aha,” concludes the delighted cognitive literary critic, “Woolf had some sort of proto-theory of recursive mind-reading!”) But with her vivid description of the catacomb-like subjectivity of the shared present moment,²¹ Woolf also manages to do something else—and that “something else” proceeds to quietly burrow into our (and her) cognitive theorizing.

This brings us to a seemingly counterintuitive but important point underlying cognitive literary analysis. Even as I map the passage featuring Richard Dalloway and Hugh Whitbread at Lady Bruton’s as a linear series of embedded intentionalities, I expect that something else present in that passage will complicate that linearity and re-pose Auerbach’s question, albeit with a difference. Will it be the phallic overtones of the description of Hugh’s pen? Or the intrusion of rhetoric of economic exchange—“credit,” “makers,” “produce,” “capital,” “margin”? Or the vexed gender contexts of the “ventriloquism” implied by the image of Millicent Bruton spouting political platitudes in Hugh’s voice?²² Or the equally vexed social class contexts of the “seating arrangements” that hierarchize the mind-reading that goes on in the passage? (After all, Woolf must have “seated” Lady Bruton’s secretary, Miss Brush, too far from the desk to be able to see the shape of Hugh’s letters so as not to add yet another level of mental embedment by having Miss Brush watch Richard watching Lady Bruton watching Hugh.) Cognitive *literary* analysis thus continues beyond the

line drawn by cognitive scientists—with the reintroduction of something else, a “noise,” if you will, that is usually carefully controlled for and excised, whenever possible, from the laboratory settings.

V. WOOLF, PINKER, AND THE PROJECT OF INTERDISCIPLINARITY

Woolf’s prose, fundamentally rooted in and tirelessly stimulating our cognitive capacities, represents such a tantalizing subject for a cognitive literary analysis that one is startled to learn that a cognitive scientist has recently characterized Woolf as having inaugurated an aesthetic movement whose “philosophy did not acknowledge the ways in which it was appealing to human pleasure” (Pinker 413). Although Steven Pinker admits that “modernism comprises many styles and artists, . . . not [all of which] rejected beauty and other human sensibilities” and that modernist “fiction and poetry offered invigorating intellectual workouts” (404), here is what he has to say about modernism as a whole and Woolf in particular:

The giveaway [explanation for the current crisis in the arts and humanities] may be found in a famous statement from Virginia Woolf: “[On] or about December 1910, human [character] changed.” She was referring to the new philosophy of modernism that would dominate the elite arts and criticism for much of the twentieth century, and whose denial of human nature was carried over with a vengeance to postmodernism, which seized control in its later decades. . . . Modernism certainly proceeded *as if* human nature had changed. All the tricks that artists had used for millennia to please the human palate were cast aside. . . . In literature, omniscient narration, structured plots, the orderly introduction of characters, and general readability were replaced by a stream of consciousness, events presented out of order, baffling characters and causal sequences, subjective and disjointed narration, and difficult prose. (409–10)²³

As literary critics, we have several ways of responding to Pinker’s claims about Woolf. We can hope that not “many students, teachers, theorists, and critics of literature will take [him] seriously as an authority on literature or the aesthetics more generally, especially since he misrepresents both Woolf and modernism.”²⁴ At first sight, this is a comfortable stance. It assumes a certain cultural detachment of literary studies and implies that cognitive scientists should just leave literature alone, acknowledging it as an exclusive playing field for properly trained professionals—us. The problem with this view is that it disregards two facts: first, that more people read Pinker (who “misrepresents” Woolf) rather than, say, *PMLA* (which could set the matter straight), and, second, that as a very special, richly concentrated cognitive artifact, literature already is fair game for scientists, including Pinker, Daniel Dennett, Paul Harris, Robin Dunbar, and others, and it will become even more so as cognitive inquiry spreads further across cultural domains.

