

Lisa Zunshine

Why Jane Austen Was Different,
And Why We May Need Cognitive Science to See It

[Forthcoming in *Style*]

Something happened to the novel “around the time of Jane Austen” (vii), argues George Butte in his compelling reintroduction of Maurice Merleau-Ponty’s discourse on phenomenology into contemporary literary and film studies, *I Know That You Know That I Know: Narrating Subjects from Moll Flanders to Marnie*. English writers began to portray a multiply-layered and mutually-reflecting subjectivity—*deep intersubjectivity*—a “change so subtle and fundamental that it has been difficult to conceive and describe” (25), particularly as today we take its impact for granted in the prose of George Eliot, Henry James, James Joyce, Virginia Woolf, Vladimir Nabokov, Ian McEwan, and others. Butte defines deep intersubjectivity as

the web of partially interpenetrating consciousnesses that exists wherever perceiving subjects, that is, human beings, collect. [T]he process begins when a self perceives the gestures, either of body or word, of another consciousness, and it continues when the self can perceive in those gestures an awareness of her or his own gestures. Subsequently the self, upon revealing a consciousness of the other’s response, perceives yet another gesture responding to its response, so that out of this conversation of symbolic behaviours emerges a web woven from elements of mutually exchanged consciousnesses. (27)

For a vivid early example of deep intersubjectivity, Butte turns to the episode in Austen’s *Persuasion*, in which Anne Elliot witnesses a silent but poignant communication between her former suitor, Frederick Wentworth, and her sister, Elizabeth, who run into each other in Molland’s bakery shop:

It did not surprise, but it grieved Anne to observe that Elizabeth would not know [Wentworth]. She saw that he saw Elizabeth, that Elizabeth saw him, that there was complete internal recognition on each side; she was convinced that he was ready to be acknowledged as an acquaintance, expecting it, and she had the pain of seeing her sister turn away with unalterable coldness. (117)

According to Butte,

When Anne Elliot watches Wentworth and Elizabeth negotiating complex force fields of memory and protocol, the enabling strategy of her story is a new layering of human consciousness, or a new representation of those subjectivities as layered in a specific way. Deep intersubjectivity has made its appearance in storytelling in

modern culture, and it has altered our sense of self and community and the discourses that construct and reflect them. (4)

Was Austen the first English writer to construct deeply intersubjective passages? According to Butte, some of her eighteenth-century predecessors came close to but stopped short of exploring the rich possibilities opened by having a character perceive the reaction of another character to the first character's mental state. In the novels of Defoe, Richardson, Fielding, and Burney, the "encounter with the other never moves beyond a two-layer exchange to multiple negotiations and perceptions," whereas in Austen the scenes "about the observation of observations" give voice to a "new way of shaping narrative" (59).

Butte's argument thus lends support to the critical view that Austen was profoundly innovative in her treatment of fictional consciousness, a view that gets obscured when her novels are treated as archetypes of the genre. Moreover, his exploration of mutually reflecting fictional subjectivities turns out to dovetail research in evolutionary psychology that focuses on cognitive challenges of processing multiple mental states embedded within each other. In other words, although Butte does not position himself as working within the new field known as cognitive approaches to literature,ⁱ his argument provides a crucial first step for recognizing Austen's prose as actively experimenting with readers' cognition. The goal of the present essay is to articulate this interdisciplinary potential of the concept of deep intersubjectivity and, more broadly, to demonstrate how a cognitive approach encourages us to see fictional narratives as engaging our evolved cognitive adaptations: playing with these adaptations and pushing them beyond their zones of comfort.

I. A Mind Within a Mind Within a Mind

To speak of mental states in works of fiction, we need to borrow from cognitive science the concept of *Theory of Mind*. Cognitive psychologists and philosophers of mind use this term interchangeably with *mind-reading* to refer to our ability to explain observable behavior in terms of underlying thoughts, feelings, desires, and intentions (for example, we see somebody reaching for a cup of water and immediately assume that she is thirsty). We attribute mental states to others and to ourselves all the time. Our attributions are frequently incorrect (for example, the person who reached for the cup of water might have done it for reasons other than being thirsty); still, reading minds is the default way by which we construct and navigate our social environment. When Theory of Mind is impaired, as it is in varying degrees in the case of autism and schizophrenia, communication breaks down.ⁱⁱ

Cognitive evolutionary psychologists working with mind-reading adaptations think that they must have developed during the "massive neurocognitive evolution" that took place during the Pleistocene (1.8 million to 10,000 years ago). The emergence of Theory of Mind 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. As Simon Baron-Cohen points out, "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).ⁱⁱⁱ In other words, mind-reading is both predicated on the intensely social nature of our species *and* makes this intense social nature possible. Lest this argument strike you as circular, think of our legs: their shape is both predicated upon the evolution of our species' locomotion *and* makes our present locomotion possible.

Note that the words *theory* in Theory of Mind and *reading* in mind-reading are potentially misleading because they seem to imply that we attribute states of mind intentionally and consciously. In fact, it might be difficult for us to appreciate at this point just how much mind-reading takes place on the level inaccessible to our consciousness. For it seems that while our perceptual systems "eagerly" register the information about people's bodies and their facial expressions, they do not necessarily make all that information available to us for our conscious interpretation. Think of the intriguing functioning of the so-called "mirror neurons." Studies of imitation in monkeys and humans have discovered a "neural mirror system that demonstrates an internal correlation between the representations of perceptual and motor functionalities" (Borenstein and Ruppin, 229). What this means is that "an action is understood when its observation causes the motor system of the observer to 'resonate.'" So when you observe someone else grasping a cup, the "same population of neurons that control the execution of grasping movements becomes active in [your own] motor areas" (Rizzolatti et al, 662). At least on some level, your brain does not seem to distinguish between you doing something and a person that you observe doing it.

In other words, our neural circuits are powerfully attuned to the presence, behavior, and emotional display of other members of our species. This attunement begins early (since some form of it is already present in newborn infants) and takes numerous nuanced forms as we grow into our environment. We are intensely aware of the body language and facial expressions of other people, even if the full extent and significance of such awareness escape us. Moreover, as I have argued elsewhere, the workings of our Theory of Mind make literature as we know it possible.^{iv} 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 would allow us to guess at their feelings and thus predict their actions. Literature pervasively capitalizes on and stimulates our Theory of Mind adaptations that had evolved to deal with real people, even as on some level readers do remain aware that fictive characters are not real people at all.

