Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

David J. Ritchie, Jr. Date

The Destabilizing Affects of Writing:
Sedgwick, Derrida, and the Critique of Cognitive Literary Studies

By

David J. Ritchie, Jr.
Doctor of Philosophy

Comparative Literature

Dr. Geoffrey Bennington
Advisor

Dr. Michael Moon
Committee Member

Dr. Elizabeth A. Wilson
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

The Destabilizing Affects of Writing:
Sedgwick, Derrida, and the Critique of Cognitive Literary Studies

By

David J. Ritchie, Jr.
B.A., DePaul University, 2006

Advisor: Dr. Geoffrey Bennington, DPhil (Oxon)

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Comparative Literature
2019

Abstract

The Destabilizing Affects of Writing:
Sedgwick, Derrida, and the Critique of Cognitive Literary Studies
By David J. Ritchie, Jr.

This project proposes an evaluation of the literary reading practices and critical habits grouped under the heading of Cognitive Literary Studies (CLS) by outlining the broad assumptions necessary for their interpretive strategies to operate. Addressing the use of "Theory of Mind" by scholars of CLS, I suggest that the application of this cognitive 'mechanism' to literary reading and analysis proceeds according to a binary consolidation of cognitive relations that produces a limited interpretive range. Taking the novels of Jane Austen as a test case, these limitations are demonstrated through a close reading of Eve Sedgwick's work on Affect Theory, and her analysis of alloerotic/autoerotic pleasure in literary writing, which challenge the structures of literature outlined by cognitive readings practices. According to an alternative theory of literary narrative developed from Genette's structural analysis of *vraisemblance*, I argue that the 'cognitive' articulation of behavior is an effect of literary writing and as such is exposed to the destabilizing effects of writing outlined in Derrida's analysis of "writing in the general sense." In this light, I critically examine Jerry Fodor's apparently rigorous elaboration of "cognitive architecture" in his account of the linguistic model of representation that grounds the computational theory of mind in cognitive psychology. Turning in conclusion to Derrida's reading of *Robinson Crusoe*, I show how Derrida's and Sedgwick's work offers an interpretation of literary affect in relation to death that is simply unavailable to Cognitive Literary Studies. Read together, Derrida's theory of 'survival' and Sedgwick model of 'reading otherwise' indicate how literary affect articulates the destabilizing effects of literary writing not legible according to cognitive analysis. Derrida's elaboration of how the structure of survival governs fictional writing and "writing in the general sense" can be read alongside Sedgwick's critical reading strategies informed by the affect theory of Silvan Tomkins, suggesting that new structures of affect can be productively combined with deconstructive analysis of literary writing.

The Destabilizing Affects of Writing:
Sedgwick, Derrida, and the Critique of Cognitive Literary Studies

By

David J. Ritchie, Jr.
B.A., DePaul University, 2006

Advisor: Dr. Geoffrey Bennington, DPhil (Oxon)

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Comparative Literature
2019

ACKNOWLEDGEMENTS

This project would not have been possible without the intellectual and emotional support of many people who, each in their own way, have guided me through this impossible period of adult life known as 'graduate school.' The years of reading and writing that led up to it contain innumerable experiences of joy and sadness, of despair and surprise, of hope and delight. The professors who have helped steer me on this path began with my undergraduate advisor at DePaul University, Jim Block, and with Elizabeth Rottenberg and Michael Naas who welcomed me to audit their graduate philosophy seminars and encouraged me (endlessly) to pursue my doctoral degree in Comparative Literature at Emory University.

Among the many people at Emory without whom I would not be here are Lisa Knisley for her patient and thoughtful critiques and sharp eye, Ania Kowalik for her kindness and shared delight in live music, and Asher Haig for his inexhaustible appetite for discussing theory of all sorts.

My advisor, Geoffrey Bennington, has stuck with me through so many unexpected turns of circumstance that it defies simple description and without his help I would not be where I am today. It was Michael Moon who led me to the work of Eve Kosofsky Sedgwick, and who pushed me to 'read otherwise' in his courses, and who pushed me to write something *interesting*. To Elizabeth Wilson I owe much of my intellectual development, I cannot adequately express my gratitude for her energy and rigor, and for her willingness to take on a student from another department (who wouldn't have gone away). Last, Alian Teach, who has been an indispensable support and delight; we are all lucky to have her and she holds us all together in more ways than we know.

The love and support of my partner, Will, is—and continues to be—without measure. Through the long nights, long months, and long years, he has always created for me that precious resource required by all writers for their success, namely, the time and space to think and read—and write. We continue our adventure together into new and yet unknown territories.

This entire adventure would not have been conceivable without my dad, and the love of reading and learning he instilled in me as a child. It is to him that I owe the stubborn determination to complete my degree. He has seen me through, from the beginning so long ago, to the end which I now share with him. Finally, this project is made possible, in more ways than can be described, by my mom, Ginny, who saw me off to graduate school, but did not see me achieve what she always believed I was capable of.

Contents

Introduction

Reading Jonathan Kramnick's essay "Empiricism, Cognitive Science, and the Novel" (2007) alongside his later polemic, "Against Literary Darwinism" (2011), offers some much needed clarity for the project of evaluating various approaches to literary criticism assembled under the disciplinary heading of Cognitive Literary Studies.[1] Kramnick's work on this particular trend in literary criticism has proven quite valuable to me because of his willingness to engage substantively with the historical developments and debates within cognitive science. The result is that his work consistently provides a model for how to coordinate the claims offered by Cognitive Literary Theorists on fictional narratives with the scientific scholarship they draw upon for support.

In "Empiricism, Cognitive Science, and the Novel" Kramnick's model for gaining access to the anxieties surrounding the use of cognitive science as a means for literary criticism begins by stating outright exactly what is as risk. It is "one of anachronism or universalism, either shoehorning recalcitrant descriptions of the mind into our current language of cognition or locating both within a timeless and unchanging account of the psyche" (Kramnick 2007, 263). Manifestly, then, Kramnick's purpose in his essay is to find a way to deal with the chronological separation between current work in cognitive science used by critics to "talk about the way in which thinking takes shape in particular works from [eighteenth-century fiction]" that does not resolve into anachronism or universalism (263). Conveniently, Jerry Fodor's book *Hume Variations*, a monograph on the work of empiricist philosopher David Hume, provides Kramnick with just such a basis on which to proceed. While Fodor's book attempts to annex Hume's *A Treatise on Human Nature* as "the foundational document of cognitive science" Kramnick's interest in Hume

[1] The definition of the term 'cognitive' in Cognitive Literary Studies, as well as the role it has in determining the methodological choices or critical reading practices developed under its moniker, varies considerably. Lisa Zunshine, in her Introduction to *The Oxford Handbook of Cognitive Literary Studies*, refers to Alan Richardson's definition, "who starts by pointing out that 'cognitive' is a broad term, referring to 'an overriding interest in the active (and largely unconscious) mental processing that makes behavior understandable'" (Zunshine 1). In the analysis that follows, I have made every effort to follow the author's use of the term.

derives from the fact that empiricist and computational models of cognition both have in common a "representational theory" of mental architecture (Kramnick 267). In this overlapping zone, both empiricist accounts of the mind and the computational model of cognition are in agreement that "the mind works by forming representations of objects and events and then implementing them in various processes of thought" (264). Having established this common ground between the two fields, specifying how these theories of cognition depart and why, is essential to track the aspects of cognition are being reformulated.

These differences begin almost immediately with the very nature of the representations, as empiricism wants to maintain that these representations are produced from one's contact with the external world. To illustrate this Kramnick quotes from Locke's *Essay Concerning Human Understanding* and the description of the mind he outlines there which presents the mind's representational structure on the basis of an image: "Closet wholly shut from light, with only some little openings left, to let in external visible Resemblances, or *Ideas* of things without" (265). For the empiricist models of thought, ideas have the representational structure of images in order to guarantee a 'transparent' relation to the external world and to ensure that the epistemology produced as a result is also derived from the external world.[2] What is important about the empiricist project is, as Kramnick has it, "that it attempted to cobble together an account of thinking with an account of epistemology" (Kramnick 2007, 256).

Diverging from this graphic, image-like structure of mental representation, the computational model of cognition maintains that ideas are structured linguistically like of a mental language. The reason for this opens onto a broader discussion of how the mind operates, whether it functions as "a screening room for pictures" or as "an instrument for processing concepts" (266). What is at stake in the difference between

---

[2] The theoretical consequences for Hume's psychology that are generated by Hume's epistemological commitment to empiricism are well known to Fodor. Here is a characteristic passage that we will return to later in chapter three: "Hume's semantic empiricism doesn't allow mental representations of anything except what can be given in a specious present, namely, the content of an experience at a time. This is, in short, yet another of those cases where Hume's epistemology prohibits the very theory that his psychology demands.

Bother epistemology, as I think I may already have pointed out, and bother empiricist epistemology most of all" (Fodor 2003, 133).

representations structured like an image or a unit of mental language entails the larger question of the concept of language at work in Fodor's thinking. Defining representations as units of a mental language enable Fodor to account of the generation of complex ideas from simple ideas. This happens as the result of a twofold change. First, by conceiving of the mind as analogous to a computer, the syntactical dimension of thought is accorded more significance than is allowed by the empiricist models of cognition in which simple thoughts (themselves impressions of the external world) are linked together to produce complex ones. Second (or simultaneously) Fodor's understanding of this syntactic "mental language" entails that the units themselves differentiated according to their structure, so that "some parts of the thought have priority over others in determining its overall meaning" (Kramnick 2007, 266).

A further but related difference that operative concept of language introduces is seen Fodor's explicit comments on the differences between the two models of cognition, Lockean and computational, with specific regard to the role of "associationism" in Hume's model. In an article published in 2006 in the journal *Daedalus,* Fodor offers this description of the difference: computational "mental representations are sentence-like rather than picture-like. This stands in sharp contrast to the traditional view in which Ideas are some kind of images. In sentences, there's a distinction between *mere* parts and constituents, of which the latter are the *semantically interpretable* parts. By contrast, every part of a picture has an interpretation: it shows part of what the picture shows" (Fodor 2006, 88).

Sticking with the comparison to the empiricist model of cognition for the moment, we can retrieve two points from Fodor's comments. First, it is clear that in this account 'images' enjoy a somewhat strange semantic homogeneity. Second, Fodor's division and weighting of "mental representations" by way of categories borrowed from linguistics marks something more of a methodological shift from the empiricism of Hume. That is, Fodor's choice of a representational model for mental representations is dictated by a logical structure embedded in the referent: "In sentences, there's a distinction between *mere* parts and constituents" that enables a specific operation of interpretation. Mental representations are no longer "picture-like" resemblances as they were for Locke and Hume, neutral, photographic copies of the things

themselves, and it is for this reason that Fodor differentiates the two models. Without difference the "picture-like" representation is explicitly without interpretable content. Mental representations do not *resemble* sentences, rather they *operate* like sentences.

Backing up slightly in Fodor's article, I'd like to draw attention to a similarity that persists between these cognitive models despite the differences we've already noted. Locke and Hume's initial theories about the structure of Ideas was derived from Associationism as the principle by which simple Ideas were joined together in the mind to form more complex thoughts. Here's Fodor:

> For nearly three hundred years, associationism was the consensus theory of cognition among Anglophone philosophers and psychologists.…Associationism was widely believed to hold, not just for thought but for language and brain processes as well: thoughts are chains of associated concepts, sentences are chains of associated words, and brain processes are chains of neuron firings. In all three cases, transitions from one link in such a chain to the next were supposed to be probabilistic, with past experience determining the probabilities according to whatever Laws of Association happened to be in fashion. (Fodor 2006, 87)

Despite the distance Fodor is at pains to put between Hume and himself, his own model of cognition also relies on an analogy to the operation of language, in the grammatical laws of its syntax. Alluding quickly to Kant's critique of Hume, Fodor continues: "Thoughts aren't *mere* sequences of ideas; at a minimum, they are *structured* sequences of ideas.…Likewise, sentences aren't just lists of words. Instead, they have a kind of internal structure such that some of their parts are grouped together in ways that others of their parts are not." (Fodor 2006, 88). What we have is a rather slippery relation between language and thought. The question becomes, how does Fodor's use of language as a model for thought differ from Hume's?

Leaving this question suspended for the moment, I want to return to Kramnick and retrace some of the larger points of interest as they connect with this discussion of cognition. Acknowledging that current

research in cognitive science understandably differs from aspects of the empiricist model of cognition, Kramnick points out that it is remains useful as rubric for the recognizing the vocabulary by which other questions circulating in seventeenth- and eighteenth-century culture appear in literary works. Decoupling historical questions from the questions of epistemological accuracy or truth allows the question to be reformulated in the following way: "If the mental token in the representational architecture of empiricism turns out on comparison to be an image of an object and thus to be object*like*, what for example might this tell us about the related model of agency, or as we now say, of mental causation?" (Kramnick 2007, 268) By shifting the focus on to the question of agency, the literary articulation become less burdened by the questions at issue for a cognitive psychology. That is, regardless of whether Hume and Locke are correct to model the cognitive processes as image-like, their models have analytic value because established the terms that informed discussions of human agency and individualism.

For empiricism as for present day cognitive psychology, mental states are understood as the precursor of an individual's behavior and thus to participate in the operation of a causative chain linked to observable behavior.[3] Kramnick distills the empiricist position with reference to Locke, who "argued that what motivates a person to take this or that action is uneasiness in the want of some absent thing. One acts to relieve this uneasiness by striving toward an object of desire or by ridding oneself of an object of distaste…And so agency on this view is an output of a person's relation to a mental token that is picture-like" (Kramnick 2007, 269). What's starting to take shape here is the way that empiricist models of agency and cognition are enfolded within a broader meditation on subjectivity, particularly an understanding of subjectivity defined in relation to object-like images. Kramnick brings these points to a close by recalling the chapter on personal identity in the *Essay on Human Understanding* in which Locke ties consciousness together with mental representations and personal experience. "To be a person," Kramnick writes, "is to have a series of connected experiences during which time one was aware of the representational nature of

[3] The question of causation, and of how to understand Hume's account of it will occupy much of Fodor's attention, see particularly Fodor 2003, 60-80 *passim*.

one's thoughts; it is to believe that the same person was in view of representations in the past as in the present, and that one ought to care about the person's fate in the future" (Kramnick 2007, 269) Individuals' behavior is thus linked to mental representations as they are "what motivates a person to take this or that action," and the actions taken, along with an individual's thoughts, compose a plot-like narrative, a temporal sequence linking a person's "past" and "present" with their "fate in the future" (Kramnick 2007, 269).

But why is the question of agency so important for literary study; why is it that Kramnick is pursuing Locke's meditation on it as an avenue for responding to Fodor's disqualification of the "picture-like" resemblance model of mental representation? Remember that Kramnick is questioning the premise Fodor has denied, that is, that "picture-like" representations have the structural requirements necessary for computational operations of cognition. Kramnick elaborates this by recourse to an epistolary novel by one of Locke's contemporary writers, Catherine Trotter, titled *The Adventures of a Young Lady*. Trotter is significant for the additional fact that she defended Locke's philosophical ideas in pamphlets of her own and, crucially, elaborated them in a way that further integrates Locke's cognitive model with the conception of agency. The question she posed specifically targets the interaction between these two facets: "I am thinking of a horse; his beauty strength and usefulness. Does this thought preserve the Idea of a Church, of Happiness or Misery? If they remain in the mind when I was only thinking of a horse where they are bestowed, it may be presum'd, there is room for that one idea more without thrusting out another to give it place" (Quoted in Kramnick, 270). Trotter's point is subtle, yet extremely important because it reintroduces the individual's agential capacity to manipulate ideas independently from the experience of the (real) objects of which they are impressions. Remaining still within the perimeter of the empiricist model, Trotter shows that the *relation* of an individual to an Idea can also become an object. The payoff being the possibility of introspection, as one can now recall previously held attitudes *to* ideas, that is, as ideas themselves (e.g. "I am thinking of a horse; his beauty strength and usefulness") independent of the corresponding objects. The ability to recall the image-idea of the horse *as* beautiful, strong, or useful is to

recall "a person's *relation* to a mental token that is picture-like" (269).