I suggest that instead of simply ignoring Pinker’s assertion that the modernist writers’ generally “difficult” prose cannot, by and large, “please the human palate,”

we should engage his argument, incorporating both the insights from our own field and those offered by cognitive scientists. By taking seriously the idea that our cognitive evolutionary heritage structures the ways in which we make sense of fictional narrative, we can gain a better understanding of why and how different “human palates” in different historical milieus can be “pleased” by quite different literary fare. Furthermore, we can show that it is by paying attention to the elite, to the exceptional, to the cognitively challenging, such as Woolf’s experimentation with the levels of intentional embedment, that we can develop, for instance, a more sophisticated perspective on the workings of our ToM. As James Phelan observes, would not Pinker himself and “those in his audience who view modernist literature as he does be more likely to be persuaded to change their dismissive view of it, if literary critics show that [Woolf’s] representations of consciousness, though initially challenging to a reader, are highly intelligible because they capture in their own ways insights that Pinker and other cognitive scientists have been offering (and popularizing)?”²⁵ And what exactly are the epistemological and ethical grounds on which we stand when we mock Pinker’s claim to being an “authority on literature” if we have not yet made this kind of good-faith effort to meet Pinker halfway and offer our literary-historical expertise to develop a more sophisticated cognitive perspective on modernist representations of fictional consciousness?

Consider again the above-discussed insights of Robin Dunbar and his colleagues. As I hope to have demonstrated in this essay, Dunbar’s research into our processing of stories that involve mind-reading above the fourth level can have far-reaching consequences for literary analysis. Yet there is no reason why, based on our knowledge of literary history, we should not ask him to qualify some of his arguments (and, indeed, would not Dunbar himself appreciate precisely this kind of response?), even if at this point, given how new the whole field is, we may have to settle for less-than-definitive answers to our criticism.

For example, Dunbar offers a fascinating speculation about the significance of his findings for our understanding of why there are generally more good readers than good writers:

The fact that people seem to experience considerable difficulty with fifth-order intentional statements, but not fourth-order ones, may explain why writing fiction is much harder than reading it, and may thus in part explain why good writers are [much] less common than good readers. . . . A novelist writing about relationship between three people has to ‘*intend* that the reader *think* that character A *supposes* that character B *wants* character C to *believe* that . . .’—five orders of intentionality. The reader, in contrast, has a much easier task: he or she merely has to ‘*think* that A *supposes* that B *wants* C to *believe* that . . .’—four orders of intentionality. (241)

Dunbar’s argument has interesting implications for our theorizing the figure of the unreliable narrator as well as the relationship between the author and the narrator. For instance, our frequently ambivalent reaction to a suddenly perceived split of the narratorial presence—we may react to it by feeling excited, intrigued, and yet unsettled—could be related, among other things, to our semiconscious realization that

we must factor in yet another level of intentionality, thus adding to the cognitive challenge already presented by the text. At the same time, as Phelan notes, Dunbar's speculation that the difficulty that we have with processing fifth-order intentional statements may provide insight into why good writers are less common than good readers is "unpersuasive" because it "would predict that until we get to fictions with five or more levels of intentionality"—which happened relatively recently in our literary history and was predicated on, among other things, the evolution of the means of textual reproduction—"the number of good writers and good readers should be approximately the same." Since the latter is clearly not the case, and since the marked paucity of literary texts going beyond the fourth level of embedded intentionality, say, in the Middle Ages, would not lead us to assume that the number of good writers and good readers in that period was approximately the same, Dunbar may want to consider how this historical dimension complicates his provocative argument.

These examples support my claim that there is now the possibility of a genuine interaction between cognitive science and literary studies, one that does not just pay obligatory lip service to interdisciplinarity while quietly assuming the superiority of science. Paradoxically, it is only while we refuse to "take seriously" the research of cognitive scientists who dare to pronounce "on literature or . . . aesthetics more generally," that we can be made to feel that our contribution to this interdisciplinary exchange would represent little or nothing of value. Once we enter the conversation and engage with respect the arguments of Dunbar, Pinker, Dennett, and others, we realize that because of their ever-increasing—and well-warranted—interest in how the human mind processes literary narratives, our expertise could make a crucial difference for the future shape of the whole field of cognitive science.