Research and publications on Theory of Mind grow apace, but for the purposes of this essay, I want to focus on one particular line of that research, which explores our ability to process multiply embedded states of mind, as in, "I know that you know that I know that . . ." Evolutionary psychologist Robin Dunbar and his colleagues have demonstrated that we have marked difficulties processing units of information that embed more than four recursive mental states. After the fourth level, our understanding plummets sixty percent.^v For a brief illustration, look at the following three sentences (the words that describe mental states are italicized):

(1) I *find* Butte's argument *compelling* on its own terms and also full of exciting possibilities when *considered* in the context of recent research on Theory of Mind and its implications for literary studies.

(2) I *thought* that she *wanted* me to *think* that she *liked* what I was saying, so I *pretended not to notice* that her eyes glazed over now and then.

(3) I *thought* that she *wanted* me to *think* that she *liked* that I *believed* that she was *interested* in what I was saying."

I am certain that if I ask you which of these three sentences is the most difficult to understand, you will say the third, even though it is shorter than the other two. And indeed, the third sentence embeds six mental states, which makes it extremely challenging for our Theory of Mind. By contrast, the first sentence, though the longest of the three, is the easiest to understand because it contains two parallel pairs of mental embedment, none raising above the third level: I *think* that Butte's argument will *convince* people; and I *think* that they will be *excited* when they *consider* his argument in the context of research on Theory of Mind. The second sentence contains four levels of recursive embedment^{vi} and as such it must strike you as more difficult—or awkward, or contrived—than the first sentence, but certainly as easier to understand than the third. For a quick illustration of this point, consider Daniel Dennett's classic example: "I suspect that you wonder whether I realize how hard it is for you to be sure that you understand whether I mean to be saying that you can recognize that I believe you want me to explain that most of us can keep track of only about five or six orders of intentionality" (243).

I have argued elsewhere that modernist writers, such as Woolf, experimented with our mind-reading capacities, by pushing their portrayals of fictional subjectivity to the sixth level of recursive embedment.^{vii} Hence the reason I am excited by Butte's argument. Putting what he does in the context of cognitive theory, we can say that he identifies both the moment in English literary history (i.e., the late eighteenth century) when the portrayal of fictional subjectivity moved from the second-third level of mental embedment to the third-fourth level and the writer (i.e., Austen) who consummated this transition.

The difference in treatment of fictional subjectivity is indeed striking when we contrast the most intersubjective moment in Daniel Defoe's *Moll Flanders* (1722) with the most intersubjective moment in *Persuasion* (1818). For example, Butte thinks that the scene in Defoe's novel when Moll and her new husband "drop their masks" and "confess their mutual schemes to marry well" is "ripe with intersubjective possibilities." Still, these possibilities remain unrealized: "never in his or Moll's or Defoe's text, does Jemmy reflect on Moll as a subjectivity and much less on her consciousness of *him*" (43). Here is one passage from that scene, following Moll's showing Jemmy all the money that she has in the world (presumably) and offering it to him, "if he would take it":

He told me with great concern, and I thought I saw tears in his eyes, that he would not touch it; that he abhorred the thoughts of stripping me and making me miserable . . . (118)

Or, to think about it in terms of Theory of Mind:

(4) Moll *knows* (based even more on Jemmy's tears than his words) that Jemmy *suffers* at the thought of making her *miserable*.

This is the third level of embedment. Contrast it now with the cognitive mapping of the already-discussed scene in *Persuasion*:

(5) Anne *realizes* that Wentworth *understands* that Elizabeth *pretends not to recognize* that he *wants* to be *acknowledged* as an acquaintance.

This is the fourth level of embedment, reaching, perhaps, even further: up to the fifth level. I suspect that moments like this—when we are faced with five levels of mental embedment—are not too frequent in Austen, but that on the whole she does operate on at least third level of embedment, with frequent forays into the fourth. Similarly, the moments in Defoe in which we have three levels of subjective embedment (as in the scene above) are much less frequent than the scenes embedding two levels.^{viii}

The realization that we can “calculate” the levels of fictional subjectivity opens up intriguing interpretive opportunities. If we can locate a historical moment—and there must have been several such moments in the last three hundred years—when the novel learns to function comfortably one intersubjective level up,^{ix} we can speculate about particular cultural circumstances behind this learning curve. Those may include the socioeconomics of textual reproduction, personal histories of the authors, the availability of readers open to this new, more challenging, but also perhaps more exciting construction of fictional consciousness, and so forth. Although we shall never be able to produce the exact and exhaustive list of these circumstances,^x discussing them implies a productive interdisciplinary dialogue. Cognitive evolutionary psychology contributes to such a dialogue the awareness of challenges involved in processing multiply-embedded subjectivities. Literary and cultural studies contribute analysis of social-political, aesthetic, and personal factors influencing the production and dissemination of texts featuring such an embedment, as well as of the far-ranging effects of the deepening of fictional intersubjectivity.

In the rest of this essay, I will look at several fictional and non-fictional representations of multiply embedded consciousness that precede those in Austen. My immediate goal is to understand how Austen's experimentation with deep intersubjectivity built upon that of her predecessors. My long-term objective is to make sure that a cognitive framework indeed offers something not available through more traditional literary-critical methodologies. Specifically, I want to see if, aided by research in Theory of Mind, I will be compelled to ask questions that I wouldn't have otherwise and trace new connections between different cultural discourses of the “long” eighteenth century.

II. Representing Greedy Self-Consciousness

Although (and perhaps *because*) Butte mentions *Clarissa* (1747-48) only in passing,^{xi} one is immediately tempted to apply his model of the new interpersonal consciousness to Samuel Richardson's *magnum opus*. Surely, a novel whose protagonists spend their waking hours trying to plan each other's emotional reactions and so to

deflect, with varying degrees of success, each other's mental gambits *must* create something of that "field of mutual consciousness" (33) that Butte sees as coming into existence only in the novels of Austen. The readers of *Clarissa* may recall, for example, the "Miss Partington" episode, in which Lovelace describes himself (in a letter to Belford) observing Clarissa closely as she speaks and inferring that she does not want him to think that she thinks that he has had some ulterior motives in wanting her earlier to share a bed with a young lady of his acquaintance.

Having discussed that episode elsewhere,^{xii} I turn now to two other texts featuring similar moments of multiply-embedded subjectivity: Rousseau's *Emile* (1762) and Sterne's *Sentimental Journey* (1768). Rousseau's philosophical-treatise-cum-novel contains passages such the one in which Emile's tutor, Jean-Jacques, aware both of his pupil's attraction to Sophie and his embarrassment at thinking that others may be aware of it too, observes Emile sitting at Sophie's parents' table unable to look up at the people surrounding him:

Confused, embarrassed, fearful, he no longer dares to look around him for fear of seeing that he is being looked at. Ashamed to let others see through him, he would like to make himself invisible to everyone in order to sate himself with contemplating her without being observed. Sophie, on the contrary, is reassured by Emile's fear. (415)

Here is one possible mapping of the mental embedment present in this scene: The tutor *observes* Sophie's feelings as she *realizes* that Emile is *afraid* that others will *understand* that he is *falling in love* with her. It is not for nothing that Alan Bloom saw *Emile* as "*Phenomenology of the Mind* posing as Dr. Spock" (5). Though not frequent, such moments of multi-level mind-reading are crucial to Rousseau's narrative.