Trotter puts this insight to work in a scene from her novel in which the main character, Olinda, reflects on her ill-fated courtship with Cloridon, in the text of a letter. The following sentence from the passage cuts directly to the core of Kramnick's overall argument, that is, that historically contemporary theories of cognition are not rendered valueless in the face of modern cognitive science's claims to truth. Here are a few lines from Olinda's letter:

> How comes it then, that I am so Griev'd and Angry that he loves another? And that I wish with such impatience for his Return? In fine, I discover'd that what I had call'd Esteem and Gratitude was Love; and I was as much asham'd of the Discovery, as if it had been known to all the World. I fancy'd every one that saw me, Read it in my Eyes: And I hated my self, when Jealousies would give me leave to Reason, for my extravagant thoughts and wishes. (Quoted in Kramnick 2007, 271)

Kramnick guides us to understand the importance this passage has for our larger discussion of object-image model cognitive representation. He writes, "Olinda's ideas become so clear to her as pictures that she imagines others must be able to see them as well.…The multiple perspective slows down the train of Olinda's ideas so that each may be separately examined as a discrete image; this is how she discovers that esteem and gratitude are really love" (Kramnick 2007, 271) Trotter here shows Olinda manipulating images of her own relation to objects. The interpretations she performs as a result seriously complicate the claims made by Fodor, were we to adhere to the implicit 'backwards compatibility' premise claimed by the computational model of cognition. Fodor's critique of the empiricist model of cognition was premised on the fact that what we now know about cognition disproves the model generated by the empiricists, because the empiricist model does not allow for the units of thought to be themselves semantically interpretable. Trotter has demonstrated that it is possible for these image-like units *to be* semantically interpretable. Fodor's position is that the present-day computation model of cognition is the truth of cognition, and thus

supersedes or underlies other misguided cognitive models. In other words, what Kramnick has done here is to show that a 'cognitive' reading of eighteenth century fiction is possible on its own terms and no less powerful for it, it is through the juxtaposition of the empirical model with the computational model of cognition we can identify two competing claims to truth at work in these projects.

Both projects are underwritten by truth claims, but these claims are differently located and differently mobilized. Fodor maintains that there is a tension between "what semantic productivity requires and what empiricism permits; the former wants the structure of a representation to 'add something' to the content of its constituents, but the latter wants it not to" (Kramnick 268). Fodor's psychology claims to more accurately model cognition because a syntactic mental architecture that operates on "units of a mental language" can account for complex ideas, while the empiricist's associationist mental architecture cannot.[4] Fodor's critique intervenes on the basis of "what semantic productivity requires," presenting the element of cognition as *observable* but unaccounted for within the empiricist model. Yet the core motivation of the empiricist's project is to account for the creation of mental representations from individual experience. On the one hand, individual experience is taken as the point of departure from which to generate a model of cognition; on the other hand, an element of individual experience is used to question the validity of a cognitive model based on experience.

The impasse reached at this point is representative of the wider set of problems to which my project responds. Attending specifically to the attitude that animates cognitive science writ large, for which Fodor is the synecdoche here, it is clearly evident that the judgment of the empiricist's cognitive model is one of inaccuracy that present day cognitive science has now corrected and set right. The argument here is

---

[4] This claim will be returned in the third chapter. Here is a representative summation of Fodor's position: "It's because association is semantically transparent that Hume can rely on the content of complex representations not to outrun the content of their constituents. If conceptual complexity reduces to association, and if the content of all the simple concepts is empiricistic, then so too is the content of all the complex concepts. So associationism gives Hume a notion of compositionality that in turn gives him the empiricism that he wants. Association explains the complexity of concepts, and the complexity of concepts explains their semantic productivity. That all seems admirably tidy; but on reflection it won't do. For (dilemma) *the consequence of association being semantically transparent is that it isn't semantically productive*. Something has gone wrong" (Fodor 2003, 91)

structured almost archeologically; the more the techniques of cognitive science steadily brush away the sand concealing our mental architecture, the more it emerges whole and intact as an ahistorical, unchanging processes. Scholars working in cognitive literary studies implicitly or explicitly accept this premise and it fuels the confidence of their critical projects.

In Lisa Zunshine's book, *Why We Read Fiction: Theory of Mind and the Novel*, she presents novels as the site of variability against a stable and universal cognitive architecture that processes them. The multiple and various generic forms of the novel throughout its historical development are collectively understood as fundamentally structured by an implicitly unified mental architecture of cognition. According to this account, the novel is defined primarily (which will come to include structurally, aesthetically, formally)—indeed enthusiastically and deliberately defined—as operating by prompting us to "posit a mind whenever we observe behavior" of literary characters, "as [novels] experiment with the amount and kind of interpretation of the characters' mental states that they themselves supply and that they expect us to supply" (Zunshine 2006, 22). Against what is, perhaps, the broadest and most basic, grounding assumption of Cognitive Literary Studies, the guiding principle and that cognitive psychology and science *should*, varying according to the theoretical requirements of the particular argument or the personal commitment of a theorist, inform the scope of literary criticism and theory. This is the price of entry, and the power of the theoretical claims made by cognitive literary studies derives directly privileged position given to the cognitive, as determining the fundamental structure of the novel. Novels, that, per Zunshine, prompt us "posit a mind whenever we observe behavior" in a literary narrative *are*, so the argument goes, explicitly structured by cognition and *are* governed by cognitive processes according to a hierarchical organizational structure.

In addition to this basic premise of Cognitive Literary Studies, the specific components of cognitive theory under examination in the coming chapters are organized around the central premise of a connection, determining and determinate, between cognitive structures such as Theory of Mind (a "mechanism" in the technical conceptual vocabulary of cognitive science; a "module" in the slightly more

figural theoretical vocabulary of cognitive psychology) and cognitive processes such as the computation, and the literary novels and to linguistic grammar. I take up Zunshine's cognitive readings as my focus deliberately because of the propositional, motivational force she assigns to cognition, as indicated in the title of her most widely circulated text, *Why We Read Fiction: Theory of Mind and the Novel*. Blasting forth from the very first word is the claim or promise that Theory of Mind, we are to understand, answers a standing question: Why do we read fiction? The preoccupation of the novel with the representation of 'other minds' is understood in cognitive terminology as a "representationally hungry problem" for our cognitive architecture, thus allowing Zunshine to claim that novels are "grist for the mill of our mind-reading capacities" (Zunshine 2006, 16).

Active in Zunshine's work, but equally so in the work of other literary critics who explicitly bolster their arguments with cognitive psychology claim to truth, is a secondary, supplementary claim that the cognitive psychology of the twentieth century (itself already supported by robust program of cognitive science) is "now" endorsed by evolutionary theory.[5] For instance, Blakey Vermeule, author of *Why Do We Care About Literary Characters?*, responds to Kramnick's critique in "Against Literary Darwinism", with a double gesture announcing "the fact that evolution is undisputedly the scientific ground" of psychology while sequestering the "contentious" "narrow-gauge discipline" of evolutionary psychology away: "Like Gould, Lewontin, and Fodor, [Kramnick] seems utterly unmoved by the fact that evolution is undisputedly

[5] See, for example, the description of unifying cognitive literary theory given by Ellen Spolsky in the preface to *The Work of Fiction: Cognition, Culture, and Complexity*: "Cognitive literary theory, these essays argue, is thus well positioned to provide insight into a question that has been occluded by the well-deserved successes of the reemergent historical and multifaceted cultural studies that proliferated after the New Criticism in the twentieth century. That question is: how does *the evolved architecture that grounds human cognitive processing*, especially as *it manifests itself* in the universality of storytelling and the production of visual art, interact with the apparently open-ended set of cultural and historical contexts in which humans find themselves, so as to produce the variety of social construction that are historically distinctive, yet also often translatable across boundaries of time and place?" (my emphasis, VIII). I draw your attention to the slippery implicit/explicit, connoted/denoted, determining effect assigned to evolution, in being both the origin and ground of cognition-as-processing in addition to its influence (however ill defined here) over "storytelling." The use of the word "storytelling" here as a metonymy for 'literature' is indicative of how literary works are redefined in anthropological terms more easily compatible with evolutionary concepts such as natural selection and adaptation. In "storytelling" is understood to signify an evolved adaptation that enabled humans to share information, which also happens to serve as the basis of aesthetic forms of literature.

the scientific ground on which academic psychology proceeds. In a broad sense, all psychology is now evolutionary psychology (although [evolutionary psychology], the narrow-gauge discipline, remains as contentious as ever)" (Vermeule 2012, 427). Vermeule's position is slightly different from Zunshine's insofar as Vermeule's objection remains located within the domain of psychology rather that specifically affirming the application of evolutionary psychology to literary interpretation; however, these two positions coincide in their reliance on the truth claim made by her identification of the 'now' for "academic psychology." Explicitly animating Vermeule's comment is the belief that the difference between the account of cognition given by cognitive psychologists and the account given by evolutionary psychologists is negligible. Vermeule's reservation is that "academic psychology" has yet to accept the laws of evolution, understood primarily in terms of the pressure exerted by natural selection that results in a consolidation of cognitive models as adaptations selected for evolutionary fitness.

These motifs at work in Cognitive Literary Studies inform the trajectory of my own project, and as we will see in more detail in the coming chapters, the methodological question of how Silvan Tomkins's theory of affect can be a productive resource for literary analysis is addressed directly by Sedgwick. My contribution is to bring the literary analysis Sedgwick's develops based on Tomkins's affect theory into conversation with Jacques Derrida's deconstruction (which, it should be noted, Sedgwick herself was no stranger to), around the structure of writing, of "writing in the general sense," but specifically around the questions of literary fiction and the "destabilizing effects of writing" (Derrida 2017, 62-3, 72). To readers familiar with Derrida's work (and even for those readers who are not), the centrality of these problems to any number of his readings of philosophical texts will be easily recognized, either with fondness or with frustration, as persistent questions that motivate Derrida's inquiry and shape his analyses.

Critical frustration with deconstruction, that is, that quickly or overtime congeals into a firmer distaste, and a ultimately a dismissal that regularly takes a familiar form (familiar, but by no means exhaustive) of a position that Derrida's analyses are 'too literary,' that Derrida's critiques are not themselves held to the

same critique, or that they are 'destructive' and do not formulate a positive alternative.[6] Less frustrated critical responses, ones that do not resolve into outright dismissal but instead acknowledge a legitimate basis of his analysis often do so only after preemptively isolating the philosophical questions about writing which Derrida distills from literary texts as limited only to works of high modernism such as those by Stephan Mallarmé, James Joyce, and Franz Kafka. All of these writers, so the argument goes, represent an overt *authorial intention* to manipulate and multiply textual effects in order to stage philosophical problems, but these problems just are not present in conventional literary works, works such as, for example, novels like Jane Austen's *Sense and Sensibility* and Daniel Defoe's *Robinson Crusoe*.

Importantly, these works of literature also prove most attractive to cognitive literary studies and thus provide a rich site of overlap in which to evaluate how cognitive reading strategies operate, and how affective reading strategies operate differently and thus yield different results. In the first chapter, I outline how the critical impulses that motivate Sedgwick's earlier work in *Between Men* and *Epistemology of the Closet* connect with the reading strategies developed from Tomkins's theory of affect in "Reading Silvan Tomkins: Shame and Cybernetic Fold," the introductory essay to *Shame and Its Sisters: A Silvan Tomkins Reader*, a volume of sections from Tomkins's larger work, *Affect, Imagery, Consciousness*. In "Affect Theory and Theory of Mind," what is so remarkable and refreshing about her critical method is evident in the opening paragraphs, reaching back to the axioms delineated at the opening of *Epistemology*. Unsatisfied with the coarse axes of difference that have been installed as fixtures in our shared critical tool belt (sex, gender, race, class) she finds that Tomkins's theory of affect "pulse[s] with possibility" not because of its claim to the truth of human affect system, but because it provides a conceptual vocabulary with which to address the complex field of relations central to literary activity of reading (Sedgwick 2011, 145). In the second chapter, I juxtapose cognitive readings of Jane Austen with the analysis of *Sense and Sensibility* developed in Sedgwick's essay, "Jane Austen and the Masturbating Girl," to indicate how the question of

---

[6] This issue frames Derrick Attridge's interview with Derrida in "The Strange Institution Called Literature," *passim*.

alloerotic/autoerotic pleasure demonstrates that Austen's narrative remains vulnerable to the destabilizing effects of writing (in the specifically Derridian articulation of the phrase)—effects that remain inaccessible to cognitive literary analyses.

In the third chapter, I step back from the close scrutiny of Cognitive Literary Studies, and review Derrida's essay "Signature, Event, Context" to succinctly articulate the structure he outlines there between writing and death, the "testamentary structure of writing." Then, turning to Gérard Genette's essay "'*Vraisemblance*' and Motivation," I offer an alternative, formal theory of the novel and narrative that directly addresses some of the central concerns motivating cognitive literary studies. Namely, that the novel's historical and aesthetic developments demonstrate a strong connection to questions surrounding characters' motivation and narrative behavior, that is, to presumptively cognitive questions. From Genette, I proceed to a sustained analysis of the "computational theory of mind" as outlined by Jerry Fodor in his monograph on Hume's cognitive psychology to show how the linguistic model of representation deployed under the banner of 'computation' remains dependent on intention, and thus does not resolve or address the questions about writing drawn from "Signature, Event, Context."

In the second half of chapter three, I turn to the reading of *Robinson Crusoe* Derrida gives during the second year of his *The Best and the Sovereign* seminar, to show how the structure of writing, of its primary relation to death can be understood in literary terms of narrative action. Derrida's delightfully provocative assertion that Robinson did die mobilizes an array of questions, both about the text and about text's conditions of possibility, that returns us to Sedgwick. What emerges from Derrida's reading of Robinson Crusoe is the suggestion that literary affect, that is, Robinson's very specific fear of "dying an living death" by being "swallow'd up or Buried alive," holds a strong connection to the "destabilizing effects of writing" and to the phantasm of sur-vival that Derrida articulates throughout The Beast and the Sovereign. In closing, I argue that Sedgwick's reading of Proust in "The Weather in Proust" outlines how Tomkins's theory of affect can be brought to bear literary works and provides the conceptual resources by which to 'read-otherwise." What she offers is way to understand Derrida's reading of Robinson's fear of "dying a

living death" as the more complex relationality he draws from it, not only as Robinson's affect of fear terror affect within the narrative, but as harboring the question of its own conditions of possibility.

What Kramnick has so effectively demonstrated about the use of cognitive science as a heuristic for literary criticism, has proceeded through a technical examination of the mechanics of cognitive science. His reservations about the use of cognitive science in literary criticism mark a sober awareness of the complications that accompany interdisciplinary projects. What is important to recognize, however, is that he remains cautionary and not prohibitory. In concluding his essay, he reflects briefly on where and how he sees these lines to be drawn:

> Those of us who work in literary study are, needless to say, not in a position to judge the relative merits of simulation theory and theory-theory as accounts of the mind itself. We ply our trade at the level of heuristics, putting modes of analysis together and seeing what emerges in the process. In this way, the appeal as well as the disadvantage of the simulation approach consist in the relation it illustrates between fictional and the philosophical versions of the mind and between the eighteenth-century materials and contemporary theory. The questions raised by this relation are, accordingly, under what sort of social, technological, and cultural pressures did the period come up with a model in which introspective mind reading became both possible and urgent, and according to what formal devices did writers evoke and render palpable a process understood to be mental and imperceptible? (Kramnick 2007, 278)

Kramnick's position is synoptic. The questions he raises are fundamental to a successful application of cognitive science to the object of literary criticism, the novel. Part historical, part formal; what he has allowed us the breathing room to do is to suspend momentarily the ahistorical truths of cognitive science and to examine literary texts as artifacts that operate according to other principles.