ENDNOTES

I am grateful to James Phelan for his thoughtful suggestions and comments, many of which I have eagerly seized upon and quoted verbatim in my essay. Parts of the present argument, particularly those dealing with ToM and autism and "effortless" mind-reading, overlap with the argument I am making in my essay "Richardson's *Clarissa* and a Theory of Mind," forthcoming in Ellen Spolsky and Alan Richardson's collection *The Work of Fiction: Cognition, Culture, and Complexity*, and I am grateful to Alan and Ellen for their patient and generous engagement with my "theory of mind and fiction" argument.

1. Like Hermione Lee, we could ground it in Woolf's position as a "pioneer of reader-response theory." Woolf, she writes, "was extremely interested in the two-way dialogue between readers and writers. Books change their readers; they teach you how to read them. But readers also change books. 'Undoubtedly,' Woolf herself had written, 'all writers are immensely influenced by the people who read them'" ("Virginia Woolf's Essays" 91).
2. For a related analysis of "representations of representations" or "metarepresentations," see Zunshine, "Eighteenth-Century Print Culture."
3. An important tenet of a cognitive approach to literature is that, as Paul Hernadi puts it, "there is no clear division between literary and nonliterary signification. . . . Literary experience is not triggered in a cognitive or emotive vacuum: modern readers, listeners, and spectators mentally process the virtual comings and goings of imagined characters as if they were analogous to remembered actual events" (60, 62). For a related discussion, see Mark Turner, *The Literary Mind*.

4. On the social intelligence of nonhuman primates, see Byrne and Whiten, *Machiavellian Intelligence* and "The Emergence of Metarepresentation"; Gomez, "Visual Behavior"; Premack and Dasser, "Perceptual Origins."
5. For a discussion of alternatives to the Theory of Mind approach, see Dennett, *The Intentional Stance*.
6. Leo Kanner first described autism in 1943. For more than twenty years after that, autism was "mistakenly thought to be caused by a cold family environment." In 1977, however, "a landmark twin study showed that the incidence of autism is strongly influenced by genetic factors," and, since then, "numerous other investigations have since confirmed that autism is a highly heritable disorder" (Hughes and Plomin 48). For the "pre-history" of the term autism, particularly as introduced by Eugen Bleuler in 1911 and developed by Piaget in 1923, see Harris 3.
7. By using the word "mechanism," I am not trying to smuggle the outdated "body as a machine" metaphor into literary studies. Tainted as this word is by its previous history, it can still function as a convenient shorthand designation for extremely complex cognitive processes.
8. For a discussion, see Leslie 120–25; Carruthers, "Autism as Mind-Blindness" 262–63; Hernadi 58; and Spolsky, "Why and How."
9. The scale of such investment emerges as truly staggering if we attempt to spell out the host of unspoken assumptions that make it possible (for a discussion, see Zunshine, "Richardson's *Clarissa*"). This realization lends new support to what theorists of narrative view as the essential underdetermination or "undertelling" of fiction, its "interior nonrepresentation" (Sternberg 119).
10. For a qualification of the term "inborn" in relation to the processing of incoming data, see Spolsky, *Satisfying Skepticism* 164.
11. For an important recent discussion of "constraints," see Spolsky, "Cognitive Literary Historicism."
12. For a discussion, see Fish, *Is There a Text in this Class?*
13. For a discussion, see Carey and Spelke and Cosmides and Tooby on domain specificity. For a recent application of the theory of domain specificity to the study of literature, see Zunshine, "Rhetoric, Cognition, and Ideology."
14. Thus bringing the findings of cognitive scientists to bear upon the literary text does not diminish its aesthetic value. As Scarry has argued in response to the fear that science would "unweave the rainbow" of artistic creation, "the fact of the matter is that when we actually look at the nature of artistic creation and composition, understanding it does not mean doing it less well. To become a dancer, for example, one must do the small steps again and again and understand them, if one is to achieve virtuosity. Right now we need virtuosity, not only within each discipline, but across the disciplines as well" ("Panel Discussion" 253).
15. For a discussion, see Easterlin, "Voyages in the Verbal Universe."
16. As a friend working with cognitive/evolutionary approaches to fiction observed recently, "literature-fiction-writing is so powerful because it eats theories for breakfast, including cognitive/evolutionary approaches" (Blakey Verneule, personal communication, 20 November 2002).
17. For a discussion of embodied cognition, see also Hart.
18. For a related discussion, see Hogan 242–43.
19. Hernadi argues that "literature, whether encountered in live performance or in textual and electronic recording, can challenge and thus enhance our brains' vital capacities for expression, communication, representation, and signification." He further connects the fictional text's capacity for developing our minds to the evolutionary history of the literary endeavor. He points out that, "the protoliterary experiences of some early humans could, other things being equal, enable them to outdo their less imaginative rivals in the biological competition for becoming the ancestors of later men and women" (56).
20. Strictly speaking, Auerbach's question refers to *To the Lighthouse*, but it is equally pertinent for our discussion of *Mrs. Dalloway*.