In *A Sentimental Journey*, we encounter Yorick writing a card to Madame de R**** in his hotel room while "the fair *fille de chambre*" is waiting for him to finish the card, so she can deliver it:

It was a fine still evening, in the latter end of the month of May—the crimson window-curtains (which were of the same colour of those of the bed) were drawn close—the sun was setting, and reflected through them so warm a tint into the fair *fille de chambre*'s face—I thought she blushed—the idea of it made me blush myself—we were quite alone; and that super-induced a second blush before the first could get off. (116)

One way of mapping this passage is to focus on the first blush, spelling out Yorick's speculation about the feelings of the young woman: Yorick blushed because he *thought* that she blushed because she *thought* about the two of them alone in the room; or even: Yorick blushed because he *thought* that she blushed because she *thought* that he *thought* about the two of them alone in the room.

Pushing the appearance of the deep intersubjectivity some thirty years back, from the 1790s to the mid-century, does not quarrel with Butte's argument. The eighteenth-century sentimental novel (of which *Clarissa* and *Emile* are prime examples) valorized attention to the body's "vocabulary . . . of gestures and palpitations, sighs and tears" and

as such carefully foregrounded its descriptions of “mutually affecting looks” (Mullan, 77). Although we do not think of Austen’s novels as sentimental, she certainly builds on Richardson’s, Sterne’s, and Rousseau’s experimentation with representing closely observed, interpreted, and misinterpreted bodies. One crucial difference between Austen and those earlier writers, however, is that they highlight their characters’ “attention to the meaning of looks and gestures” (Mullan, 77), whereas she takes that attention for granted.

For example, Clarissa’s friend exhorts her to be “vigilant” (41), and Clarissa repeatedly assures her that she is, and Lovelace testifies that he and Clarissa “are both great watchers of each other’s eyes” (460). Rousseau’s Jean-Jacques is convinced that his “true function” is to be “the observer and philosopher who knows the art of sounding hearts while working to form them”; accordingly, he reads in Emile’s “face all the movements of his soul,” and by “dint of spying them out, . . . gets to be able to foresee them and finally to direct them” (226). Sterne’s Yorick prides himself on being “quick in rendering the several turns of looks and limbs, with all their inflections and delineations, into plain words” and boasts that he does it constantly:

For my own part, by long habitude, I do it so mechanically, that when I walk the streets of London, I go translating all the way; and have more than once stood behind in the circle, where not three words have been said, and have brought off twenty different dialogues with me, which I could have fairly wrote down and sworn to. (79)

This is very different from Austen, whose prose seems to have fully internalized the assumption that some people make a careful study of the “turns of looks and limbs.” When she does portray characters who describe themselves as particularly attuned to the emotions of others and their own, she does it to parody the sentimental novels. Hence Laura in *Love and Freindship* (1790), written when Austen was fourteen, informs her correspondent that a “sensibility too tremblingly alive to every affliction of my Freinds, my Acquaintance and particularly to every affliction of my own, was my only fault, if a fault it could be called” (3). By contrast, in *Persuasion*, Anne, who apparently watches people very closely, is never explicitly described as doing so. Instead we get the results of her observation, such as the exchange of looks in Molland’s bakery shop (above), or the description of the expressive glance of Captain Wentworth as he turns quickly to see Anne’s face when he notices that a strange man looks at her “with a degree of earnest admiration” (77).

Similarly, neither Elinor in *Sense and Sensibility* nor Fanny in *Mansfield Park* are explicitly characterized as being over-sensitive to other people’s body language, even though it is precisely their heightened sensitivity that makes possible the intersubjective moments present in those novels. In fact, in *Sense and Sensibility*, Elinor is understood to be much less “sensitive” than her sister Marianne—a conventional understanding of sensitivity that appears particularly ironic when we realize that no moments of deep intersubjectivity can originate with Marianne because for two-thirds of the novel she is too enclosed within her own emotions to reflect and re-reflect those of other people. In other words, it seems that by freeing her heroines from the compulsion to register and praise (or lament) their emotional responsiveness to themselves and others, Austen opens up new venues for exploring such responsiveness.

Studies in Theory of Mind suggest why this may work. When the character does not have to *contemplate* her *contemplating* other people's emotions, it frees up one level of intentionality, which can then be used to add an extra level of embedment on another end. Compare the two following mappings. One, already familiar to you, describes Anne's observing Elizabeth and Wentworth in the bakery shop; another (hypothetical) adds to it Anne's awareness of her perceptive self:

(6) Anne *realizes* that Wentworth *understands* that Elizabeth *pretends not to recognize* that he *wants* to be *acknowledged* as an acquaintance.

(7) Anne is *aware* that her keen powers of observation allow her to *realize* that Wentworth *understands* that Elizabeth *pretends not to recognize* that he *wants* to be *acknowledged* as an acquaintance.

The second sentence is a mouthful (or a *mindful*) because it pushes us to the sixth level of mental embedment, and the only recompense that we get for our pains as we strive to grasp its overall meaning is being reminded what a smart girl our Anne is. Apparently, *she* has no problems keeping track of who thinks what here. This is enough to make us feel a twinge of resentment toward the hitherto favorite heroine.

If we are in the pre-Austenian world of the sentimental novel and thus insist on keeping that self-congratulatory tag, we have to simplify the rest of the sentence in order to render it comprehensive. A simplified version embedding four levels of subjectivity might look like this:

(8) Anne is *aware* that her keen powers of observation allow her to *realize* that Wentworth *understands* that Elizabeth *pretends not to recognize* him.

Or like this:

(9) Anne is *aware* that her keen powers of observation allow her to *realize* that Elizabeth *pretends not to recognize* that Wentworth *wants* to be *acknowledged* as an acquaintance.

Note how both of these simplified versions make Anne appear less self-satisfied than in example 7. The cognitive perspective helps us to account for this slight shift of tone. In example 7, we had to operate on the difficult sixth level of mental embedment, and it seemed that all that difficulty originated with Anne, who smugly appreciated that she could keep track of who was thinking what, and whose pleased self-awareness seemed particularly out of place given all the negative emotions experienced by Elizabeth and Wentworth.^{xiii}

By contrast, examples 8 and 9 do not present us with the same cognitive challenge—we are now on the fourth level of embedment—which apparently makes us amenable to a different reading of Anne. We don't perceive her as being smugly self-satisfied anymore. Instead, she seems to be *grateful* for being capable of keeping track of who thinks what. After all, had she missed or misread either Elizabeth's or Wentworth's reaction to each other, she could have done or said something that would have resulted in

mutual awkwardness and unease. You can see how tweaking the sentence to make it either more or less challenging for our mind-reading adaptations can contribute, at least to some extent, to adjusting its emotional timbre.