"Each book is a pedagogy aimed at forming its
reader."
— Jacques Derrida, *Learning to Live Finally*, p. 31

Chapter 1

Queer Theory and Theory of Mind

Proceeding directly from the central proposition of Eve Kosofsky Sedgwick's text, *The Epistemology of the Closet*, that "many of the major nodes of thought and knowledge in twentieth-century Western culture are structured—indeed, fractured—by a chronic, now endemic crisis of homo/heterosexual definition, indicatively male, dating from the end of the nineteenth century," this chapter will outline a connection between the analytic strategies used to read the effects of the epistemological crisis of sexual definition amplified by post-Stonewall sexual politics and the literary critical reading practices enacted in the essays "Reading Silvan Tomkins: Shame and the Cybernetic Fold," "Affect Theory and Theory of Mind," and "The Weather in Proust" (1).[7] Sedgwick's identification of the problem of sexual definition as a 'crisis' might initially read as a overdramatic, rhetorical flourish, yet the preface to the 2008 edition of *Epistemology* confirms the durability of her terminological choice. Reminding her readers that "in the latter 1980s, when *Epistemology of the Closet* was written, something called 'sodomy' was illegal in half of the United States, and the precise definition of that slippery by highly charged term was a front-page issue in the national debate around gay rights," Sedgwick uses the preface to emphasize that the epistemological crisis of homo/heterosexual categorization, less immediate and palpable, "is not a natural given" dictated by "immemorial, seemingly fixed discourses of sexuality" but rather "a historical process, still incomplete today and ultimately impossible but characterized by potent contradictions and explosive effects" (Sedgwick 2008, xiii, xiv, xvi).

---

[7] Sedgwick parses her use of the concept of "fractures" in *Between Men*, informing us there that "by fractures [she] mean[s] the lines along which quantitative differentials of power may in a given society be read as qualitative differentials with some other name" (Sedgwick 1985, 10).

Yet the analytic figure of 'the closet' and the epistemological crisis of sexual definition in which it is rooted is by no means an abstract, academic debate, but issues forth from the indescribable horrors of the height of the AIDS crisis in American in the late 1980s. It would be a dishonor to the lives lost to allow this fact to fade from present view, and thus I quote Sedgwick at length:

> I think anyone who was around gay communities in the eighties would agree that, far beyond the changes in legal interpretation [of *Lawrence v. Texas*], the difference that's harder for younger people now to fully imagine has to do with the ten-enveloping pressure of the AIDS emergency. The first reports of the disease had come out only in 1981, and its sheer newness, its untreatableness, and its ballooning mortality brought a sudden, encompassing devastation into the lives of urban gay men and their friends. It was common experience then to be in the room of vibrant young people, conscious that within a year or two, all but a few of them would have sickened and died. Furthermore, the excessively potent fusion of homophobic stigma with deadly medical mystery resulted in uncanny fractures within families as well as society at large. If gay communities were experience an unremitting horror comparable to that of wartime, it seemed to be a war full of disowned losses without a home front, generating grievous news that no one was willing to receive. (xiv-xv)[8]

Signaled above in the reference to *Lawrence v Texas*, the 2003 decision which overturned the legal precedent set by *Bowers v Hardwick* in 1986 upholding a Georgia law that criminalized oral and anal sex, Sedgwick's critical interest in the figure of the closet is "impelled by the distinctively indicative relation of homosexuality to wider mappings of secrecy and disclosure, and of the private and the public, that were

[8] In addition to the reasons already stated, it is not insignificant to observe here that the epistemological crisis of sexual definition that drives the analyses in *Epistemology* is shaped by the crucible of a set of rapidly shifting relations to death. The relation of literary writing to death will play an important role in later chapters.

and are critically problematical for the gender, sexual, and economic structures of the heterosexist culture at large" (Sedgwick 2008, 70). This chapter will unfold the implications this generalized epistemological crisis holds for the complexities of literary analysis that result when leveraging scientific knowledges as mechanisms for literary interpretation. My choice to pair my critique of Cognitive Literary Studies with an exploration of Sedgwickian models of critical reading practices is guided by Sedgwick's patient, persistent pursuit of seemingly intractable problems. Almost Bartlebian, the sense of calm fascination shaping her engagements with conceptual problems which are widely considered to be 'dead-ends' has demonstrated the immense value of intellectual projects of "thinking otherwise,"[9] against or beside dominant critical ideologies. Confronting a critical enterprise such as Cognitive Literary Studies, whose glacial advance proceeds undisturbed by the terrain crushed smooth under the weight of its scientism, requires exactly the type of earnest tenacity that circulates throughout Sedgwick's thought.

The broad outlines of this task anticipate the work I undertake in the following chapters, wherein my focus will be narrowed to specific applications of cognitive reading practices to the literary texts. This chapter will model how to create some analytic space between the explanatory discourses of cognitive science and psychology and the interpretative questions regarding literary narrative and character. The methodology Sedgwick elaborates in *Epistemology of the Closet* outlines how "the figures of 'the closet' and 'coming out'" each mobilize a related series of diacritically opposed, "epistemologically charged pairings" that organize and structure political and cultural relations, such as "secrecy/disclosure, and public/private"

---

[9] cf. "Foucault's searching critical analysis of the persistence of the repressive hypothesis through so many, supposedly radical and discontinuous discourses—Marxist, psychoanalytic, and libertarian, as well as liberal—certainly indicates that the project of thinking otherwise remained a prime motivation of his study. …But the triumphally charismatic rhetorical force of *Volume 1* also suggests that Foucault convinced himself—certainly he has convinced many readers—that that analysis itself represented an exemplary instance of working outside of the repressive hypothesis. Rather than working outside of it, however, *Volume 1*, like much of Foucault's earlier work, might better be described as propagating the repressive hypothesis ever more broadly by means of displacement, multiplication, and hypostatization….My discontent with the interpretations listed above [which claim, but ultimately fail, to fulfill Foucault's implicit promise that there might be ways of stepping outside the repressive hypothesis] is not, either, that they are too pessimistic or insufficiently utopian. Instead, impressed by Foucualt's demonstration of the relentlessly self-propagating, adaptive structure of the repressive hypothesis, I came to see a cognitive danger in these interpretations: a moralistic tautology that became increasingly incapable of recognizing itself as such." *Touching Feeling*, 11-12

but also "masculine/feminine, majority/minority, innocence/initiation, natural/artificial, new/old, growth/decadence, urbane/provincial, health/illness, same/different, cognition/paranoia, art/kitsch, sincerity/sentimentality, and voluntarity/addiction" (Sedgwick 2008, 72). Embedded in this series of oppositional pairings are axes of conceptual definition that will play a central role in affect based analyses formulated later, and demonstrate how the antihomophobic and gay-affirmative analytic techniques generated here can inform my critique of Cognitive Literary Studies.

By attending to these locations in Sedgwick's corpus I argue that the axioms she elaborates in *Epistemology of the Closet*, the set of "otherwise implicit methodological, definitional, and axiomatic groundings" shaped by "the long crisis of modern sexual definition" model a strategy of reading that can be productively applied to other, similarly structured epistemological crises (Sedgwick 2003, 22, 1). If the epistemologies and discursive resources mobilized around this definitional task are also complicit in the social and political power structures of oppression, it will be necessary to formulate different strategies of inquiry and interrogative tactics to think otherwise. It is this nexus at which structures of oppression are soldered to epistemological supports that serves as the guiding paradigm for Sedgwick's investigation of literary texts. After locating deployments of this paradigm within the modern literary field, her examination begins to map out "how the definitional stranglehold [of homo/heterosexual definition] works, and for whom" but also "where the points of volatility or leverage in it might be, and, again, for whom" (Sedgwick 2003, 92). Transposed into the register of closed and open systems that will be introduced later, I am arguing that Sedgwick's strategy here reads these points of stricture as an open system that has been restricted to operating as a closed system.

The impaction and deadlock that characterizes the relation of Cognitive Literary Studies to other modes of critical analysis in the Humanities indicates a deeply situated divide regarding fundamental questions posed by the literary text. The stakes of this divide are easily visible in the responses solicited by *Critical Inquiry* to Kramnick's critique of cognitive critical reading practices, entitled "Against Literary Darwinism," in which he lays out "the central premises of the Darwinian program in literary studies" and

offers a comprehensive inventory of the scientific and psychological concepts they utilize, as well as a rigorous critique of the conclusions advanced by its proponents (Kramnick 2011, 316). While recognizing that the field of Cognitive Literary Studies cannot justifiably be reduced to the claims of Literary Darwinists, the responses to Kramnick's essay offer us a unique view of the more widely shared basis from which work in the Cognitive Literary Studies proceeds. Introduced above, the response offered by Blakey Vermeule of Stanford University, author of texts that utilize concepts from evolutionary cognitive psychology for literary analysis, provides a picture of which propositions are mostly strongly held and defended. Initially soft-pedaling the stakes of their disagreement over the accuracy of evolutionary psychology's account of the human mind, Vermeule claims to join Kramnick in harboring a healthy skepticism that evolutionary psychology can parlay its claims of cognitive architecture far enough to offer "a compelling account of particular products of the human imagination" (Vermeule 2012, 426). Yet quickly bypassing this shared skepticism, Vermeule gives the following caricature of Kramnick's critique in which we can discern a certain interpretive fatigue. Legible in what Vermeule finds appealing about evolutionary psychology is the streamlined interpretive mechanism the theory offers for literary analysis, secured by defining them as "products of the human imagination." Yet in acknowledging her own questions, she offers little in response:

> Like Gould, Lewontin, and Fodor, [Kramnick] seems utterly unmoved by the fact that evolution is undisputedly the scientific ground on which academic psychology proceeds. In a broad sense, all psychology is now evolutionary psychology (although EP, the narrow-gauge discipline, remains as contentious as ever). What, Kramnick seems to wonder, does evolutionary psychology have to do with literary study? And he is right to wonder. I struggle with this question all the time even as I read and absorb ever more evolutionary theory and evolutionarily inflected cognitive science. But then I used to worry about it even more (and with much more reason) when our field was given over to psychoanalysis, a fascinating rival case. As a theory, psychoanalysis is undeniably rich. As a story about the

> mind, however, it is laughable. Evolutionary psychology has different problems. As a story about the mind it is true; more details are emerging all the time to buttress its central claims. Yet its relevance to the kinds of art objects that humans make is not in the least obvious. (Vermeule 427)

The position Vermeule takes here is valuable for two reasons: first, the group of scholars among whom she situates Kramnick are by no means "unmoved" by the impact of evolutionary theory on related scientific fields. Although a more accurate description of these scientists' reservations would necessitate a lengthy detour, it can be roughly stated that this group is representative of those engaged with the complexities of how evolution differently impacts[10] biology, sociology, genetics, and cognitive science.[11] Second, Vermeule's juxtaposition of evolutionary psychology with psychoanalysis delivers the explanatory payload of a false choice between supposedly mutually exclusive options. The performative thrust of her comparison successfully distinguishes evolutionary psychology as a potentially valuable tool for literary criticism solely on the basis of an implicitly widely shared and supposedly self-evident dissatisfaction with the psychoanalytic account of the mind, yet without advancing her case for evolutionary psychology's literary fitness.

What I want to draw from this encounter is the degree to which the field of inquiry loosely grouped under the heading of Cognitive Literary Studies is overdetermined by the presumed value of science as an explanatory framework for literary interpretation that supersedes others.[12] In his reply to Vermeule,

---

[10] In many ways this interface between evolution and other fields of science also conforms to the closed/open system paradigm, in that what Vermuele (and others) grow impatient with is these scholars unwillingness to assume that the relation between evolution and sociology will be the same as the relation between evolution and genetics.

[11] Stephen Jay Gould, author of *The Structure of Evolutionary Theory*, held the Alexander Agassiz chair in Zoology at Harvard University for most of his career before leaving to teach biology and evolution at New York University. Richard Lewontin is a leading geneticist and evolutionary theorist who also occupied the Alexander Agassiz chair in Zoology at Harvard, and collaborated with Gould extensively. Jerry Fodor is a leading cognitive psychologist whose work is situated precisely at the intersection of cognitive science and computational theory.

[12] Fodor provides a concise and effective summary of this position: "The Scientific World View (by which is meant, I suppose, some sort of commitment to a physicalist ontology…" (Fodor 2003, 24).

Kramnick draws our attention to this distortion by isolating two of Vermeule's matter-of-fact claims and displaying the synergistic (performative?) effect produced by their combination. First, he classifies Vermeule's claim that "all psychology is now evolutionary psychology" as historical; second, he classifies Vermeule's statement that "[a]s a story about the mind [evolutionary psychology] is true" as metaphysical. In layering these claims together, Vermeule amplifies the force they exert over the "interdisciplinary landscape" so that when cognitive science is brought to intersect with questions of literary criticism the ground for epistemological diversity has been eliminated.

Yet without proceeding much further we can discern an additional factor at work in Vermeule's "historical claim" that falls outside Kramnick's purview. While his interpretation of the historical dimension as pertaining to "the state of play in psychology departments" remains true, he does not attend to the peculiar temporal distortion entailed when evolutionary time is invoked. In other words, to identify this as a historical claim accepts as fact, however specious, a claim as to the standing of evolutionary theory within the present composition of academic psychology—but there is more at stake in this claim than the mere details of disciplinary history. What Kramnick chooses not to attend to are the implicit consequences that the incorporation of evolutionary theory into psychology entails. The combination of these epistemological systems allows one to bypass the individuated variations of mindedness by subsuming this diversity within the evolutionary concept of the 'population'. The result is that evolution's influence on mental function is understood, or rather imagined to have produced a universally shared and homogenous cognitive architecture, the biological 'hardware' of the brain. The evolutionary flavor that makes this psychology so irresistible is the radical simplification it promises to enact on the synchronic level of population. The uninterrogated assumption here is that, to put it plainly, everyone born in the present, that is, at this moment in evolutionary time, shares the same universal mental architecture.

This brief scansion charts the rough outlines of a critique of Literary Darwinism. Additionally, it indicates how epistemological circuits are fabricated in which different knowledges are combined to amplify their explanatory wattage, but with potentially damaging effects. Certainly, the example of Literary

Darwinism, which depends highly upon maximizing the linkages between cognitive psychology and evolutionary theory, is particularly vulnerable to Kramnick's critique. Objections may legitimately be raised to the relation invoked between Literary Darwinism and the arguably wider field of Cognitive Literary Studies. To his credit, Kramnick gestures toward this question, mentioning work undertaken in cognitive literary criticism that, as he puts it, "[keeps] the principles of evolutionary biology no closer than the laws of physics" (Kramnick 2012, 453). However, what seems more important here is that Vermeule's comment makes visible the degree to which 'evolution,' considered as a signifier designating an 'evolving,' dynamic and thus unstable epistemological field, crystallizes elements both *ideological* and epistemological.

Under the intense pressure exerted by proponents of the Computational Theory of Mind first theorized by Alan Turing and later elaborated by Noam Chomsky through his work on syntactic structures in language, evolutionary psychologists have been eager to leverage developmental concepts from evolutionary theory such as mutation and adaptation to provide a causal foundation upon which limited and localized data on cognition can be integrated.[13] Indeed, Kramnick is quite right to say that Vermeule's hardline position on the central role of evolution in cognitive science "both undersells the diversity of cognitive science, much of which proceeds independent of any evolutionary framework, and oversells the verdict on evolutionary psychology, at least some of which has proven tremendously controversial" (Kramnick 2012, 453). Yet Kramnick's crisp formulation of this division obscures the fundamental issue at stake by presenting evolution itself as the problematically freighted term in question. Alternatively, it is my contention that the problem is instead *how* evolutionary theory is combined with cognitive psychology, expanding its claims about cognition to a species-wide, ahistorical register.