21. A remarkable new study by George Butte, *I Know That You Know That I Know: Narrating Subjects from Moll Flanders to Marnie*, offers a fascinating perspective on a writer's interest in constructing a "present moment" as a delicate "connection" among the characters' subjectivities. Applying Maurice Merleau-Ponty's analysis of interlocking consciousnesses (*Phenomenology of Perception*) to a broad selection of eighteenth- and nineteenth-century novels, as well as to the films of Hitchcock, Hawks, and Woody Allen, Butte argues compellingly that something had changed in the narrative representation of consciousness at the time of Jane Austen: writers became able to represent the "deep intersubjectivity" of their characters, portraying them as aware of each other's perceptions of themselves and as responding to such perceptions with body language observable by their interlocutors, which generated a further series of mutual perceptions and reactions. Although Butte does not refer in his work to cognitive science or the Theory of Mind, his argument is in many respects compatible with the literary criticism that does.
22. On Woolf's definition of narrative ventriloquism, see DiBattista 132.
23. Pinker actually misquotes Woolf in his book to make his point stronger. According to Pinker, Woolf wrote that "In or about December 1910, human nature changed."
24. I quote here an anonymous reader for *PMLA*.
25. The quotations of Phelan are from a personal communication from 17 April 2003.

WORKS CITED

- Abbott, Porter.** "Humanists, Scientists and Cultural Surplus." *Substance* 94/95: 30 (2001): 203–17.
- Auerbach, Erich.** *Mimesis*. Princeton: Princeton Univ. Press, 1991.
- Baron-Cohen, Simon.** *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge: MIT Press, 1995.
- Benjamin, Walter.** *Illuminations*. New York: Harcourt, Brace & World, 1955.
- Brook, Andrew, and Don Ross.** *Daniel Dennett*. Cambridge: Cambridge Univ. Press, 2002.
- Butte, George.** *I Know That You Know That I Know: Narrating Subjects from Moll Flanders to Marnie*. Columbus: The Ohio State University Press, forthcoming.
- Byrne, Richard W., and Andrew Whiten.** "The Emergence of Metarepresentation in Human Ontogeny and Primate Phylogeny." In *Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading*, edited by Andrew Whiten, 267–82. Oxford: Basil Blackwell, 1991.
- . *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford: Oxford Univ. Press, 1988.
- Carey, Susan, and Elizabeth Spelke.** "Domain-Specific Knowledge and Conceptual Change." In *Mapping the Mind: Domain Specificity in Cognition and Culture*, edited by Lawrence A. Hirschfeld and Susan A. Gelman, 169–200. New York: Cambridge Univ. Press, 1994.
- Carruthers, Peter.** "Autism as Mind-Blindness: An Elaboration and Partial Defense." In *Theories of Theories of Mind*, edited by Peter Carruthers and Peter K. Smith, 257–73. Cambridge: Cambridge Univ. Press, 1996.
- Cosmides, Leda, and John Tooby.** "Origins of Domain Specificity: The Evolution of Functional Organization." In *Mapping the Mind: Domain Specificity in Cognition and Culture*, edited by Lawrence A. Hirschfeld and Susan A. Gelman, 85–116. New York: Cambridge Univ. Press, 1994.
- Crane, Mary Thomas.** *Shakespeare's Brain: Reading with Cognitive Theory*. Princeton: Princeton Univ. Press, 2001.
- Dennett, Daniel.** *The Intentional Stance*. Cambridge: MIT Press, 1987.