In general, what the contrast between example 7, on the one hand, and 8 and 9, on the other, demonstrates is that a writer enters a treacherous territory when she begins to work above the fourth level of recursive mental embedment. Something's got to give if she is to keep a given passage both deeply intersubjective and reader-friendly, *and* to avoid making the heroine seem self-satisfied about her powers of penetration.

Let me make clear what I am *not* saying here. I am not saying that a writer who does *not* depict characters strongly aware of their own sensibility will necessarily end up exploring deep intersubjectivity. The case in point is Defoe, whose prose is neither sentimental nor deeply intersubjective. I am saying rather that it so happened that during this particular period of literary history, writers who came to be identified with sentimentalism (e.g., Richardson, Rousseau, Sterne) experimented with multiply-embedded subjectivity by depicting characters whose awareness of their sensibility claimed two levels of mental embedment. To adapt Butte's memorable title, certain scenes from sentimental novels could be mapped not as "I know that you know that I know," but as "I know that I know that you know that I know." Austen disliked sentimentality, so she eliminated that greedy self-consciousness from her own experimentation with multiply-embedded subjectivity. (I call the sentimentalist self-consciousness greedy because it requires two levels of mental embedment).

Moreover, it is *not* the case that in the novels of Richardson, Rousseau, and Sterne, *all* deeply intersubjective passages cultivate this greedy self-consciousness by featuring protagonists who are intensely aware of the importance of observing others' and their own body language and who carefully monitor their progress in this social game. Some passages do and some don't. It is important, however, that we do not seem to encounter any such passages in Austen, and here is why it is important:

I strongly agree with Butte that Austen's treatment of fictional intersubjectivity *is* different. I want to demonstrate, however, that even if we do find instances of deep intersubjectivity in earlier eighteenth-century novels, it does not invalidate his point. Richardson, Rousseau, and Sterne all occasionally operated on the fourth (or even fifth) level of mental embedment, but Austen "learned"^{xiv} new ways of doing it, so that her protagonists on these occasions would appear neither Machiavellian (as the two main characters sometimes do in *Clarissa*^{xv}), nor smug (as the narrator in *A Sentimental Journey*), nor overbearingly controlling (as Jean-Jacques in *Emile*).

III. Downgrading Deep Intersubjectivity

Freeing up one level of embedment for other uses by making the protagonist through whose eyes we see the scene less sensitive to her own sensitivity was just one strategy adopted by Austen in her construction of "reader-and-character-friendly" deep intersubjectivity. I am about to consider two other strategies, but I want to stress that these three are not exhaustive. In fact, when it comes to understanding how Austen experimented with our Theory of Mind in general and multiply-embedded subjectivities in particular, what I am doing here merely scratches the surface of that discussion.

And so to a non-fictional text published around the time of Richardson's *Pamela*: David Hume's *Treatise of Human Nature* (1739-40). Hume's *Treatise* contains what seems to be a perfect instantiation of Butte's image of the self that "upon revealing a consciousness of the other's response, perceives yet another gesture responding to its response, so that out of this conversation of symbolic behaviours emerges a web woven from elements of mutually exchanged consciousnesses" (27):

In general we may remark that the minds of men are mirrors to one another, not only because they reflect each other's emotions, but because those rays of passions, sentiments and opinions, may often be reverberated, and may decay away by insensible degrees. Thus the pleasure, which a rich man receives from his possessions, being thrown upon the beholder, causes a pleasure and esteem; which sentiments again, being perceived and sympathized with, encrease the pleasure of the possessor; and being once more reflected, become a new foundation for pleasure and esteem in the beholder. (365)

The last sentence feels a bit peculiar, and mapping it out in terms of embedded intentionalities explains why. "The beholder is *happy* to see the possessor *happy* to see the beholder *happy* to see the possessor *happy* at his possessions." This is the fourth level of embedment, which, as we know, is not the easiest level for us to function on. Moreover, my teaching experience suggests that presented with this passage and asked to convey its meaning, students respond with, "people are happy to see other people happy." This simplification makes perfect sense in light of Dunbar's findings: it still captures the gist of Hume's example but it also downgrades the levels of intentionality from the difficult four to the comfortable two.

Hence my second surmise about the difference between this type of multi-level embedment and the one encountered in Austen. It turns out that it is rather difficult to downgrade her fourth-level subjectivity to the second level and still convey the meaning of the scene. For example, were we to do it with the encounter in the Molland's bakery shop, it would go something like this, "Anne notices the exchange of looks between Wentworth and Elizabeth." This sounds not just simplified, but plainly wrong, because the first interpretation of this sentence that comes to mind is that Wentworth and Elizabeth might be interested in each other.

What happens if we try conveying the gist of the scene by downgrading it not to the second but to the third level of embedment? We can say, "Anne notices the exchange of looks between Wentworth and Elizabeth and realizes that Elizabeth still dislikes Wentworth," or, "Anne notices the exchange of looks between Wentworth and Elizabeth and realizes that Wentworth is trying to be friendly with Elizabeth." Both of these still misconstrue the meaning of the passage! To get finally to the correct meaning we would have to say, "Anne notices the exchange of looks between Wentworth and Elizabeth and realizes that even though Wentworth is trying to be friendly with Elizabeth, she still dislikes him, and he is now becoming aware of her continuous dislike." We are back in our fourth-to-fifth level embedment, even if the phrasing does not sound like Austen and does not convey other important nuances of the episode.

Again, I am wary of generalizations here. I am not saying that *none* of Austen's intersubjective scenes can be paraphrased by downgrading its levels of mental embedment to the super-comfortable second level or to the still-rather-comfortable third. Perhaps some of them can, and I would be happy to hear from the readers who find such scenes. What I am saying is that Austen learned to construct her deeply-intersubjective passages in such a way that the deep intersubjectivity is organic to their meaning: eliminate one or two levels of mental embedment, and you lose the meaning. But once Austen had this particular representational tool in her writing toolbox, she could use it on some occasions and not on others, just as she would any other rhetorical strategy.

IV. The Importance of Emotional Investment

To figure out yet another rhetorical strategy used by Austen in her construction of deep intersubjectivity, we need to switch genres and go further back into literary history, for it seems that at least some English playwrights waxed deeply intersubjective much earlier than their novel-writing counterparts.