When Vermeule lumps Kramnick in among Gould, Lewontin, and Fodor, as a group of scholars "utterly unmoved by the fact that evolution is undisputedly the scientific ground" of psychology," she does this *because* these scholars have each worked differently to disrupt the neat alignment of Evolution

[13] For more on the historical developments, see Elizabeth Hart's article, "The Epistemology of Cognitive Literary Studies," especially pages 316-319. See also, Elizabeth Wilson's book, *Affect and Artificial Intelligence*

with adaptation.[14] This view of evolution-as-adaptation, shorn of the messy details with which Gould, Lewontin, and Fodor busy themselves, is an easily mobilized, conceptually elegant 'just-so' story that in fact responds directly to the ambient ideology of biology[15]. In the process of this transposition, natural selection is unproblematically replaced by sexual selection, or more succinctly, this conception of evolution records in a scientific register so many popularly received ideas of gender and sexuality, ratifying them with an origin story cast in a scientific vernacular. In establishing some analytic space between evolutionary theory and the discourses of cognitive science and psychology provides a resource for interdisciplinary scholarship, enabling literary theorists some leverage within discussions otherwise deemed 'above their pay grade'.[16]

At this point a more detailed understanding of how Sedgwick's work is useful for an analysis of the reading practices proposed by Cognitive Literary Studies has begun to emerge. The transition we will be making here connects Sedgwick's analytic work on gender and sexuality undertaken in *Epistemology of the Closet* with the reading practices she deploys in "Shame and the Cybernetic Fold," "Affect Theory and Theory of Mind," and "The Weather in Proust," based on affect theory and Post-Freudian psychoanalytic theories of mind such as object relations. I've chosen to read these two locations in her work alongside one another for two reasons: first, Sedgwick's extensive discussion of methodology in *Epistemology of the*

---

[14] cf. Fodor, "Steven Pinker's *How the Mind Works* and Henry Plotkin's *Evolution in Mind*. These books suggest, in quite similar terms, how one might combine CTM with a comprehensive psychological nativism and with biological principles borrowed from a neo-Darwinist account of evolution."

[15] For an extensive and cogent analysis of the signifying power held by biological explanations, see Richard Lewontin's exceptional text *Biology as Ideology*.

[16] cf. Spolsky, Ellen, "Darwin and Derrida: Cognitive Literary Theory As a Species of Post-Structuralism" p. 60, n. 15: "Literary scholars can inoculate themselves against the naïve overestimation of what social science or evolutionary biology can offer by remembering to ask themselves: What is the probability that their field (as compared with mine) is not riven by competitive hypotheses? What is the probability that, while I struggle to deal with apparently irreconcilable complexities, they know exactly what they're doing, so that I may borrow their theories and empirical data as unimpeachable evidence to resolve my controversies? Indeed the field of evolutionary biology is in the throes of several different controversies, which literary scholars are not professionally trained to evaluate."; also, Kramnick, Jonathan, "Empiricism, Cognitive Science, and The Novel" p. 278, "Those of us who work in literary study are, needless to say, not in a position to judge the relative merits of simulation theory and theory-theory as accounts of the mind itself."

Closet demonstrates how literary critical reading practices can provide a durable challenge to institutionalized discourses recirculating compromised and damaging definitions of gender and sexuality transposed into the register of cognition; second, the readings done in these essays model how theories of affect and mind can be introduced within literary analysis in ways that prioritize textual and narrative questions. My argument is that the axiomatic propositions articulated in the introduction to *Epistemology of the Closet* delimit a carefully organized 'analytic space,' a sort of conceptual testing ground in which we may critically evaluate multiple, distinct models of reading practice. Bringing these axioms into closer contact with the reading practices enacted in the essays of *The Weather in Proust* enables us to see more clearly how they record and respond literary or textual questions as opposed to questions of critical ideology with which they have been conflated.

Before taking a closer look at the specific axioms Sedgwick defines, it is worth pausing for a moment to note the particular weight carried by the term 'axiom.' At its broadest definition, an axiom is defined as "a proposition that commends itself to general acceptance; a well-established or universally-conceded principle; a maxim, rule, law." Yet the function ascribed to an axiom as a technical term in mathematics and logic is slightly more interesting, in that it defines and delimits the realm of analysis, and as such it "requir[es] no formal demonstration to prove its truth." When a system of knowledge is "axiomatized" it demonstrates that the claims of the system are derived from a limited set of propositions. In mathematics, an axiomatic structure indicates a shift in the level of complexity, the point at which calculation gives way to formal abstraction so that the system, considered as a set of formal, epistemic propositions, can now be applied to multiple contexts and hence capable of accounting for multiple different meanings. This structuring function of the axioms is important to underline because the analytic space of inquiry opened by them is organized relationally as opposed to a hierarchy of priority. These seven axioms function in a lateral, mutual relation to one another, instead of operating as subordinated, hierarchical structure. Given

this, the 'analytic fertility' of this method of inquiry is not exhausted by the literary readings which follow[17] in later chapters, but has only just barely begun to be explored.

Contemplating the book's potential usefulness for thinking through some of the political issues nascent at the time of its writing, Sedgwick steps back from these immediate concerns to underscore the more general ambition driving her work: "A point of the book is *not to know* how far its insight and projects are generalizable, not to be able to say in advance where the semantic specificity of these issues gives over to (or: itself structures?) the syntax of a 'broader' or more abstractable critical project" (Sedgwick 2003, 12). Following closely our discussion of the axiomatic structure above, this sentence captures something quite remarkable about Sedgwick's critical enterprise: that it remains in contact with a sense of its own limits while simultaneously open to further iterations and transformations of the critical project now set in motion.

This project is located precisely at this point. What is at stake here in the subtle transition from antihomophobic inquiry to gay-affirmative thinking? How does a critical project become generative one? If we accept Sedgwick's admission that her book's only categorical imperative "is the very broad one of pursuing an antihomophobic inquiry," we must understand this imperative primarily in relation to the unimaginably wide field of interpretative possibility it opens up, instead of limiting these reading practices to the terms of sexuality (Sedgwick 2003, 14). To proceed according to this 'antihomophobic imperative' necessarily moves us beyond the scope of the institutionalized discourses of sexuality and toward the infinitely rich field of the literary text. And such a move is anticipated for us: "If the book were able to fulfill its most expansive ambitions, it would make certain specific kinds of readings and interrogations, perhaps new, available in a heuristically powerful, productive, and significant form for other readers to

---

[17] Evidence of the degree to which this mathematical definition of axioms carries over into Sedgwick's analysis is found, initially and immediately, in the articulation of the axioms themselves. For example, Axiom 2 states "But we cannot know in advance how they will be different." Axiom 3: "There can't be any a priori decision…" Axiom 5: "The historical search for a Great Paradigm Shift *may* obscure the present conditions of sexual identity." (emphasis added); Axiom 7: "The paths of allo-identification *are likely to be* strange and recalcitrant. So are the paths of auto-identification." (emphasis added); I highlight the conditional grammar here because it indicates a 'long-view' that anticipates without predicting the content of the analysis the axioms make possible.

perform on literary and social texts with, ideally, other results" (Sedgwick 2003, 14).

At the outset of this chapter I introduced the argument that the axioms for antihomophobic analysis model a strategy of inquiry that can be productively applied to other epistemological 'crises,' similar in structure to the problem of homo/heterosexual definition. Here the 'crisis' of definition is to be understood as an identifiable epistemological structure, a systemic "double bind" legible in "the internal incoherence and mutual contradiction of each of the forms of discursive and institutional 'common sense'" (Sedgwick 2003, 1). These sites of internal incoherence and mutual contradiction are thus markers of a peculiarly dense confrontations of meanings, tectonic shifts in the epistemological bedrock upon which the power relations of the present culture stand. What is crucial to recognize, however, is that the two contradictions specific to the question of sexual definition and central to the book as whole that Sedgwick isolates in two neat formulations are *also* coimbricated/interdigitated with each other:

> "The first is the contradiction between seeing homo/heterosexual definition on the one hand as an issue of active importance primarily for a small, distinct, relatively fixed homosexual minority (what I refer to as a minoritizing view), and seeing it on the other hand as an issue of continuing, determinative importance in the lives of people across the spectrum of sexualities (what I refer to as a universalizing view). The second is the contradiction between seeing same-sex object choice on the one hand as a matter of liminality or transitivity between genders, and seeing it on the other hand as reflecting an impulse of separatism—though by no means necessarily political separatism—within each gender. (Sedgwick 2003, 1-2)

Why has Sedgwick focused on these contradictions? An answer is suggested by the fact that both contradictions contain organizational implications for groups of persons within these epistemologically informed power structures. This much is clearly evident and easy to grasp in her use of 'minoritizing' and 'universalizing,' but the second contradiction does not present itself as so readily congruent with potential

shapes of affiliation for both men and women. The fact that these formally symmetrical propositions fail to map correspondingly symmetrical realities does not, paradoxically, invalidate their epistemological positions. With these mutually entangled formulations in-hand we can more rigorously inquire into what *does* emerge from this "complex and contradictory map of sexual and gender definition," that, as Sedgwick herself points out, form the potential bases for "alliance and cross-identification among various groups [that] will also be plural" (Sedgwick 2003, 89).

In the review of the axiomatic propositions, I suggest that Sedgwick is engaged in a process of forging exactly these lines of 'alliance and cross-identification' by shaping antihomophobic inquiry according to the principles of feminist theory. I propose that the axioms also organize her of theorization of reading practices based on theories of affect and Object Relations. Our goal here is to make some of analytic strategies elaborated in *Epistemology of the Closet* available for our critique of the reading practices promoted by Cognitive Literary Studies and to prepare the way for a more thorough theorization of how the affect theory developed by Silvan Tomkins productively informs literary reading practices that attend to the middle ranges of agency often lost to analysis.

> Axiom 2: *The study of sexuality is not coextensive with the study of gender; correspondingly, antihomophobic inquiry is not coextensive with feminist inquiry. But we can't know in advance how they will be different.* (Sedgwick 2008, 27)

In the exposition of Axiom 2, Sedgwick works to clear some conceptual ground for the type of analysis she labels "antihomophobic," one that takes "sexuality" as its primary term of inquiry.[18] She begins this section by first taking stock of the current state of affairs regarding critical/theoretic accounts of the terms sex, gender, and sexuality, from within feminist theory, and then (evident in the wording of the axiom)

---

[18] cf. "*Epistemology of the Closet* is a feminist book mainly in the sense that its analyses were produced by someone whose thought has been macro- and microscopically infused with feminism over a long period of time. At the many intersections where a distinctively feminist (i.e., gender-centered) and a distinctively antihomophobic (i.e., sexuality-centered) inquiry have seemed to diverge, however, this book has tried consistently to press on in the latter direction." (Sedgwick 2008, 15-16).

moves to establish a relation between feminist inquiry and the nascent field of antihomophobic inquiry. Throughout its development, at times more strongly stated than others, a persistent reference to the tools of feminist theory is cultivate and underlined, yet as these tools are regrafted, a clarified picture of the differences between these disciples is discernable. Here's how Sedgwick outlines the situation:

> Sex, gender, sexuality: three terms whose usage relations and analytical relations are almost irremediably slippery. The charting of a space between something called "sex" and something called "gender" has been one of the most influential and successful undertakings of feminist thought. (Sedgwick 2008, 27)

These two sentences capture a great deal, both about the historical moment in which Sedgwick was writing and about the complexities of interdisciplinary scholarship. To begin with, Sedgwick's recognition of the fundamental slippage active in the "usage relations and analytical relations" of sex, gender and sexuality records is a combination of referential crises. Sedgwick grounds this claim simply by observing the plurality of feminist models available from three schools or waves of feminist analysis. She draws from the work of Constructivist Feminists, Radical Feminists, and (what she refers to as) Foucault-influenced analysis, to show the different treatments given to the terms sex, gender, and sexuality, and the new insights into the structures of oppression that interlock around these terms produced as a result.

For Sedgwick, the signal example of this feminist pushback against oppression is Gayle Rubin's famous theorization of the "sex/gender system" deployed by feminist thinkers to "gain analytic and critical leverage on the female-disadvantaging social arrangements that prevail at a given time in a given society, by throwing into question their legitimative ideological grounding in biologically based narratives of the 'natural'" (Sedgwick 2003, 28). The brilliance of Rubin's intervention is that it provided an analytic mechanism with which feminists could now separate the social forms of oppression from their biological alibis. By postulating a difference between sex and gender Rubin thereby created a space within which to catalogue, examine and analyze the articulations of this difference.

In connection with this point, it is important to take note the choice of phrasing when she writes "something called 'sex' and something called 'gender'" because this formulation records the variable relation underlying these terms, so profitably exploited by each mode of feminist inquiry. Indeed, it is because these terms have been abstracted and differently formalized by successive models of feminist analysis that, through Sedgwick's conceptual collage work, we can see how "sex" and "gender" overlap and diverge, naming distinctly structured nodes of social organization to which the effects of power attach. Yet Sedgwick's purpose here is not to adjudicate between the competing truth claims, implicit or explicit, fueling feminist theories of gender oppression, but rather to demonstrate by way of these examples that "there is always at least the potential for an analytic distance between gender and sexuality" (Sedgwick 2003, 30).

Approaching gender oppression with Rubin's sex/gender distinction enabled a mode of 'reading otherwise' that posed a durable challenge to the 'status quo.' Seen through the lens ground by Rubin's sex/gender analytic, the forms of female oppression became legible *as* so many emulsions of biological and cultural elements in unstable suspension. When feminist analysis deploys the sex/gender analytic at the site of "female-disadvantaging social arrangements" it actively disrupts the legitimacy delivered by a biological endorsement. The spatial language Sedgwick uses to describe how this mechanism of reading enabled "[t]he charting of a space between something called 'gender' and something called 'sex,'" underlines an important component. Acting as a capacitor,[19] the space created interrupts and displaces the charge circulating within the epistemological-ideological circuit of oppression.

Having now prepared the way Sedgwick quietly initiates a terminological renovation, a retooling of the 'sex/gender' rubric so that it will be responsive to the needs of antihomophobic inquiry. Now 'gender' will

---

[19] A capacitor (originally known as a condenser) is a passive two-terminal electrical component used to store energy electrostatically in an electric field. The forms of practical capacitors vary widely, but all contain at least two electrical conductors (plates) separated by a dielectric (i.e., insulator). The conductors can be thin films of metal, aluminum foil or disks, etc. The 'nonconducting' dielectric acts to increase the capacitor's charge capacity. A dielectric can be glass, ceramic, plastic film, air, paper, mica, etc. Capacitors are widely used as parts of electrical circuits in many common electrical devices. Unlike a resistor, a capacitor does not dissipate energy. Instead, a capacitor stores energy in the form of an electrostatic field between its plates.

denote "the whole package of physical and cultural distinctions between men and women," that is, either biological or cultural, and 'sex' will denote "what modern culture refers to as sexuality" (Sedgwick 2003, 29). Yet 'sex' is no simple matter either; in short order Sedgwick's inventory of the [field of] possible meanings ascribed to sex/sexuality that "cluster most densely around certain genital sensations but are not adequately defined by them" culminates in a critical framework of closed and open systems, and which will be reiterated in the literary engagements of "The Weather in Proust" and "Affect Theory and Theory of Mind". Citing the career-long demonstrations by Freud and Foucault of the innumerable ways that human sexuality spills over the borders of biological procreation, it's clear that sex-uality cannot be neatly mapped within the space opened by the "feminist-defined sex/gender distinction," that is, as the opposition between a biological order and a cultural order. Indeed, the following sentence anticipates the need for a different analytic mechanism than the feminist sex/gender distinction, one equipped with a differently structured relation between what Sedgwick calls 'gender' and 'sex'. Having seen that the formal capacity of either 'biology' or 'culture' is now overloaded and unable to intelligibly account for the range of referents mobilized by the word sexuality, we must fabricate more flexible tools of analysis. To grapple with sexuality requires the means to address a conceptual field structured like a spectrum, which accommodates a range of contradictory positions, markedly different from the field of inquiry opened up by the feminist sex/gender distinction. Here's how she puts it:

> To note that, according to these different findings [of Freud and Foucault], *something* legitimately called sex or sexuality is all over the experiential and conceptual map is to record a problem less resolvable than a necessary choice of analytic paradigms or a determinate slippage of semantic meaning; it is rather, I would say, true to quite a range of contemporary world views and intuitions to find that sex/sexuality *does* tend to represent the full spectrum of positions between the most intimate and the most social, the most predetermined and the most aleatory, the most physically rooted and the most symbolically infused, the most innate and the most learned, the most autonomous and the

most relational traits of being. (Sedgwick 2003, 29)

Notice the care with which the question of 'sexuality' has been separated from the feminist analysis of 'gender' as cultural effect of chromosomal sex: it is not a matter of selecting the correct paradigm forged in feminist foundries, or of a determinate amount of semiotic historiography necessary to unearth the 'true' meaning of sexuality. Rather for antihomophobic inquiry it is occasion sensitive, whereby the axis of difference called sexuality must be carefully designed in order to accommodate what is "true to quite a range of contemporary world views and intuitions" about sexuality.