- DiBattista, Maria.** "Virginia Woolf and the Language of Authorship." In *The Cambridge Companion to Virginia Woolf*, edited by Sue Roe and Susan Sellers, 127–45. Cambridge: Cambridge Univ. Press, 2000.
- Dick, Susan.** "Literary Realism in *Mrs. Dalloway*, *To the Lighthouse*, *Orlando* and *The Waves*." In *The Cambridge Companion to Virginia Woolf*, edited by Sue Roe and Susan Sellers, 50–71. Cambridge: Cambridge Univ. Press, 2000.
- Dunbar, Robin.** "On the Origin of the Human mind." In *Evolution and the Human Mind: Modularity, Language, and Meta-Cognition*, edited by Peter Carruthers and Andrew Chamberlain, 238–53. Cambridge: Cambridge Univ. Press, 2000.
- Easterlin, Nancy.** "Making Knowledge: Bioepistemology and the Foundations of Literary Theory." *Mosaic* 32 (1999): 131–47.
- . "Voyages in the Verbal Universe: The Role of Speculation in Darwinian Literary Criticism." *Interdisciplinary Literary Studies: A Journal of Criticism and Theory* 2, no. 2 (Spring 2001): 59–73.
- Fish, Stanley.** "How to Recognize a Poem When You See One." In *American Criticism in the Poststructuralist Age*, edited by Ira Konigsberg, 102–15. Ann Arbor: Univ. of Michigan Press, 1981.
- . *Is There a Text in this Class?* Cambridge: Cambridge Univ. Press, 1980.
- Gomez, Juan C.** "Visual Behavior as a Window for Reading the Mind of Others in Primates." In *Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading*, edited by Andrew Whiten, 195–208. Oxford: Basil Blackwell, 1991.
- Grosz, Elizabeth.** "Feminist Futures?" *Tulsa Studies in Women's Literature* 21, no. 1 (Spring 2002): 13–20.
- Harris, Paul L.** *The Work of Imagination*. Oxford: Blackwell Publishers, 2001.
- Hart, F. Elizabeth.** "The Epistemology of Cognitive Literary Studies." *Philosophy and Literature* 25 (2002): 314–34.
- Hayles, N. Katherine.** "Desiring Agency: Limiting Metaphors and Enabling Constraints in Dawkins and Deleuze/Guattari." *Substance* 94/95: 30 (2001): 144–59.
- Hemingway, Ernest.** *A Farewell to Arms*. New York: Charles Scribner's Sons, 1929.
- Herman, David.** "Regrounding Narratology: The Study of Narratively Organized Systems for Thinking." In *What is Narratology?*, edited by Jan-Christoph Meister, Tom Kindt, and Hans-Harald Müller. Berlin: de Gruyter, forthcoming.
- . "Scripts, Sequences, and Stories: Elements of a Postclassical Narratology." *PMLA* 112 (1997): 1046–59.
- Hernadi, Paul.** "Literature and Evolution." *Substance* 94/95: 30 (2001): 55–71.
- Hogan, Patrick Colm.** "Literary Universals." *Poetics Today* 18 (1997): 223–49.
- Hughes, Claire, and Robert Plomin.** "Individual Differences in Early Understanding of Mind: Genes, Non-Shared Environment and Modularity." In *Evolution and the Human Mind: Modularity, Language, and Meta-Cognition*, edited by Peter Carruthers and Andrew Chamberlain, 47–61. Cambridge: Cambridge Univ. Press, 2000.
- Kanner, Leo.** "Austistic disturbances of affective contact." *Nervous Children* 2 (1943): 217–50.
- Lee, Hermione.** *Virginia Woolf*. New York: Alfred A. Knopf, 1997.
- . "Virginia Woolf's Essays." In *The Cambridge Companion to Virginia Woolf*, edited by Sue Roe and Susan Sellers, 91–108. Cambridge: Cambridge Univ. Press, 2000.
- Leslie Alan.** "ToMM, ToBY, and Agency: Core Architecture and Domain Specificity." In *Mapping the*