George Etherege's *The Man of Mode; or, Sir Fopling Flutter* premiered at the Duke's Theatre in 1676, remained popular with audiences until the 1750s, and was then "dropped from the repertory" apparently because "changing tastes made its sexual frankness seem objectionable" (O'Neill, 526). The play showcases the sexual charm and "Machiavellian" wit (Robert D. Hume, 96) of Mr. Dorimant, a character often compared to John Wilmot, the First Earl of Rochester. Dorimant simultaneously breaks up with one woman, Mrs. Loveit, seduces another, Bellinda, and falls in love with a third, Harriet. In the concluding scene, he faces and has to mollify all three of them, while getting ready to follow Harriet to her family's countryseat, where (we assume) he will court and marry her. It is in figuring out how the actions of Dorimant and Bellinda are influenced by what they know about each other's thoughts that we become aware how deeply intersubjective this scene might be. I say might be because some of its intersubjective possibilities are easily lost in the reading, but they can be, and I think *must be*, magnified in the performance.

To tease out those possibilities let us establish who knows what throughout the play. Mrs. Loveit knows that Dorimant has been courting another woman and suspects that this is why he is breaking up with her, but she does not know who that other woman is. The other woman is Bellinda, but because she is Mrs. Loveit's avowed friend and, moreover, is anxious to preserve her reputation, she does not want anybody to know that she is about to become Dorimant's mistress. Moreover, Bellinda wants proof that the affair between Dorimant and Mrs. Loveit is over. To get this proof, she first fans Mrs. Loveit's jealousy about the masked woman with whom Dorimant has been recently seen in the theatre (who was Bellinda herself) and then arranges it so that she is present when Dorimant quarrels with Mrs. Loveit on false pretenses and breaks with her. Witnessing Dorimant's ruthless handling of Mrs. Loveit makes Bellinda unhappy because she is afraid that he might later treat her the same way, but she still finds him irresistible and so spends a night with him. When she leaves his house at dawn, the chairman in attendance at Dorimant's door mistakenly takes her not to her own house but to the house of Mrs. Loveit, since he is used to returning a woman to that particular lodging in the morning. While Bellinda is at Mrs. Loveit's, Dorimant comes in, to see how much emotional power

he still has over Mrs. Loveit. Bellinda and Dorimant are both surprised by the meeting and begin to suspect each other of double-dealing.

And all this while, Dorimant has been pursuing Harriett, who is beautiful, witty, and rich, but will not succumb to his charms without marriage. In the last scene, Mrs. Loveit and Bellinda learn that Dorimant is about to marry Harriet. To mollify Mrs. Loveit (who can damage his standing in the eyes of Harriet's mother), Dorimant confides to Mrs. Loveit that Harriet *is* the masked woman of Bellinda's story, and, moreover, that he has been pursuing Harriet not because he loves her—he claims that he loves Mrs. Loveit still—but because he needs to marry this heiress “to repair the ruins of [his] estate.” Dorimant tells all this to Mrs. Loveit within the hearing of Bellinda, who is both relieved and mortified by these lies. She is relieved because her reputation is safe—Mrs. Loveit will not accuse her of betraying their friendship and the town won't know that she has slept with Dorimant—and she is mortified because Dorimant has now abandoned her too, and for the woman (Harriet) whom he apparently loves. Here is how it is depicted in the play:

Dorimant. [to Mrs. Loveit] . . . Be satisfied—this is the business; this is the mask has kept me from you.

Bellinda. (Aside.) He's tender of my honor, though he's cruel to my love.

Mrs. Loveit. Was it no idle mistress, then?

Dorimant. Believe me, a wife, to repair the ruins of my estate that needs it.

Mrs. Loveit. This knowledge of this makes my grief hang lighter on my soul, but I shall never more be happy.

Dorimant. Bellinda—

Bellinda. Do not think of clearing yourself with me. It is impossible—Do all men break their words thus?

Dorimant. Th'extravagant words they speak in love. 'Tis as unreasonable to expect we should perform all we promise then, as do all we threaten when we are angry. When I see you next—

Bellinda. Take no notice of me, and I shall not hate you.

Dorimant. How came you to Mrs. Loveit?

Bellinda. By a mistake the chairmen made for want of my giving them directions.

Dorimant. 'Twas a pleasant one. We must meet again.

Bellinda. Never. (585)

To grasp the full irony of the scene, the spectators have to understand that Dorimant lies brazenly to Mrs. Loveit within Bellinda's earshot because he knows that Bellinda cannot contradict his story because she is afraid that Mrs. Loveit will realize that Bellinda lied to her earlier. Moreover, Dorimant understands that Bellinda knows that he knows what a vulnerable position she is in and that she resents him for happily exploiting her present vulnerability. This is why Dorimant is not surprised when Bellinda tells him that he cannot clear himself with her and it is all over between them. Mapping this scene in terms of embedded levels of subjectivity gives us the following:

(10) Bellinda *knows* that Dorimant *knows* that Bellinda will not contradict his story because she is *afraid* that Mrs. Loveit will *realize* that Bellinda was *deceiving* her.

And if this were not complicated enough already, we can uncover yet another level of mental embedment here if we ask ourselves why Dorimant chooses this exact moment to ask Bellinda what she was doing at Mrs. Loveit's the morning after their assignation. Dorimant was "confounded" (579) by that meeting and suspected that his new mistress was playing some double game by conferring with his old mistress behind his back. By putting the question to her *now*, as she tells him that she can barely stand him ("Take no notice of me, and I shall not hate you"), he attempts to turn the tables on her. By asking "How came you to Mrs. Loveit?" Dorimant reminds Bellinda that he, too, has a reason to feel injured by her (presumed) double-dealing and that his present "cruelty to her love" is but a fit recompense for her past behavior. It is likely that Dorimant does not really feel as injured by Bellinda as he pretends to be because as soon as she explains what happened that morning ("By a mistake the chairmen made for want of my giving them directions"), he immediately relents and asks for another assignation ("Twas a pleasant one. We must meet again"), to which Bellinda, who has finally learned her lesson (one hopes), responds "never."

Thus, if we consider Dorimant's attempt to make Bellinda think that he knows perfectly well that he is being "cruel to [her] love" but feels that as an injured lover he has a right to treat her badly, we get the following:

(11) Dorimant *wants* Bellinda to *know* that he *knows* that Bellinda will not contradict his story because she is *afraid* that Mrs. Loveit will *realize* that Bellinda was *deceiving* her.

We are now in such depths of deep intersubjectivity that I am reminded of what Robert D. Hume once wrote about *The Man of Mode*. Responding to critics' speculations about the "central [philosophical] concern" of the play, Hume observed, "I suspect that Etherege's 'central concern' was to display his own wit" (96). Indeed, imagining our author hard at work constructing his plot so that the last scene will allow for six levels of mental embedment makes one wonder what philosophical issues he might have also hoped to smuggle in as his audience struggled to keep up with Dorimant's and Bellinda's multilevel representations of each other's mental states.