Allow me to pause a moment in order to clarify the connection I gestured to above between the reformulation of the feminist-defined sex/gender distinction and the literary reading practices deployed in the essays. Sedgwick began this section by proposing a degree of initial overlap between the tools of feminist inquiry and those yet-to-be assembled for antihomophobic inquiry. Immensely productive, the "sex/gender" analytic mechanism created by Gayle Rubin for her analysis of gender oppression was taken by Sedgwick as a die with which to cast a similarly powerful analytic tool for antihomophobic inquiry. Accounting for the mutual entanglement of 'gender' with 'sexuality,' Sedgwick posits a heavily trafficked diacritical frontier between these analytic axes whenever questions of same-sex desires and relationships are asked. Yet what distinguishes Sedgwick's theorization of the sexuality/gender system from Rubin's sex/gender system is that, as a mode of reading, Sedgwick's mechanism is capable of reading the relation between two open systems, that is, across just such a diacritical frontier. For Sedgwick, both sexuality and gender name fields that are themselves differently constituted emulsions of biological and cultural phases, and thus neither term designates an 'origin' of meaning that is then processed by the other. It is the interface between these open systems that will be the primary locus of antihomophobic inquiry, so that it will remain highly possible to read for meanings of "something called sexuality" that are not determined by the gender of object choice.

In choosing to label these freshly minted analytic axes as 'open systems' I draw a direct connection to Sedgwick's essay "The Weather In Proust," where she initiates a discussion of the important question "of

how open systems relate to closed ones, or perhaps better put, of how systems themselves move between functioning as open and closed" (Sedgwick 2011, 3). The spare formulation given here in the language of systems, I argue, bears the imprint of Axiom 2 while operating as the organizing thematic for reading practices ostensibly unrelated to antihomophobic inquiry. My suggestion is that this more general articulation of her methodological approach to reading Proust, it the development anticipated earlier: "not to be able to say in advance where the semantic specificity of these issues gives over to (or: itself structures?) the syntax of a 'broader' or more abstractable critical project" (Sedgwick 2003, 12).

As we have seen from our analysis of Axiom 2, the space of inquiry it delimits is radically open, enabling us to map a wide range of meanings that are eliminated or bypassed under available modes of critical reading. This aspect of Sedgwick's epistemological labor, that charts multiple interactions between or even among open systems, is difficult to recognize at work in the tightly focused literary readings and yet it remains fundamental to her analytic style and creative insight.

This specifically literary moment in Sedgwick's writing communicates powerfully with a problematic trend in critical reading practices that utilize a closed system to constrain an open system. This coarsely sketched outline approximates characteristic stance that entails a problematic over confidence that concepts from cognitive psychology may be imported into literary analysis as a correction to the non-rigorous interpretive methods produced by literary criticism of post structuralism and deconstruction. Such is the mood that animates Alan Richardson's comments that "cognitive science can…provoke needed self-reflection among literary theorists and critics," offered in the introduction to a collection of representative essays on Cognitive Literary Studies (Richardson 3). Yet into which specific issues can cognitive science lend assistance? Richardson helpfully provides a brief primer:

> The Saussurean postulate of systemic linguistic arbitrariness, for example, though relied upon throughout structuralist and poststructuralist criticism, no longer seems tenable in light of post-Chomskian linguistics. Psychoanalytical accounts of early development, which underwrite any number of literary interpretations, have been rendered problematic

> at best by the past thirty years of empirical research in developmental psychology. The notion of 'interpellation' cited in many literary-theoretical accounts of subject formation seems remarkably thin and abstraction in relation to the vast amount that has been discovered concerning child development and language acquisition over the past several decades. (Richardson 2004, 3)

While it is arguably not surprising to find Richardson stumping for cognitive science given the context, his manner of delivery indicates that a sharp distinction is being drawn between the 'then' and 'now' of "the cognitive revolution."[20] Neatly repackaging the historicism at issue in the exchange between Vermeule and Kramnick noted earlier, Richardson lays siege to the theoretical resources that inform other literary reading practices, implicating these critical discourses as outdated and thus obsolete when compared to the presumably vastly superior explanatory power of cognitive psychology informed by cognitive science.

Against this background Sedgwick's framework of closed and open systems emerges as distinctively nuanced and delicate, capable of implementing different analytic strategies without jockeying for mastery, particularly as she negotiates the terrain between literary, psychological and scientific discourse. Notice the striking difference to her description of the engagement:

> Rather than trying to bring Proust into explicit relation with the science of chaos and complexity, however, or with the science of his own day, I'm trying in the present project to use these topics in contemplating some characteristically Proustian modes of being, of relation to self and the world. Like, I think, many readers of Proust, I especially want to understand his continuing access to a psychology of surprise and refreshment, as well as his nourishing relation to work. (Sedgwick 2011, 3)

Unlike the landscape depicted by Richardson, in which cognitive science monopolizes the ground of

---

[20] cf, Patrick Colm Hogan, *Cognitive Science, Literature, and the Arts: A Guide For Humanists*, 1-3.

legitimate inquiry, Sedgwick explicitly suspends the expectation that scientific insights will deliver satisfactory answers once brought into proper alignment with the questions of the literary text. Moreover, Sedgwick's inquiry preserves the possibility that different epistemological systems may differently support an interpretive project. The topics of chaos and complexity help enact a practice of reading otherwise that allows for a *thinking otherwise* about complexity in literary systems, decentering common sense critical practices so that new insight might be glimpsed.

At this point I want to back up to indicate how the steps we've gone through thus far constellate into the shape of a broader but no less significant question about the status of the literary text. To speculate as Sedgwick does here on the transformative potential of the 'inexhaustible interest' of a psyche, a landscape, or a sentence, challenges some of the central critical assumptions that underwrite the reading practices of cognitive literary studies. Indeed, the entire premise of Axiom 2 is that different strategies of inquiry are needed to access different problems of interpretation; that feminist inquiry and antihomophobic inquiry, along with the concepts of gender and sexuality they mobilize, while related, are ultimately not the same set of tools and questions.

The motif characterizing the reading practices developed by cognitive literary studies is the belief that the authority of this science is also operative for textual analysis, and furthermore that these methods provide exclusive purchase on the single—or most fundamental—meaning of a literary text. For cognitive literary studies the belief that the truth of the text is in short supply works simultaneously to ratify its 'discoveries' while delegitimating other, differently tooled analytic projects.

How then might we address the 'scarcity of meaning' premise? In order to succeed at this task it will be necessary first to more fully elaborate how Cognitive Literary Studies constitutes its discursive authority and then to formulate an alternative mode of reading whereby we can shift the conceptual deadlock that endlessly reabsorbs the energy of critique. Sedgwick sets the tone for precisely this form of engagement in two related passages in *Epistemology*, first, in which she demonstrates the need for reconceptualizing the power afforded by knowledge as instead a relational field in which multiple knowledges and a

corresponding plurality of ignorances are active, and, second, in which she formulates the 'minoritizing vs universalizing' rubric as an alternative to the more familiar terms of the essentialist vs constructionist debate. Taken together, these passages are instructive as a model for containing the explanatory sprawl of cognitive literary studies, as its strength stems from a subtle conflation of ignorance with critique and knowledge with truth.

Concerning the question of power, Sedgwick's analysis proves valuable for our analytic project for the succinct explanation it gives of the intricate relations between 'knowledge' and 'ignorance' understood as performative structures of power. Viewed from a different angle, we might rephrase this same point to say that knowledge and ignorance do not denote corresponding positions of power and powerlessness; ignorance can also be deployed to gain power over a situation, a person, or a text.

Citing the works of Michel Foucault, Jacques Derrida, Thomas Kuhn, and Thomas Szasz, as thinkers representative of this awareness, Sedgwick warns us not to appeal "too directly to the redemptive potential of simply upping the cognitive wattage on any question of power" in want of a resolution (Sedgwick 2008, 8). Recognizing this allows for a reassessment of the strategies available for engaging with an epistemological discourse confident of its power. In the following paragraph Sedgwick outlines how to exploit this 'deadlock' and loosen the grip of power upon knowledge by multiplying forms of ignorance. By implementing these ignorances in the practice of asking questions about the legitimacy of the status-quo, they effectively flank a well-fortified hold on power. The result is a veritable 'Socratic coup' whereby interlocutors are held to account for the alignment of word and deed through a thoroughly deconstructive reading practice that weakens the epistemological joists supporting the infrastructure of power.

> Rather than sacrifice the notion of "ignorance," then, I would be more interested at this point in trying, as we are getting used to trying with "knowledge," to pluralize and specify it. That is, I would like to be able to make use in sexual-political thinking of the deconstructive understanding that particular insights generate, are lined with, and at the same time are themselves structured by particular opacities. If ignorance is not—as it

> evidently is not—a single Manichaean, aboriginal maw of darkness from which the heroics of human cognition can occasionally wrestle facts, insights, freedoms, progress, perhaps there exists instead a plethora of *ignorances*, and we may begin to ask questions about the labor, erotics, and economics of their human production and distribution. Insofar as ignorance is ignorance *of* a knowledge—a knowledge that may itself, it goes without saying, be seen as either true or false under some other regime of truth—these ignorances, far from being pieces of the originary dark, are produced by and correspond to particular knowledges and circulate as part of particular regimes of truth. We should not assume that their doubletting with knowledge means, however, that they obey identical laws identically or follow the same circulatory paths at the same pace. (Sedgwick 2008, 8)

Echoing the sentiments we have remarked above from literary critics who are quick to plead ignorance regarding specific on-going debates within cognitive psychology, the lesson learned from deconstruction that "particular insights generate, are lined with, and at the same time are themselves structure by particular opacities" also applies here. Literary critics who apply cognitive psychology to a text to determine its 'truth' frequently cite their own ignorance of the scientific basis as a way of qualifying their claims from being invalidated on the grounds of lack of expertise.[21] While cognitive science as a mode of scientific inquiry has proven itself to be an extremely powerful tool for addressing certain questions of human cognition, the 'truths' it produces are nonetheless limited. Furthermore, the degree to which Cognitive Literary Studies presents itself as a powerful explanatory mechanism for textual analysis over and against other forms of analysis is perhaps better understood mobilization of ignorances that, together, enable a radical simplification of textual problems. In making these multiple 'ignorances' visible and distinct, we can "being to ask questions about the labor, erotics, and economics of their human production and distribution" so

[21] The dangers of such formulations are strikingly evident in Zunshine's exchange with Disabilities Studies scholar, Ralph James Savarese in "The Critic as Neurocosmopolite; Or What Cognitive Approaches to Literature Can Learn from Disability Studies: Lisa Zunshine in Conversation with Ralph James Savarese".

that we may begin to approach these problems not oppositionally, but from the side, to work on the laws under which they operate *diagonally*. This process of charting and mapping power relations opens the space for reading otherwise and thus thinking otherwise, destabilizing the existing pathways between knowledge and ignorance through which power circulates.

> Axiom 4: *The immemorial, seemingly ritualized debates on nature versus nurture take place against a very unstable background of tacit assumptions and fantasies about both nurture and nature.* (Sedgwick 2008, 40)

Sedgwick's work is exemplary for her engagement with seemingly intractable problems and her ability to propose strategies for 'thinking otherwise' by way of practices for 'reading otherwise'. Yet how is it that a problem is identified as 'intractable'? In the opening pages of *Epistemology* Sedgwick describes the conditions that have led to her inquiry, articulating in the first sentence the central proposition of the book: "*Epistemology of the Closet* proposes that many of the major nodes of thought and knowledge in twentieth-century Western culture as a whole are structured—indeed, fractured—by a chronic, now endemic crisis of homo/heterosexual definition, indicatively male, dating from the end of the nineteenth century" (1). The grounds for this claim are plainly stated: the peculiar interrelation of thought and politics, spheres which operate according to distinct laws and principles nonetheless overlap significantly. This overlap makes possible a violently dramatized series of struggles all of which orbit around the question of sexual definition and extend forward from the turn of the twentieth century. The "internal incoherence and mutual contradiction of each of the forms of discursive and institutional 'common-sense'" becomes a politically volatile combination because it creates the conditions by which multiple truths are operative. Sedgwick identifies the contours of these contradictions succinctly:

> The first is the contradiction between seeing homo/heterosexual definition on the one hand as an issue of active importance primarily for a small, distinct, relatively fixed homosexual minority (what I refer to as a minoritizing view), and seeing it on the other

> hand as an issue of continuing, determinative importance in the lives of people across the spectrum of sexualities (what I refer to as universalizing view). The second is the contradiction between seeing same-sex object choice on the one hand as a matter of liminality or transitivity between genders, and seeing it on the other hand as reflecting an impulse of separatism—though by no means necessarily political separatism—within each gender. (Sedgwick 2008, 1-2)

The space delineated by these oppositional epistemological spectrums is inseparable from the political projects to which they are put, but rather than supplanting one another, these discourses—past, present, and future—churn and recirculate their 'common-sense' truths unevenly throughout the history of the 20th century. Ranging from brutal to banal, legal regulation of acts and discrimination of persons identified as prohibited sexualities was legitimated politically by using these "forms of discursive and institutional 'common-sense'" as backing for their enforcement.

Sedgwick's choice to "demur vigorously" the task of adjudicating between these two understandings of sexual definition marks a second substantial difference to Feminist analysis. Recognizing that the "gay essentialist/constructivist debate takes its form and premise" from the long "history of other nature / nurture or nature/culture debates," places this rubric in direct relation to the sex/gender system formulated by feminist analysis. The critical leverage this mode of reading secured for feminist politics was found in, as Sedgwick puts it, the "implication that the more fully gender inequality can be shown to inhere in human culture rather than in biological nature, the more amenable it must be to alteration and reform" (Sedgwick 2008, 41).

Thus did constructivist understandings of homosexuality followed in step, drawing from the tradition of nature/culture thinking, viewing "culture as more malleable relative to nature" and therefore arguing that biology does not dictate a person's proper sexuality. Although this intervention, much like its feminist forerunner, targeted the forms of homosexual oppression based on biology, the argument that sexual aim is "not a hard-wired biological given" has proven dangerous in a political landscape dominated by the

"unimaginatively large" "scope of institutions whose programmatic undertaking is to prevent the development of gay people" (Sedgwick 2008, 42). Differently put, the feminist sex/gender system was politically effective in a way that is unavailable to gay-affirmative thinking, because the structures of oppression have developed according to different necessities and regimes of enforcement.

This gap between the highly focused literary critical readings undertaken in *Between Men* and *Epistemology of the Closet* and the distinctively different theoretical approach seen in the essays collected in *Touching Feeling*, is suggestive of the development of a literary critical tool set informed by Tomkins's affect theory. In the materials collected and published posthumously in the volume *The Weather in Proust*, the essay "Affect Theory and Theory of Mind" marks Sedgwick's most sustained argument for the use of Tomkinsian affect theory in literary critical analysis. Originally delivered on April 14, 2008, for a conference at the University of Leiden titled "Affective Operations in Art, Literature, and Politics," the essay is noteworthy for the engagement that its title blatantly signals. In the abstract submitted to the conference for "Affect Theory and Theory of Mind" Sedgwick explains that the paper "will discuss two usages of the term 'theory' that have some attractive commonalities" (Sedgwick 2011, 161). The commonalities being, first, a "democratic understanding of 'theory'" as primarily the "lay theorizing" done by individuals to navigate their daily lives, and second, that both theories emerge from the possibility of wrongness (Sedgwick 2011, 161).