- Mind: Domain Specificity in Cognition and Culture*, edited by Lawrence Hirschfeld and Susan Gelman, 119–48. New York: Cambridge Univ. Press, 1994.
- Origgi, Gloria, and Dan Sperber.** “Evolution, Communication and the Proper Function of Language.” In *Evolution and the Human Mind: Modularity, Language, and Meta-Cognition*, edited by Peter Carruthers and Andrew Chamberlain, 140–69. Cambridge: Cambridge Univ. Press, 2000.
- Pinker, Steven.** *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking, 2002.
- Premack, David, and Verena Dasser.** “Perceptual Origins and Conceptual Evidence for Theory of Mind in Apes and Children.” In *Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading*, edited by Andrew Whiten, 253–66. Oxford: Basil Blackwell, 1991.
- Richardson, Alan.** *British Romanticism and the Science of Mind*. Cambridge: Cambridge Univ. Press, 2001.
- Richardson, Alan, and Francis Steen.** “Literature and the Cognitive Revolution: An Introduction.” *Poetics Today* 23 (2002): 1–8.
- Sacks, Oliver.** *An Anthropologist on Mars*. New York: Alfred A. Knopf, 1995.
- Scarry, Elaine.** *Dreaming by the Book*. New York: Farrar, Straus, Giroux, 1999.
- . “Panel Discussion: Science, Culture, Meaning Values.” In *Unity of Knowledge: The Convergence of Natural and Human Science*, 233–57. New York: The New York Academy of Sciences, 2001.
- Spolsky, Ellen.** “Cognitive Literary Historicism: A Response to Adler and Gross.” *Poetics Today*, forthcoming.
- . *Satisfying Skepticism: Embodied Knowledge in the Early Modern World*. Aldershot: Ashgate, 2001.
- . “Why and How to Take the Wheat and Leave the Chaff.” *Substance* 94/95 30, nos. 1–2 (2001): 178–98.
- Starr, Gabrielle G.** “Ethics, Meaning, and the Work of Beauty.” *Eighteenth-Century Studies* 35 (2002): 361–78.
- Sternberg, Meir.** “How Narrativity Makes a Difference.” *Narrative* 9 (2001): 115–22.
- Turner, Mark.** *The Literary Mind*. New York: Oxford Univ. Press, 1996.
- Vermeule, Blakey.** *The Party of Humanity: Writing Moral Psychology in Eighteenth-Century Britain*. Baltimore: The Johns Hopkins Univ. Press, 2000.
- Whitworth, Michael.** “Virginia Woolf and Modernism.” In *The Cambridge Companion to Virginia Woolf*, edited by Sue Roe and Susan Sellers, 146–63. Cambridge: Cambridge Univ. Press, 2000.
- Woolf, Virginia.** *The Diary of Virginia Woolf*. 5 vols. Edited by Anne Olivier Bell. London: Penguin, 1977–84.
- . *Jacob’s Room*. London: Hogarth, 1976.
- . *The Letters of Virginia Woolf*. Vol. 2. Edited by Nigel Nicholson. London: Hogarth Press, 1975–80.
- . *Mrs. Dalloway*. San Diego: Harcourt Brace, 1981.
- Zunshine, Lisa.** “Eighteenth-Century Print Culture and the ‘Truth’ of Fictional Narrative.” *Philosophy and Literature* 25 (2001): 215–32.
- . “Rhetoric, Cognition, and Ideology in Anna Laetitia Barbauld’s 1781 *Hymns in Prose for Children*.” *Poetics Today* 23 (2001): 231–59.
- . “Richardson’s *Clarissa* and a Theory of Mind.” In *The Work of Fiction: Cognition, Culture, and Complexity*, edited by Ellen Spolsky and Alan Richardson. Aldershot: Ashgate Press, forthcoming.