Although mapping out those representations the way I did above shows that *The Man of Mode* can be "cognitively challenging"—the sixth level of embedment is well beyond our comfort zone—simply reading the play still does not convey and may in fact misrepresent its full intersubjective potential. That potential only comes alive in actual performance. To get a flavor of the difference, let us then imagine one small part of the exchange quoted above as performed by actors on stage.

Butte provides us with an excellent framework for such visualization by reminding us that in the embodied conversation, words and gestures play against each other in the ever-increasing complexity of intended and misinterpreted meanings. He does not deal with theatre in his book, focusing instead on the movies of Howard Hawks,

Alfred Hitchcock, and Woody Allen and thus factoring into his analysis of screen subjectivities the point of view of the camera. Still, his argument about cinematographic theatricality is highly relevant to our discussion:

Theatricality is inherently intersubjective because gestures and language in 'performance' embody for or against the other the framer's intentions. Theatricality becomes deeply intersubjective when its narrative tracks the embodiments of gestures responding to gestures that are themselves interpretations or distortions of yet prior signs. (134)

In particular, what Butte says about the construction of deep intersubjectivity in Hitchcock movies works well when applied to the last scene of *The Man of Mode*: "In such scenes a public conversation both conceals and exposes a private conversation that transacts its business by way of double meanings in public language and private gestures" (144). When Dorimant is assuring Mrs. Loveit that he has been pursuing Harriet because he needs a rich wife, their conversation is "public" because it includes Bellinda (who, as Mrs. Loveit's bosom friend, is allowed into her secrets). At the same time, this public conversation also "conceals and exposes" a private conversation between Bellinda and Dorimant, which indeed "transacts its business by way of double meanings . . . and private gestures." For example, Bellinda may visibly start at Dorimant's claim that Harriet was the "mask [that] has kept [him] from [Mrs. Loveit]," and this gesture will be interpreted differently by Mr. Loveit and by Dorimant. Mrs. Loveit will think that Bellinda as her friend is startled by the revelation that Dorimant was hunting for a wife all this time and concealing it from Mrs. Loveit, and we will register this "public" (and incorrect) interpretation. Dorimant, however, will interpret that gesture correctly, and he will turn to Bellinda to start reproaching her for *her* presumed double-dealing.

Moreover, Mrs. Loveit continues to dwell within the realm of "public" conversation that includes herself, Dorimant, and Bellinda. Her body language must thus reflect what Butte would call her response to "misinterpreted gestures." We presume that she does not hear the part of the conversation in which Dorimant vindictively asks Bellinda how she happened to be that morning at Mrs. Loveit's house. This means that she has to have some kind of explanation for the fact that Bellinda and Dorimant are still talking together. If Bellinda looks angry during that conversation—and most likely she does, given what she says to Dorimant—Mrs. Loveit will interpret her angry posture as her indignation at Dorimant's shabby treatment of Mrs. Loveit. The body language of Mrs. Loveit may thus reflect some satisfaction at seeing her friend angry on her behalf, and the viewers will be aware that her interpretation is both correct in the context of the "public" meaning of the conversation and incorrect in the context of the private exchange between Dorimant and Bellinda. And, on top of it, Bellinda might be aware of Mrs. Loveit's incorrect interpretation and further respond to it by her own body language, for as she leaves Dorimant and rejoins Mrs. Loveit, she may continue to look indignant, presumably *still* feeling bad for her friend.

At least this is what I would ask actors to convey, were I to direct this scene. No matter how you direct it, however, it is clear that the actual performance deepens further and calibrates the deep intersubjectivity scripted by Etherege. The *performed* deep

intersubjectivity thus reaches out to levels and nuances of mental embedment that are not captured by my earlier mapping of the *read* deep intersubjectivity.

What I find most fascinating about the gap between the two is that the constraints discussed by cognitive scientists must operate differently when the deep intersubjectivity is performed, or rather, *performed in some particular way*. We saw already that our reading comprehension of passages featuring multiply-embedded intentionality goes sharply down after the fourth level of embedment and continues to decline after that. In fact, this applies to some forms of embodied representations as well. Think of Bruce Eric Kaplan's cartoon from *The New Yorker*, in which the gloomy husband assures the equally gloomy wife: "Of course I care about how you imagined I thought you perceived I wanted you to feel." Although this is a *somewhat* embodied sentiment—we can *see* the alienated bodies of the speakers—its level of embedment (sixth!) renders it almost undecipherable. (Which, of course, is exactly the point of the joke: the statement about mutual sensitivity, caring, and understanding is literally incomprehensible.) Or think of the "The One Where Everybody Finds Out" episode from the fifth season of *Friends*, in which Phoebe and Rachel plot to play a practical joke on Chandler and Monica, but then the plotting and counterplotting get slightly of out hand. The ever-eloquent Phoebe captures the end result as follows: "They thought they could mess with us! They're trying to mess with *us*? They don't know that we know they know we know!" Again, as in the case of *The New Yorker* cartoon, much of the humor of the episode stems from its growing incomprehensibility, and that's in spite of the fully embodied nature of this representation.

In other words, there is no guarantee that if you take the cognitively challenging level of mental embedment—say, sixth—and *perform* it instead of merely *describing* it, the audience's understanding will be enhanced. In fact, the opposite could be true. The playwright who wishes to "display his own wit" may end up flat on his face because embedding more than four levels of subjectivity in a given scene runs the danger of rendering that scene incomprehensible. The author may choose to emphasize that incomprehensibility to comic effect—as the creators of *Friends* did—but to render the action intelligible *while at the same time increasing and enriching its deep intersubjectivity* via the interplay of word and gesture requires some extra work.

What kind of work might it be?

We are a far way off at this point from answering that question, not least because of how little we still know about our cognitive processes. For example, it is clear to me that the functioning of our mirror neurons—when our brain makes sense of other people's gestures by "pretending" that we are making those gestures too—must play a role in enhancing our comprehension of performance. Similarly, it must matter that when we see bodies in action, we register several intersubjective exchanges almost simultaneously, while when we read about such exchanges, we process them one at a time. For example, when I teach a class, I often have the feeling that at any given moment I am aware of the body language of several students, whereas when I read descriptions of looks and gestures, I have to finish one description before I move on to another. That is, I am not sure that I can really see one student shrugging her shoulders in response to something I just said at *exactly the same time* as I see the other student leaning forward and whispering something to her neighbor, but it is obvious that it takes less time for me to register both than for you to read these two descriptions.