The essay is important for at least two additional reasons. First, it is in this essay that Sedgwick directly engages with literary criticism that uses Theory of Mind as an analytic tool and thus provides some clues as to how exactly Tomkinsian affect theory is a better tool for addressing literary questions; second, the essay begins by contextualizing the excitement she felt for the possibilities offered by Tomkins's affect theory through recourse to the first axiomatic proposition of *Epistemology of the Closet*, written almost twenty years prior in 1990.

Redeploying the first axiom that framed the literary critical project undertaken in *Epistemology*, suggests that "Affect Theory and Theory of Mind" may reasonably be thought of as a further iteration, a feeding-forward of these reading practices, and, moreover, as continuing to develop theoretical support for

engaging with texts that avoids the deadlock of critical discourse around a set of binarisms. Sedgwick explains that Tomkins's work on affect theory seemed well equipped to respond to axiomatic assumptions that have informed her work from as far back as *Between Men*. In striking contrast to "inconceivably coarse axes of categorization" available in critical discourse, Tomkins's work "made a case for affect as the system that uniquely mediates between the imperceptibly fugitive, almost impersonal nuance of a feeling, at one extreme, and at the other the most inveterate bent of a lifetime's motive and style of being, thinking, and relating" (Sedgwick 2011, 145). This exceptionally scalable quality of Tomkins's theory of affect, that it is equally capable of addressing the "nuance of a feeling" as well as a "style of being," suggest that affect can be leveraged as a descriptive vocabulary for ontological relations without proscribing them.

She continues on to say that "[i]t was also exciting how firmly and flexibly, at once, Tomkins rooted such processes in the body and brain—at that time a profoundly unfashionable move in critical-theory discourse, though now the opposite" (Sedgwick 2011, 145). This comment prepares the way for her consideration of a growing trend in popular and theoretical discourse that draws on "Darwinian neuroscience" for its analytic energies and concepts, chief among them being the concept of Theory of Mind. The ostensible similarity between these two theories, each based on "processes in the body and brain," will be examined with reference to Axiom 1: "People are different from each other," in order to determine the potential of each theory "to support any understanding of truly diverse developmental possibility" (Sedgwick 2011, 145). Suspending for the moment a deeper investigation of the concept of Theory of Mind as well as its historical development, our present purpose is to clarify some of the broader ways these theories answer to the requirements outlined by the axiomatics of Sedgwick's earlier work and so we will restrict our view to her presentation of these theories in this essay.

The concept known as 'Theory of Mind' derives from work done in a variety of fields including "animal behavior, child development, and cognitive neuroscience," and designates a "purportedly delineated and unitary cognitive ability that is said to develop in some few animals, and also in most people by the age of four" (Sedgwick 2011, 145). It is the "stable understanding" that one "either has or has not" that "others

(and ourselves) have their own thoughts, feeling, perception, beliefs, and so forth—as opposed to thinking one's own 'mental' experience [is] identical to objective reality" (Sedgwick 2011, 145; Blackburn 2000). The supplement to this internal aspect of Theory of Mind is "the ability to attribute to other people as well as oneself mental states like 'pretending, thinking, knowing, believing, imagining, dreaming, guessing, and deceiving" (Baron-Cohen, quoted in Sedgwick 2011, 146). It is to this aspect of Theory of Mind that Sedgwick draws our attention because it overlaps with a central premise in Tomkins's affect theory, that is, "the attribution of theoretical activity to lay people" (Sedgwick 2011, 146). On this point Sedgwick emphasizes that one of the things she "value[s] most in Tomkins, in his exposition of 'affect theories' and some associated ideas, is the sense of irreducible continuity between what people with affects do, on the one hand, and on the other, what experts—people like Tomkins—do when they think about affects." (Sedgwick 2011, 146)

The importance of this "irreducible continuity" between lay persons and experts is easy to overlook given the truncated citation Axiom 1 with which Sedgwick begins the essay. Returning to the full entry in *Epistemology* we can better discern how Sedgwick understands this aspect of Tomkins's theory to be crucial for responding to the pressure of the theoretical vacuum outlined there. In the paragraphs that follow we are presented with a demonstration of the limitations to our critical vocabulary of differences that transect our current ways of understanding sexuality. To accomplish this, Sedgwick enumerates a cursory list of thirteen differences that "can differentiate even people of identical gender, race, nationality, class, and 'sexual-orientation—each one of which…retains the unaccounted-for potential to disrupt many forms of the available thinking about sexuality" (Sedgwick 2008, 24-5). Yet the differences enumerated on this list "depend radically on a trust in the self-perception, self-knowledge, or self-report of individuals" which is all but disregarded data under the dominant rubric we have available for thinking about sexuality: the presumed centrality of gender of object choice (Sedgwick 2008, 26). The additional differences adhering to individuals' experience of their sexuality are neglected and devalued to the extent that they find no expert theoretical discourse for support. Simply put, knowledge of sexuality and sexual desire is oddly

structured by an overwhelming differential of authority that privileges expert—that is, institutional discourse.

The immediate consequence of this analytic negligence is a "delegitimation of felt and reported differences" that makes it possible to "alienate, conclusively, *definitionally*, from anyone on any theoretical ground the authority to describe and name their own sexual desire" (Sedgwick 2008, 27, 26). By contrast, the "irreducible continuity" between the theorizing done by lay persons and that done by experts is fundamental to the structure of Tomkins's affect theory and avoids this wholesale alienation of the personal authority to theorize about affects. Yet it is crucial to grasp just *how* fundamental this "irreducible continuity" is to Tomkins's ideas, because it forms the basis for an "understanding of truly diverse developmental possibility" (Sedgwick 2011, 146).

While Tomkins's work was primarily dedicated to an exploration and mapping of the affect system, he maintained throughout that affect and cognition were "not just complexly interfused, but in human beings essentially inextricable" (Sedgwick 2011, 147). He posited that the linkage between these two systems can be best understood as the various operations of many different types of feedback mechanisms that govern the processes and subsystems of the body. This basic orientation guarantees the centrality of "self-perception, self-knowledge, and self-report" as the very foundation of Tomkins's extended meditations on further aspects of the affect system. By taking feedback mechanisms as an organizing principle of the affect system, the status of error is transformed and given new value as the necessary data for further calibration. With this configuration error-making functions through the body's feedback mechanisms to provide the 'data' of the self and of experience. As individuals grow older they accumulate more experiences of their affects, steadily learning which things stimulate them to excitement or to anxiety and consolidating this data into "'theories,' in much the same way that theories are constructed to account for uniformities in science or in cognition in general" (Tomkins, quoted in Sedgwick 2011, 148). The picture of how "affect theories" emerge from a person's experiences is thus decidedly democratic, in that the meaning of these affect theories as "a simplified and powerful summary of a larger set of affect

experiences" is not dictated by expert authority.

Having briefly examined how Sedgwick's turn to the work of Silvan Tomkins connects with the axioms outlined in her earlier literary critical work, I want to return now to the text of the later essay to establish her reservations regarding Theory of Mind and some of its literary applications. Proceeding in this way will have the additional benefit of providing a glimpse of the broader issues circulating around the intersection of the humanities with cognitive psychology, while also allowing us to remain close to Sedgwick's specific engagement here.

Recall that Theory of Mind refers in general to the "unitary cognitive ability" to understand that other people have their own thoughts, feelings, perceptions and beliefs which may differ from our own. And that along with this awareness of the opacity of other minds, Theory of Mind includes the "ability to attribute to other people as well as oneself mental states like 'pretending, thinking, knowing, believing, imagining, dreaming, guessing, and deceiving" (Baron-Cohen, quoted in *Weather* 146). Broadly speaking, when an individual develops the cognitive "mechanism" called a Theory of Mind it "is understood as an access to sociality" whereby interaction among persons is organized by this shared capacity to know that others' see things differently while also attributing to them any number of potential mental states (*Weather* 146). What's important to observe already at this early stage is that Theory of Mind, as both a cognitive mechanism and organizing principle of sociality, appears to straddle a division between self and others while simultaneously instantiating it. The effect it has is almost one that transforms the opacity of other minds translucent. If viewed from a different angle we might reasonably think of Theory of Mind as having a community building function, in that the other persons who share this cognitive capacity are presumably thinking about your mind in the same way you're thinking about theirs (or, at least, in the same way that you think they're thinking about your thinking).

This subtle alignment of cognitive structure with the capacity for (a) sociality takes place as a presumably logical extension of Theory of Mind, and yet this complicates the sense of difference between persons that grounds Theory of Mind in the first place. However, Sedgwick is quick to identify an area in

which in the alignment of cognitive structure with a capacity for sociality is anything but subtle, and instead operationalizes Theory of Mind as a diagnostic tool of still greater reach:

> And in terms of human cognition, it has become common over the last decade—although, as we'll see—controversial—to define Theory of Mind in an exact negative relation to autism: Theory of Mind is the thing that autistic individuals don't have, and autism is the thing that people who don't have a Theory of Mind have. In fact, many of the defining moves for Theory of Mind have come from autism researchers like Uta Frith and Simon Baron-Cohen, and the popularization of this negative equation is one of the things that have consolidated understandings of both Theory of Mind and autism as actually existing, unitary, and neurobiologically grounded things in the world. (Sedgwick 2011, 146)

Axiom 1: *People are different from each other* (Sedgwick 2008, 22)

Before proceeding further with Sedgwick's evaluation of Theory of Mind with particular regard to its suitability for literary relevance, I'd like to denote some significant similarities between the structuring effects exercised by Theory of Mind as a principle of categorization and the gender-of-object-choice as a principle of sexual categorization. While it may initially appear that no possible connection might here be drawn between these two ostensibly different registers, I argue that making this connection is precisely what Sedgwick intended to enable with Epistemology of the Closet. Indeed, if we return once more to the introduction, we find Sedgwick explicitly discussing this outcome:

> Any critical book makes endless choices of focus and methodology, and it is very difficult for these choices to be interpreted in any other light than that of the categorical imperative: the fact that they are made in a certain way here seems a priori to assert that they would be best made in the same way everywhere. I would ask that, however sweeping the claims made by this book may seem to be, it not be read as making that particular claim. Quite

> the opposite: a real measure of the success of such an analysis would lie in its ability, in the hands of an inquirer with different needs, talents, or positionings, to clarify the distinctive kinds of resistance offered to it from different spaces on the social map, even though such a project might require revisions or rupturings of the analysis as first proffered. The only imperative that the book means to treat as categorical is the very broad one of pursuing an antihomophobic inquiry. If the book were able to fulfill its most expansive ambitions, it would make certain specific kinds of readings and interrogations, perhaps new, available in a heuristically powerful, productive, and significant form for other readers to perform on literary and social texts with, ideally, other results. (Sedgwick 2008, 14)

Some care is required to understand exactly how this later work on affect theory and Theory of Mind corresponds to her earlier work on sexuality, but the resources for doing so are in fact located entirely within her explication of Axiom 1. If we recall the manner with which Sedgwick approached sexuality as a category of difference, we remember that her two strongest objections were that sexuality was defined exclusively according to the criterion of gender-of-object-choice, and that classifying persons with this rubric overrode a set of differences that destabilizes its precise taxonomy of homo/heterosexual categorization. By identifying an cursory range of differences over which our critical discourse on sexuality provides no leverage because it is tuned singularly to the note of gender, Sedgwick demarcates a "reservoir" of untapped epistemological potential circulating through literary texts:

> This is among other things a way of saying that there is a large family of things *we know* and need to know about ourselves and each other with which we have, as far as I can see, so far created for ourselves almost no theoretical room to deal. The shifting interfacial resistance of "literature itself" to "theory" may mark, along with its other denotations, the surface tension of this reservoir of unrationalized nonce-taxonomic energies; but, while

> distinctively representational, these energies are in no sense peculiarly literary. (Sedgwick 2008, 24)

Lacking a conceptual vocabulary to account for these differences, these "unrationalized nonce-taxonomic energies," we are unable to even begin mapping the knowledges "literature itself" withholds from critical analysis. Indeed, these two sentences nicely capture both the explicit project of *Epistemology of the Closest*, to elaborate a literary critical discourse on sexuality that is primarily antihomophobic, as well as the implicit result, that it would model a different type of engagement between "theory" and "literature itself," one attentive to this representational excess. The success of this project, Sedgwick writes, "would require framing its questions in such a way as to perform the least possible delegitimation of felt and reported differences and to impose the lightest possible burden of platonic definitional stress. Repeatedly to ask how certain categorizations work, what enactments they are performing and what relations they are creating, rather than what they essentially mean[, has been my principal strategy]" (Sedgwick 2008, 27).

When seen in this light, the two specific points of contrast made between Theory of Mind and Tomkins's affect theory in her later essay gain new significance as evaluations of the potential these two types of 'theory' may have for literary relevance. The first concerns each theories' capacity for change or alteration over time:

> Unlike Theory of Mind, that is—which once acquired is supposed to be quite a stable element of the psyche—affect theory, as Tomkins posits is, continues over the individual lifespan to behave *as* a theory: that is, to a greater or lesser extent it remains subject to revision, reorganization, or even revolutionary disruption, though such internal events may become decreasingly probable. (Sedgwick 2011, 149)

Sedgwick's criticism of Theory of Mind on this point is that it proves to be remarkably *in*flexible as a theory, but rather appears to be a reified "element of the psyche" that one either has or doesn't. Tomkins's theory, structured from the ground up around feedback between systems within the body, remains

remarkably equipped to respond to variations that take place within an "individual lifespan." It is thus responsive to a range of "felt and reported differences," whereas Theory of Mind produces a binary diagnostic. Why, then, is this so unsuitable? The answer is brought more clearly into focus with the next point of contrast.

> We are given to understand, about Theory of Mind, that there is only one of them—shared by, well, almost everybody. In contrast to what we are told by about Theory of Mind, in Tomkins's understanding of affect theory, though it's similar in being a form of lay theory, there can be and are very different affect theories. (Sedgwick 2011, 150)

Not only does the theory of Theory of Mind separate those without it into obscurity, the homogenizing effect of the positive valence (those of us who have a theory of mind) yields no traction over further differences. A strangely paradoxical result: Theory of Mind arrives as the cognitive development that 'people are different from each other' only to resolve this difference in a universally shared cognitive structure.

When Sedgwick turns her discussion to Lisa Zunshine's book, *Why We Read Fiction: Theory of Mind and the Novel*, to evaluate the results of a critical engagement informed by Theory of Mind, the effect of this vertically integrated critical discourse is brought fully into light. Here it is important to take account of how Sedgwick situates her critique of Zunshine, or perhaps better said, of Theory of Mind's literary purchase. Of paramount concern is the point of contact between the problematics consolidated under the concept of Theory of Mind and those evident in a novel's intersubjective relations. In other words, by which mechanisms is a concept from the cognitive sciences applied to literary representation and with what results?

The process by which Theory of Mind makes the transition from theorized cognitive structure to literary analytic involves a damaging (or at the very least dangerous) consolidation of the set of hypotheses about cognition, sociality, and neurological development form which it emerged. The discipline crossing

mobility granted to Theory of Mind is primarily the result of its historical connection to the so-called Sally-Anne test, what's known in psychology more generally as a False-Belief Test. As Sedgwick explains, "[f]rom the term's very inception, the presence of Theory of Mind has been tested by whether or not the subject understands that somebody else could make an inference that the subject knew to be mistaken" (Sedgwick 2011, 150). The Sally-Anne test itself is significant for two interrelated reasons, the first of which pertains to its limitations as a diagnostic tool, while the second is a conceptual-theoretical concern that will ultimately return us to questions framing Sedgwick's essay. Here is the description of the Sally-Anne test Sedgwick provides, which is given by Jared Blackburn from a paper delivered at Autism Europe's 6th International Congress:

> In this test, the child gets to see two dolls, Sally and Anne. Sally puts a marble into the basket and then leaves the room. Anne enters the room, takes the marble our of the basket and puts it into a box. The child them gets asked where Sally would look for the marble when she comes back to the room. It is seen as a sign of ToM [Theory of Mind] if the child understands that Sally doesn't know what Anne knows, and therefore answers that Sally will look in the basket where she put the marble herself. (Sedgwick 2011, 150)

The Sally-Anne test functions as diagnostic for Theory of Mind by ostensibly testing if the subject can 'theorize' a difference between the minds of Sally and Anne, and thus answer that Sally will look for her marble in the basket. Individuals who answer in this way are said to have Theory of Mind, but they are presumed to understand deception. The 'commonsense' assumption that underlies the test is that a person's behavior is the result of their knowledge, and so in spite of Anne's deception Sally's behavior remains *logical*. To rephrase once again, Sally makes the wrong choice *because* Anne deceived her by hiding the marble. While further criticism of the Sally-Anne test might be valid, it remains accurate to observe that the test itself plays out in a fiction field of intersubjective relations among characters.