So far I have found one thing that Etherege does differently from the creators of *Friends* and which, I think, enables him to enhance his spectators' comprehension even as he constructs the scene featuring not just deep but apparently ever-deepening intersubjectivity. In *Friends*, the multiple embedment is built around a manifest trifle—a practical joke whose outcome has no bearing upon the characters' fate. The stakes of understanding exactly who knows what are so very low that we can afford to give it all up and merely laugh at the enthusiasm with which Phoebe negotiates and builds the mental labyrinths with her “they don't know that we know they know we know!”

By contrast, in *The Man of Mode*, the stakes of figuring out exactly who knows what are extremely high for the two characters, Bellinda and Dorimant, whose mutual readings create most of the deep intersubjectivity of the last scene. Bellinda stands to lose her reputation (if the town finds out that she is Dorimant's mistress, and now a cast-off mistress, too), her friend (Mrs. Loveit), and the man she loves. Dorimant stands to lose the woman he loves (Harriet) and his reputation as a wit. By keeping close track of the increasingly complex web of intersubjectivities, Dorimant escapes unscathed and Bellinda gets off with one serious loss instead of three. Note, too, that Mrs. Loveit, who has nothing to lose at this point—her reputation is gone and Dorimant has already left her—is excluded from this inner cycle of mutual readings. She is navigating the circumference of the “public” conversation, misreading gestures, and missing private meanings.

It seems then that the authors wishing to facilitate their audiences' comprehension of deeply intersubjective scenes should significantly raise the stakes for the protagonists and thus ratchet up our emotional involvement with their cogitations. Conversely, the authors wishing to impede their audiences' comprehension and render the deep intersubjectivity an amusing spectacle in and of itself, should lower the stakes and thus decrease our emotional involvement. Because (as Butte reminds us) “theatricality is inherently intersubjective,” a number of the late seventeenth-century playwrights must have intuitively figured out this rule. I used as my case in point Etherege's *Man of Mode*, but equally fascinating examples of deep intersubjectivity, calibrated now to enhance the viewer's comprehension now to impede it, can be found in the plays of John Dryden, for example in Act III of his *Marriage a la Mode* (1673).

I cannot claim that Austen *consciously* adapted this theatrical strategy of dealing with deep intersubjectivity to her novels. Still, it does constitute a crucial component of her construction of multilevel mental embedment. When Anne Elliot observes Elizabeth and Captain Wentworth in Molland's bakery shop, the stakes of her getting just right the private meaning of their mutually reflecting body language are very high. Austen pulls us into being profoundly emotionally invested in knowing that Anne knows that Wentworth knows that Elizabeth pretends not to recognize that he wants to be acknowledged as an acquaintance. We rise up to this cognitive challenge because we are made to believe that it *matters*.

V. Do We Really Need Cognitive Science?

Did I *have* to poach on the territory of cognitive scientists in the case of Jane Austen when our discipline seems to be doing just great on our own? With such studies as Butte's *I Know That You Know That I Know*, which brilliantly explores Austen's

innovative use of deep intersubjectivity, why borrow the vocabulary of evolutionary psychologists and talk about “Theory of Mind” and “recursive mental embedment”? The immediate, practical answer to this question is easy. First, Dunbar’s research on cognitive challenges posed by scenes that embed more than four levels of recursive subjectivity allows me to explain why is it that when *Phoebe knows that Monika knows that Chandler knows, etc.*, it is funny, but when *Bellinda knows that Dorimant knows that Bellinda knows, etc.*, or when *Anne knows that Wentworth knows that Elizabeth knows, etc.*, it is rather sad. The former takes frank pleasure in its incomprehensibility; the two latter encourage us to identify emotionally with the heroines who can decipher the complex social situations but can’t keep from losing the men they love.

Second, the studies of recursive embedment allow me to explain why Austen made sure that her deeply intersubjective passages cannot be successfully paraphrased unless we retain their original level of mental embedment. Austen knew (intuitively, of course) that readers would try to downgrade that level from the challenging fourth to the easier third or even second, and she crafted prose that would resist that simplifying impulse.

Third, I now have a new way of approaching the familiar topic of Austen’s complicated relationship with eighteenth-century sentimental novels. I can demonstrate that those novels’ treatment of multilevel mental embedment made possible Austen’s own exploration of deep intersubjectivity *at the same time* that it made her intuitively wary of the self-conscious sensibility cultivated by some earlier writers. We know that Austen disliked fictional sentimentality, but now we also can see how this dislike enabled her to free the extra level of mental embedment that she could then use elsewhere. “I know that you know that I know” comes easier for a writer who has an allergy to the (hyper-sensitive) “I know that I know that you know that I know.”

These are three immediate observations about Austen’s prose that follow from appropriating concepts of cognitive scientists who study our mind-reading adaptations. Excited as I am about them, however, I am even more interested in a larger theoretical insight that emerges from bringing together cognitive and literary studies:

It seems that fictional narratives endlessly *experiment with* rather than *automatically execute* our evolved cognitive adaptations. When cognitive scientists succeed in isolating a certain regularity of our information-processing (such as an apparent constraint on the number of levels of embedded subjectivity that we can process with ease), we can take that constraint and see how it plays itself out in a fictional narrative. What we discover is that where there is a cognitive constraint, there is a “guarantee” of sorts that writers will intuitively experiment in the direction of challenging that constraint, probing and poking it and getting around it. The exact forms of such probing and poking will depend on specific cultural circumstances, including mind-reading profiles of individual writers and their readers. The culturally enmeshed cognitive “limits” thus present us with creative openings rather than with a promise of stagnation and endless replication of the established forms. This realization marks the possibility of a genuine interaction between cognitive psychology and literary studies, with both fields having much to offer to each other.

Let me close with another question. Assuming that you agree that works of fiction constantly experiment with our cognitive constraints, we can speculate about the payoffs of experimenting with the specific constraint discussed above. Ask yourself: What

exactly is achieved when the writer does manage to pack four or five levels of recursive mental embedment into a scene yet makes this difficult embedment feel natural—that is, not funny, forced, or incomprehensible? Perhaps we can speak here about some form of subconscious pleasure experienced by readers who thus are enabled to sail smoothly through a clearly demanding cognitive construction. It is as if we are made to feel that we are dealing with a genuinely complex, nay, almost intractable, social situation, but we are navigating it beautifully. Can we then say that the scenes such as the one in *Persuasion* extend us a promise—or rather an illusion, but a highly pleasing one—that we will be all right out there in the real world, where our social survival depends on attributing states of mind and constantly negotiating among those bewildering, approximate, self-serving, partially wrong or plainly wrong attributions? Is this lovely illusion of sociocognitive well-being one reason that some writers persist in constructing such scenes and some readers seek out texts containing them?

ⁱ For a review of the new field of cognitive literary studies, see Alan Richardson, “Studies in Literature and Cognition.”

ⁱⁱ But see Gallagher, “Understanding Interpersonal Problems in Autism,” for a suggestive alternative view.