By situating the Sally-Anne test in this way its apparent suitability for literary application becomes clear,

yet this is only half of the problem. Stepping back from the close examination of the mechanics of Zunshine's method and taking stock of the arguments present by the book as a whole, Sedgwick delivers an incisive observation of the limitations to this reading practice:

> …I'm dismayed by how much the bulk of [Zunshine's] readings have to depend on that rackety warhorse of high-school English classes, the Unreliable Narrator….I don't want to attack Zunshine for being philistine or writing a popularizing book. However exquisite or abstruse she wanted to get, the problem is that her readings really do *have to* cleave to the most reductive version of the Unreliable Narrator problematic. The constraint lies in Theory of Mind itself. (Sedgwick 2011, 151)

Under Zunshine's labor the complex intersubjective relations enabled by the "knowledge that others have minds of their own" is streamlined and repackaged as "the knowledge that [others] are capable of ignorance, concealment, or misrepresentation, or of being stupid, crazy, or methodically vicious" (Sedgwick 2011, 150). Under this optic, just as in the Sally-Anne test, readers are delivered a logical template that correlates the knowledge and behavior of characters, however we given no access to how exactly a reliable narrator is constituted, "as though 'reliable' itself is a single thing to be, a single kind of normative transparency or relationality" (Sedgwick 2011, 151).

"This history of impoverished "Jane Austen" readings is
not the result of a failure on readers to "contextualize
historically…"
—Eve Kosofsky Sedgwick,
"Jane Austen and the Masturbating Girl"

Chapter 2

Cognition, Literary Pleasure, and Masturbation

This chapter develops and refines a strategy for literary analysis under the heading of 'reading otherwise,' drawn primarily from Eve Kosofsky Sedgwick's essay, "Jane Austen and The Masturbating Girl." In elaborating this method, I seek to challenge the presumption held by Cognitive Literary Studies that novels are best, or 'properly,' read by applying findings from cognitive science and psychology to literary artifacts and characters. This challenge proceeds by problematizing the conventional understanding of the relation of the reader to the text by tracing the significant-yet-forgotten dependence of the realistic novel's point-of-view upon clinical modes of observation. As this textual structure is revealed to exercise considerable influence over the interpretive outcomes of literary analysis, the need to develop modes for 'reading otherwise' becomes more urgently felt. The paper argues for a renewed consideration of the text itself and its specific effects as the medium of novelistic narrative and argues against attempts by cognitivist readings to render the text invisible or inconsequential by transforming it into a simple record of thought or thinking.

Suspending for the moment the questions raised thus far regarding the scientific basis (cognitive evolutionary psychology) of CLS, the task of evaluating the literary uses of the cognitive approach remains. While Zunshine frequently and happily asserts that our evolved cognitive adaptation for Mind-reading "makes literature as we know it possible," such observations supply few details of the interpretive or analytic payoff said to be made available by cognitive methods (Zunshine 2006, 10). If we entertain Zunshine's claim, we are entitled to ask what, specifically, does evolutionary cognitive psychology enable

the literary critic to accomplish? Once this question has been substantively answered it will then be possible to isolate and evaluate cognitive insights in terms of their *literary* value.

Here I will examine two articles by Zunshine which propose to answer these questions directly by way of the work of Jane Austen: "Why Jane Austen Was Different, And Why We May Need Cognitive Science to See It" (2007) and "Mind Plus: Sociocognitive Pleasures of Jane Austen's Novels" (2009) Alongside these articles, I will compare Zunshine's insights into Austen's novels to those made by Eve Kosofsky Sedgwick in her essay, "Jane Austen and the Masturbating Girl". I've selected these texts because their attention to the same literary figure serves as a control, allowing for a careful comparison of the results produced by cognitive reading strategies to the results of interpretive methods developed by feminist theory and gay and lesbian studies.[22] If Zunshine's "long-term objective is to make sure that a cognitive framework indeed offers something not available through more traditional literary-critical methodologies," our procedure will put this to the test (Zunshine 2007, 146).

Austen and Cognitive Science

Zunshine's 2007 article, "Why Jane Austen Was Different, and Why We May Need Cognitive Science to See It," opens by introducing a central claim of George Butte, Professor of English at Colorado College, that "[s]omething happened to the novel 'around the time of Jane Austen,'" when English authors "began to portray a multiply-layered and mutually-reflecting subjectivity" in novels (Zunshine 2007, 141). Drawn

[22] One of the more general questions circulating through this comparison of interpretive methods is the degree to which Cognitive Literary Studies presents a particular model of cognition as 'simply,' and thus falsely, natural and universal, and extends this 'neutral' objective construct into literary studies as structurally necessary, as the very infra-structure of narrative, homogenous and unitary. To the extent that this occupation of the literary by the cognitive remains unproblematic to scholars working within it, is set into relief by Sedgwick's description of the conditions motivating her and the other participants in MLA panel "to begin an exploration of literary aspects of autoeroticism." The choice "makes sense," she says, "because thinking about autoeroticism is beginning to seem a productive and necessary switch-point in thinking about the relations—historical as well as intrapsychic—between homo- and heteroeroticism: a project that has not seemed engaging or necessary to scholars who do not register the antiheterosexist pressure of gay and lesbian interrogation" (Sedgwick, 2003b, 110, 111-112). The point here being the need for queer theory to produce something other than simply "antiheterosexist" "interrogation," that is, not to rest its critical laurels in a defensive position *against* heterosexism but to develop a different contribution.

from Butte's 2004 book *I Know That You Know That I Know: Narrating Subjects from Moll Flanders to Marnie* that advances the uses of Merleau-Ponty's phenomenology for literary analysis, Butte names this [revolutionary] literary construct "*deep intersubjectivity"* and defines the narrative technique as "a new layering of human consciousness, or a new representation of those subjectivities as layered in a specific way" (Butte 4, quoted in Zunshine 2007, 141). Zunshine asserts that the "impact" of this narrative technique is "today" taken "for granted in the prose of George Eliot, Henry James, James Joyce, Virginia Woolf, Vladimir Nabokov, Ian McEwan, and others," thus presenting the figure of Jane Austen as the beginning of a sea change in narrative technique in literary history, and making her texts exemplary of (and exemplary because of) deep intersubjectivity.[23]

Zunshine contends that Butte's argument "supports the critical view that Austen was profoundly innovative in her treatment of fictional consciousnesses, a view that gets obscured when her novels are treated as archetypes of the genre," and that his argument regarding Austen's "mutually reflecting fictional subjectivities" demonstrates a congruence with "research in evolutionary psychology...on the cognitive challenges of processing multiple mental states embedded within each other" (Zunshine 2007, 142). Departing from Butte, Zunshine argues that Butte's work "provides a crucial first step for recognizing Austen's prose as actively experimenting with readers' cognition" in a manner fundamentally distinct from other eighteenth-century authors who preceded her. In what way does this experimentation with readers' cognition take place?

Subtly if bluntly authorizing the necessity of a cognitive framework for fictional analysis, Zunshine mandates that to "speak of mental states in works of fiction, we need to borrow from cognitive science the concept of *Theory of Mind*" (Zunshine 2007, 142, emphasis original). Introducing Theory of Mind alongside its more discursively amiable alternative nomination, "*mind-reading*," Zunshine reports (without

[23] "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'" (Zunshine 2007, 142).

remarking on the the difference between cognitive science and psychology) that "[c]ognitive *psychologists*" use this (or these) term(s) to "refer to our ability to explain observable behavior in terms of underlying thoughts, feelings, desires, and intentions" (Zunshine 2007, 142, my emphasis). Here two distinct elements are put into play while masquerading as one: the concept is rendered as a biological organ, and as an array of response patterns, a set of 'explanations' for observed behaviors, which comprise the epistemology of sociality. *Theory of Mind* is thus hypostatized as a cognitive "mechanism" [24]— which Zunshine models on a muscle as the ground of "our ability"— that collects within cognition all explanatory frameworks for observable behavior and unifies them under the rubric of "mind-reading".[25]

Baron-Cohen, gives a fuller picture of the choice of terminology in the closing two pages, under the heading "Modularity and Mindreading." With a calm knowingness Baron-Cohen acknowledges that "[s]ome people get very nervous when there is talk about modules." And he 'want[s] to say just a few words about this slippery concept, especially in relation to the mechanisms [he has] been describing" (56). Here, Baron-Cohen explains that the concept of 'modules' is largely the result of Jerry Fodor's "notion that the mind and the brain have modular organization" cf. *The Modularity of the Mind.* Summarizing a list of the attributes Fodor theorizes to define a module, Baron-Cohen focuses our attention on "only the last three tenets that are true of 'biological' modules'" in light of which "the systems [he] has been discussing" have been "referred to...as 'neurocognitive mechanisms,' rather than as 'modules' in the strict Fodorian sense" (57). This attention by Baron-Cohen to the biological plane establishes a relation between the Fodorian psychological module and Baron-Cohen's *descriptions* of four systems as "neurocognitive mechanism" "four mechanisms that might underlie the universal human capacity to mindread" (57, 31). While Baron-Cohen cites Fodor in this context as the relevant reference for the concept of a module, earlier, when 'qualifying' his linking of the 'biological' with the 'universal,' Baron-Cohen speaks exclusively of Steven Pinker, who

---

[24] Zunshine worries this term slightly, seemingly more of an echo of Baron-Cohen than one capable of producing some interesting complications for her work. After discussing Pinker's work, Baron-Cohen states plainly that "[o]ne key aim of evolutionary psychology is to describe the evolution of neurocognitive mechanisms" and thus we see the 'logic' required of the biological (Baron-Cohen 1997, 56)

[25] It remains an open question for me as to just how significant the double nomination of this concept is for the literary applications of the brain sciences. What's most worth noting is the fact that this term retains connections to discursive systems, and to writing specifically.

"using exactly the logic" that underwrites Baron-Cohen's method "adopted the evolutionary approach to language" argues that "language is a part of human biology" (10).

The mind-reading "ability" guaranteed by this muscle's *tonus*[26] is to execute automatic and unconscious explanations of "observable behavior," and these explanations, Zunshine contends, are themselves evolutionary adaptations to pressures exerted by natural selection upon us, the result of "the intensely social nature of our species" (Zunshine 2007, 143).

Keeping this in mind is important, because the scope of Zunshine's project oozes beyond a simply cognitive analysis of literature, for she intends to "demonstrate how a cognitive approach" to literary criticism and analysis "encourages us to see fictional narratives as engaging our *evolved* cognitive adaptations" (Zunshine 2007, 142, my emphasis). Thus, the rhetorical light-touch of encouragement that Zunshine's cognitive approach initially lends us, soon gives way (in the space of about two pages) to an authoritative categorical shove. The evolutionarily necessary cognitive adaptations that "may have developed during the 'massive neurocognitive evolution' that took place during the Pleistocene" are now confidently put forth as the basis of literature "as we know it" (142-143, 143). With passing mention of "the intriguing function of the so-called 'mirror neurons,'" Zunshine implies that a neuronal basis underwrites her conclusion that "[a]t least on some level, your brain does not seem to distinguish between your actions and a person you observe doing them" (143). Thus, we are left to intuit, the blindness of these so-called 'mirror neurons' to the difference between our own actions and observed ones, authorizes Zunshine's observation that "[l]iterature 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" (Zunshine 2007, 144).

This quick run-through of Zunshine's foundational premises has moved swiftly away from Austen and toward categorical claims about literature precisely because Zunshine's own analysis pursues a similar

---

[26] In physiology, medicine, and anatomy, muscle tone (residual muscle tension or tonus) is the continuous and passive partial contraction of the muscles, or the muscle's resistance to passive stretch during resting state.

trajectory. This observation should not provoke objections because, manifestly, that is the very purpose of her article, telegraphed by the title itself: Jane Austen's novels are different and to "see it" we "may need" cognitive science. They are different, that is, from those novels which came before her; because as we are reminded, the specific difference Austen *was*, is today taken for granted in the prose of modernity.

Given that Zunshine begins her article with George Butte's work on Austen's use of a narrative technique he calls *deep intersubjectivity*, the task remains to isolate the specifically cognitive (as distinct from the phenomenological). Here Zunshine introduces "one particular line" of research by evolutionary psychologist Robin Dunbar that focuses on "our ability to process multiply embedded states of mind," and which has "demonstrated that we have marked difficulties processing units of information that embed more than four recursive mental states" (Zunshine 2007, 144).[27] This is not the first time that Zunshine's analyses have utilized a rubric of mental "embedment" constructed from Dunbar's work,[28] as she subjects a scene from Virginia Woolf's *Mrs. Dalloway* to the same analysis of the first section in her book, *Why We Read Fiction*. The insight Zunshine draws there is that Woolf's novels along with, as she reiterates here, those of other modernist writers "experimented with our mind-reading capacities by pushing their portrayals of fictional subjectivity to the sixth level of recursive embedment" (Zunshine 2006, 144). *Modernist* fiction, according to Zunshine's cognitive approach, is distinguished by techniques of sixth-level

---

[27] "After the fourth level," Zunshine summarizes, "our understanding plummets sixty percent" (Zunshine 2006, 144).

[28] In the pages on Woolf from *Why We Read Fiction*, Zunshine's terminology is "embedded intentionality," which she links to the work of Daniel Dennett: "Dennett, who first discussed this recursiveness of the levels of intentionality in 1983, thought that it could be, in principle, infinite. A recent series of striking experiments reported by Dunbar and his colleagues have suggested, however, that our cognitive architecture may discourage the proliferation of cultural narratives that involve 'infinite' levels of intentionality" (28). The take away of this for Zunshine's present purposes is, thus: "Cognitive scientists knew that this 'failure on the mind-reading task [was] not simply a consequence of forgetting what happened, because subjects performed well on the memory-for-facts embedded into the mind-reading questions.' The results thus suggest that people have marked difficulties processing stories that involve mind-reading above the fourth-level" (28-29). In a review of Zunshine's book published in *Philosophy and Literature* from 2006, Brian Boyd draws attention to a problematic "fuzziness of the counting" exhibited in Zunshine's reading (or calculation?) of the passage (593). Examining her reading according to her "template of embedded of intentionality," Boyd declares it a misreading and thus a "fatal flaw" in her "second, implicit but no less central, claim: that analyzing fiction in terms of [Theory of Mind] offers criticism greater explanatory power, precision and clarity" (590).

recursive embedment. Constituting it categorically thus makes peculiarly 'visible' our/modern readers' "hunger" for a "cognitive workout," and on this analogy reading becomes 'exercise,' and novels a sort of cognitive step-aerobics class of escalating intensity.

Dialing back from Woolf to Austen, Zunshine distills from Butte's study an essential link between evolutionary cognitive psychology and literary history:

> Putting what [Butte] 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-forth level and the writer (i.e., Austen) who consummated this transition. (144-145)

Having previously outlined the general logic of cognitive evolutionary psychology, which serves as the foundational origin story of Theory of Mind, and having aligned this logic with the more local, "particular line" of Dunbar's research on the limitations to Theory of Mind, Zunshine seemingly coordinates 'cognitive theory' with 'literary history'; an evolutionary *why* with a narratological *how*. With Butte's [interpretation of the] data on the literary field of the eighteenth-century in hand, Zunshine plots his identification of Austen's innovation 'chronologically' in relation to her analysis of the scene from *Mrs. Dalloway*. On this view, the "eighteenth-century sentimental novel" portrayed fictional subjectivity as a "second-third level of mental embedment," with Austen's novels marking the portrayal of a "third-fourth level," while Woolf and her contemporary authors "push[ed] their portrayal of fictional subjectivity to the sixth level (148, 145, 144).