ⁱⁱⁱ For a related discussion, see also Dennett, “True Believers.”

^{iv} Zunshine, *Why We Read Fiction*, 10.

^v Dunbar, “On the Origin of the Human Mind,” 241.

^{vi} I am focusing here only on its first part because the second part—“so I pretended . . .

“—contains only one attribution of a mental state.

^{vii} See *Why We Read Fiction*.

^{viii} Compare to Vermeule’s argument about the “low” and “high” Theory of Mind traditions in the novel (“God Novels”).

^{ix} Note that when I speak of moments in literary history when the novel moves one intersubjective level up, I do not mean to imply that once it happens, novels operating on the lower level of mental embedment become extinct. For example, today Woolf’s *Mrs. Dalloway* coexists peacefully with novels that never move beyond the third level of

mental embedment. Different levels of embedment appeal to different readers at different times and also answer the demands of particular genres and styles.

^x For discussion, see Butte, 237.

^{xi} *Clarissa* figures in Butte's discussion of *Pamela* (1739). Butte sees that novel's treatment of its title protagonist as subject to "curious limitations":

[Pamela] is powerfully aware that she is the target of [Mr. B's and her parents'] interpretation, but what perceptions, experiences, and tones of feeling give rise to those interpretations in the other? How are they part of the fabric of that consciousness in her parents or Mr. B.? Neither Pamela nor *Pamela* asks these questions of their world. Mr. B. reads Pamela's letters, but she does not read *his* (in this matter *Clarissa* is a step forward). More important, Pamela does not internalize Mr. B.'s perceptions of her—she deflects them—and does not ask what consciousness those perceptions emerged from and to which her responses could be seen to return, as a thread in what could have been a tapestry of responses. (69)

^{xii} See Zunshine, "Can We Speak."

^{xiii} I am grateful to the anonymous reader from *Style* for pointing this out to me.

^{xiv} I put this term in quotation marks because I don't want to claim that this process of learning is necessarily conscious.

^{xv} See Zunshine, "Richardson's *Clarissa*."

Works Cited:

- Austen, Jane. *Love and Freindship and Other Early Works*, ed. G. K. Chesterton. New York: Frederick A. Stokes Company, 1922.
- . *Northanger Abbey and Persuasion*, ed. R. W. Chapman (London: Oxford University Press, 1969).
- Baron-Cohen, Simon. *Mindblindness: An Essay on Autism and Theory of Mind* Cambridge: The MIT Press, 1995.
- Bloom, Allan. "Introduction," in Jean-Jacques Rousseau, *Emile or On Education*. New York: Basic Books, 1979. 3-29.
- Borenstein Elhanan and Eytan Ruppin, "The Evolution of Imitation and Mirror Neurons in Adaptive Agents." *Cognitive Systems Research* 6.3 (September 2005), 229-242.
- Butte, George. *I Know That You Know That I Know: Narrating Subjects from Moll Flanders to Marnie*. Columbus: The Ohio University Press, 2004.
- Defoe, Daniel. *Moll Flanders*, ed. Albert J. Rivero. New York: Norton, 2004.
- Dennett, Daniel. *The Intentional Stance*. Cambridge: The MIT Press, 1987.
- . "True Believers: The Intentional Strategy and Why It Works." In John Haugeland, ed., *Mind Design 2: Philosophy, Psychology, Artificial Intelligence*, 2nd edition. Cambridge: The MIT Press, 1997. 57-79.
- Dunbar, Robin. "On the Origin of the Human mind," in *Evolution and the Human Mind: Modularity, Language, and Meta-Cognition*, eds. Peter Carruthers and Andrew Chamberlain. Cambridge: Cambridge University Press, 2000. 238-253
- Etherege, George. *The Man of Mode; or, Sir Fopling Flutter*, in *The Broadview Anthology of Restoration and Eighteenth-Century Drama*, ed. J. Douglas Candfield. Peterborough, Ontario: Broadview Press, 2001.
- Friends*. Episode 5.14, "The One Where Everybody Finds Out." Written by Alexa Junge, directed by Michael Lembeck. Aired 2/11/99, 5/18/99, 6/17/99, 3/18/2004.
- Gallagher, Shaun. "Understanding Interpersonal Problems in Autism: Interaction Theory as An Alternative to Theory of Mind." *Philosophy, Psychiatry, & Psychology* 11.3 (September 2004). 199-217.
- Hume, Robert D. *The Development of English Drama in the Late Seventeenth Century*. Oxford: Clarendon Press, 1976.

- Kaplan, Bruce Eric. "Of course I care about how you imagined I thought you perceived I wanted you to feel." *The New Yorker*, October 26, 1998.
- Mullan, John. *Sentiment and Sociability: The Language of Feeling in the Eighteenth Century* (Oxford: Clarendon Press, 1988).
- O'Neill, John H. "Introduction to *The Man of Mode*," in *The Broadview Anthology of Restoration and Eighteenth-Century Drama*, ed. J. Douglas Candfield. Peterborough, Ontario: Broadview Press, 2001.
- Richardson, Alan. "Studies in Literature and Cognition: A Field Map." *The Work of Fiction: Cognition, Culture, and Complexity*. Eds. Alan Richardson and Ellen Spolsky. Aldershot: Ashgate, 2004. 1-30.
- Richardson, Samuel. *Clarissa, or the History of a Young Lady*, ed. Angus Ross. London: Penguin, 1985.
- Rizzolatti, Giacomo Leonardo Fogassi, and Vittorio Gallese. "Neuropsychological Mechanisms Underlying the Understanding and Imitation of Action." *Nature Reviews. Neuroscience* 2.9 (2001), 661-670.
- Rousseau, Jean-Jacques *Emile or On Education*. Introduction, Translation, and Notes by Allan Bloom. New York: Basic Books, 1979.
- Laurence Sterne, *A Sentimental Journey*, ed. A. Alvarez. London: Penguin, 1986.
- Vermeule, Blakey. "God Novels," in *The Work of Fiction: Cognition, Culture, and Complexity*, eds. Alan Richardson and Ellen Spolsky. Aldershot: Ashgate, 2004. 147-166.
- Zunshine, Lisa. "Can We Teach the 'Deep Intersubjectivity' of Richardson's *Clarissa*?" *New Windows on a Woman's World: A Festschrift for Jocelyn Harris. Otago Studies in English*, 9. Dunedin, New Zealand: University of Otago, 2005. 88-99.
- . "Richardson's *Clarissa* and a Theory of Mind," *The Work of Fiction: Cognition, Culture, and Complexity*, eds. Alan Richardson and Ellen Spolsky (Ashgate Press, 2004), 127-146.
- . *Why We Read Fiction: Theory of Mind and the Novel* (Columbus: The Ohio University Press, 2006).