But what does this developmental narrative, all but explicitly stated, *do*? If Zunshine is, as she says, "excited by Butte's argument" because of its similarity to her own work: what does "putting what [Butte] does in the context of cognitive theory" accomplish for literary interpretive methods? In other words, what I am asking about is how we are to understand the congruence between phenomenological literary analysis

and a evolutionary-cognitive literary analysis; what does it signify or authorize? Are we to understand a determinant relation whereby the cognitive element makes possible the phenomenological? Is this a scene of validation, the congruence legitimating the 'traditional' vocabulary of phenomenology by affiliation with a cognitive framework? It is somewhat surprising that Zunshine's discussion of the "intriguing interpretive opportunities" opened-up by the "realization that we can 'calculate' the levels of fictional subjectivity" in a narrative is notably scant with [specifically cognitive] details:

> 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[29] to function comfortably one intersubjective level up, we can speculate about particular cultural circumstances behind this learning curve. Those may include the socioeconomics of the 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 consciousnesses, and so forth. Although we shall never be able to produce the exact and exhaustive list of these circumstances, 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-

[29] The subjunctive mood Zunshine uses here is not, it would seem, a simple slip of the tongue. Two pages further on, when again writing on this point of transition that Austen inaugurates, Zunshine says "Austen 'learned'[12] new ways" of treating fictional subjectivity, and explains in the note that "[she (Zunshine)] put this term in quotation marks because [she doesn't] want to claim that this process of learning is necessarily conscious (151, 151n12). This hesitation pops up consistently under the alternative guise of intuition; on "...a number of the late seventeenthcentury playwrights must have *intuitively* figured out this rule" (158, my emphasis), on the way her novels would be received, "Austen knew (*intuitively*, of course) that readers would try to downgrade the level" (159, my emphasis), on Austen's dislike for sentimental novels which "made her *intuitively* wary of the self-conscious sensibility cultivated by some earlier writers" (159, my emphasis), on the relation between cognitive constraints and craft "we discover...,there is a 'guarantee' of sorts that writers will *intuitively* experiment in the direction of challenging that constraint..." (160, my emphasis).

ranging effects of the deepening of fictional intersubjectivity. (Zunshine 2007, 146).

As it stands here, the cognitive approach appears a bit underwhelming, especially given that Butte was able to "'calculate"' the levels of fictional subjectivity in Austen's texts quite well with phenomenology alone. If cognitive evolutionary psychology simply "contributes...the awareness of challenges involved in processing multiply-embedded subjectivities" the value of such a contribution is rather ill-defined, given that according to gestalt of Zunshine's logic these complex fictional subjectivities would remain challenging to readers either way, with or without cognitive interpretive tools (Zunshine 2007, 146).

The Sentimental Novel and "greedy self-consciousness"

At the outset of this section I outlined the aim to evaluate Zunshine's "long-term objective": "to make sure that a cognitive framework indeed offers something not available through more traditional literary-critical methodologies" (Zunshine 2007, 146). From what we've seen thus far, the cognitive framework has, more or less, simply concurred with the literary-critical work done using these 'more traditional' methods of reading and textual analysis. If we turn now to the "immediate goal" of Zunshine's essay, "to understand how Austen's experimentation with deep intersubjectivity built upon that of her predecessors," it may be possible for us to identify the cognitive contribution and its literary-critical value.

Drawing on John Mullan's well known work on the eighteenth-century sentimental novel, *Sentiment and Sociability: The Language of Feeling in the Eighteenth Century*, Zunshine proceeds to amplify the cognitive transition she sees Austen's novels to mark in literary history. "Although we do not think of Austen's novels as sentimental," she writes, "[Austen] 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" (148). Without belaboring the point too much, the reason I've selected this moment for sustained

contemplation is that it nicely captures a switch-point that exists between fictional characters and persons which Zunshine's cognitive-interpretive machinery obscures to the degree that it succeeds in simplifying and reorganizing the differences within the texts she references around the diacritical distinction: Sentimental novel vs. Modern psychological novel. As we've already discussed above, the cognitive evolutionary psychology humming along quietly in the background of Zunshine's analysis requires that characters be formatted as persons. So formatted, a cognitive model (or, let's say, a 'Theory of Mind') may then be installed and characters interpreted correctly according to an evolutionary 'granted' rubric. Whatever the differences between Austen's novels and those of her predecessors, documented by Butte and Mullan among many others, are thus compressed and smoothed out by the evolutionary weight of Zunshine's cognitive interpretation.

To spell out what I mean by the 'switch-point that exists between fictional characters and persons' mentioned above, I want to introduce the point of comparison from Sterne's *Sentimental Journey*, which Zunshine presents to us as exemplary of what Austen *is not*. Here's Zunshine:

> 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." (Zunshine 2007, 148)

The transition from text and person at stake overall for Zunshine's cognitive interpretive framework is here encapsulated in a scene-of-reading given to us by Sterne's Yorick as he strolls through the streets of London. That is, Sterne presents the reader with a scene-of-reading(-and-translating), that among other things serves as instructions to the reader for *how to read*, both persons and characters (persons as characters). Demonstrating to the reader of his novels the very activity of reading done by his protagonist, Sterne's *prose* is doubly marked as *writing*. Indeed, Yorick emerges in this passage as a writer himself, a character who "could have fairly wrote down…twenty different dialogues…and sworn to" them.

The choice of this passage from Sterne matches up quite well with the specifications of the sentimental novel which Zunshine takes from Mullan, as a type of novel that "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)" (Zunshine 2007, 148). However, when Zunshine compares Sterne to Austen's writing, the interpretation she pushes overlooks the persistent discursive element in favor of the observational. When she writes that Austen's "*prose* seems to have internalized the assumption that some people make a careful study of the 'turns of looks and limbs,'" Zunshine overrides the fact of words in favor of (represented) "looks and limbs" (Zunshine 2007, 148, my emphasis). Differently put, the cognitive framework gives the impression that the significant difference between sentimental novels and Austen's is the degree to which novels call attention to themselves as writing, as words, as having-been-written, as novels at all. The sentimental novel—complete with its overwritten and overwrought characters contemplating of their own contemplation of others—is replaced by Austen's innovation, with a narrative 'unaware' of 'itself' as writing.

While I've chosen to emphasis the discursive element under development in Austen's work, Zunshine has confidently pressed on in the other direction; and following a quick canvas of Austen's 'heroines' from *Persuasion* to *Mansfield Park* to *Sense and Sensibility*, she announces 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" (Zunshine 2007, 149). Falling back upon

the "[s]tudies in Theory of Mind" already introduced in connection to George Butte's work, Zunshine now runs a quasi-counterfactual comparison of alternative "mappings" of the well-known scene at Molland Bakery from *Persuasion* in which Anne Elliot and Captain Wentworth encounter one another, in order to demonstrate the differential narrative effects produced when Austen "frees up one level of intentionality" (Zunshine 2007, 149). In the 'sentimentalized version' of the scene Zunshine adds a layer of self-reflection to the narrative mapping in the form of "Anne's awareness of her perceptive self," and uses italics to mark the levels of intentionality:

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

7. (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. (Zunshine 2007, 149; emphasis original)

Running this through the cognitive framework, Zunshine calculates that the second sentence "pushes us to the sixth level of mental embedment," thus nearing the upper limit of our evolved cognitive comprehension. Contrasting this with the first sentence mapping Austen's original scene which remains at the fourth or fifth level of embedment,[30] Zunshine then accounts for the qualitative difference between the two sentences:

> ...the only recompense that we get for our pains as we strive to grasp its overall meaning is being reminded of 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" (Zunshine 2007, 149; emphasis original).

[30] "This is the fourth level of embedment, reaching, perhaps, even further: up to the fifth level" (Zunshine 2007, 145)

I've reproduced these passages from Zunshine at length because they open up the compression enacted by the cognitive framework upon narrative, whereby fictional characters are hardened into persons. If Zunshine's gloss of 'our' reaction to this passage seems rather harsh on Anne—don't worry: the "cognitive perspective" clarifies why this is the case. For the cognitively challenging second mapping, "all that difficulty" began when Anne "smugly appreciated that she could keep track of who was thinking what and whose pleased self-awareness seemed particularly out of place..." (Zunshine 2007, 149-50). Yet, for this interpretation to hold we are required to overlook the false comparison between the *text* of the narrative and ourselves; that is, we must strip the textual-discursive element away and replace it with a person whose cognitive abilities outpace our own. On this view, Anne gets in *our* way, blocking our *view* of the scene as she refuses to yield her agency and be reduced to a set of Galilean binoculars.

In this zero-sum game of 'paranoid' cognitive structures we readers, says Zunshine, rage against Anne's "greedy self-consciousness" because 'she' takes two levels of mental embedment all for herself. This "sentimentalist self-consciousness" consumes too much of *our* cognitive resources, and thus is not "reader-and-character-friendly," producing protagonists who are either "smug" or "Machiavellian" (Zunshine 2007, 151). With seemingly only so much cognition to go around, Anne's over consumption disproportionately reshapes the narrative scene "making the heroine seem self-satisfied about her powers of penetration" (Zunshine 2007, 150). Recasting this 'sentimental effect' in the terms from Butte's title, Zunshine suggests that "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" (Zunshine 2007, 150).

Thus we have arrived at the nexus of cognitive and narrative structures which answers in large part the question posed in the title of Zunshine's essay. The cognitive approach provides greater insight into why Jane Austen was different, Zunshine argues, because it offers "a new way of approaching the familiar topic of Austen's complicated relationship with eighteenth-century sentimental novels" (Zunshine 2007,

159).[31] Although I'm condensing Zunshine's claims regarding the utility of a cognitive framework, this move is authorized by the essay's manifest purpose: to specify how Austen's novels inaugurated a transition in literary historical periodization by dispensing with the narrative techniques of the sentimental novel. And yet here's where the waters begin to muddy: the cognitive reading has not surpassed the other interpretive methods in analytic payoff. While Zunshine maintains that her insights into Austen "follow from appropriating concepts from cognitive scientists who study our mind-reading adaptations," the *necessity* of a cognitive framework remains an open question given the centrality of Butte's phenomenological reading of Austen to her own argument (Zunshine 2007, 159). Zunshine's mobilization of cognitive terminology and concepts such as mind-reading and Theory of Mind lends her argument an elevated, technical, and modern vocabulary, but this vocabulary is largely utilized to bypass or override textual questions in favor of a presumed cognitive content. Here, for example, is Zunshine's own summation of why exactly the cognitive approach was necessary:

> I can demonstrate that [eighteenth-century sentimental] 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 to a writer who has an allergy to the hyper-sensitive "I know that I know that you know that I know."] (Zunshine 2007, 159, emphasis original)

---

[31] Zunshine enumerates a total of three insights produced by the use of cognitive science: first, to explain why use of high levels of mental embedment is sometimes funny because its incomprehensibility creates a "comic effect" (which she sees active in an episode of the television show *Friends*) while in other cases its use "encourages us to identify emotionally with the heroines" from Jane Austen. This 'emotional identification' spoken of here is simply the narrative style of Austen, oddly rephrased in non-cognitive terminology; second, to explain why Austen "made sure that her deeply intersubjective passages cannot be successfully paraphrased" by "downgrading" the level of mental embedment; third, as stated above.

Plainly, the analytic thrust of this demonstration is heavily compromised to the degree that the insights remains soldered to Austen's 'intuitive' choices as a writer. While Zunshine speculates on Austen's intuition, the support she musters for this claim is decidedly *non*-cognitive but rather discursive, the evidence residing well within the words printed on the page.

Mind + Prose

In her article for *Studies in the Literary Imagination*, "Mind Plus: Sociocognitive Pleasures of Jane Austen's Novels," published two years after "Why Jane…" in 2009, Zunshine gives a somewhat expanded account of her previous argument for the necessity of a cognitive reading of Austen's novels. Importantly, "Mind Plus" dispenses with the task of placing Austen's novels in relation to either the works of her sentimental predecessors or to modern 'psychological' novels, and instead focuses almost exclusively on Austen's prose in order to refine and elaborate what Zunshine calls "Austen's trademark as a writer": the "strategy of imposing an extra mind onto a scene *after* the configuration of minds has apparently been completed" (Zunshine 2009, 105; emphasis original).

Derived from a scene from *Northanger Abby*, during the first meeting of Catherine Morland and Henry Tilney, and two scenes from *Mansfield Park*, the first describing the introduction of Henry and Mary Crawford to the Bertram household and the second in which Henry Crawford reads a passage from Shakespeare, Zunshine further refines her contention as to Austen's trademark by narrowly applying it to the construction of her novels' protagonists. For Austen's novels "whose protagonists are relatively 'clueless' (for example, Catherine Morland, Emma Woodhouse), the subjectivities of other people are often added on," but "in the novels whose heroines are more 'penetrating' (e.g., Fanny Price, Anne Elliot), their subjectivities are often brought in this way," that is, these heroines are added on (105). Condensing this and yoking it to the broader assertion of Austen's "trademark as a writer," Zunshine specifies, "[o]r, to put it differently, this is one way Austen's narrative constructs a perceptive heroine—she gets to be *the* mind

added to the seemingly completed scene" (Zunshine 2009, 105; emphasis original).[32]

The reorganization of her argument about Austen's novels around the question of the protagonist is a subtle yet highly significant shift, because it establishes an apparent primacy for questions of cognition and perception for literary interpretation at the expense of thinking through discursive elements and textual effects. That the grade of this "epistemological slide"[33] enacted by Zunshine upon Austen's work is sloped definitively toward the cognitive and perceptual should come as no surprise, but any consideration of the consequences of this realignment are suppressed as the cognitive register is presented as the fundamental basis of literary production. This is accomplished by a series of substitutions or equations of terms which continuously push a thematics of cognition, perception, observation and "seeing" forward as the foundation upon which Austen's work depends: subjectivity is equated with mind, mind is equated with a character's perspective (in both senses of the word), and observation and "mutual observation" are transformed into awareness.[34] With these substitutions in place the processes of literary interpretation are thereby soldered to 'our' cognitive processes. "Tracing these patterns of mutual awareness," Zunshine summarizes, "is integral to our interpretation of the characters, even if we never think about it in such terms" (Zunshine 2009, 104).

But how does this now familiar argument about character scale up to encompass "Jane Austen's Novels" *tout court* as the title of Zunshine's essay indicates? The priority of the protagonist for the wider considerations of narrative structure and its 'sociocognitive pleasures' is necessary, because the privileged

---

[32] Curiously, Zunshine makes a distinction between observing and seeing in a footnote to this sentence in order to mark her disagreement with Bharat Tandon's argument, given in *Jane Austen and the Morality of Conversation*, regarding the status of Fanny Price: "I thus have to disagree with Tandon's suggestive argument that, "Fanny forms only one apex of a triangle of peripheral observers and silent auditors, alongside her Aunt—and Pug' (215). Neither Lady Bertram nor Pug is shown to be able to reflect other minds; unlike Fanny, they may 'observe,' but they do not *see*" (121, emphasis original).

[33] This term is taken from Roland Barthes essay "From Work to Text," p. 155 in *Image, Music, Text*

[34] "The last sentence adds another subjectivity or, let us say, *mind* (that of Mr. Allen) to the configuration of the minds hitherto present in the chapter" (104, emphasis original); "…depending on where in the scene we decide to place Mr. Allen's observation of Henry, Catherine, and his wife, our perception of that moment will be adjusted. We become aware of another mind—another perspective of what's going on…" (104).

position of the protagonist among other characters is constructed, according to Zunshine, as a position of cognitive dominance and superiority. The process by which this dominance occurs is when a narrative scene presented to the reader is suddenly or subtly annexed by the protagonist after-the-fact as a scene of contemplation, dissolving the narrator and thus the narrative scene into the narrative seen by the protagonist.[35] Zunshine demonstrates this process of Austen's by way of two scenes from *Mansfield Park* wherein "Fanny [Price]'s subjectivity is added onto a seemingly completed scene" (Zunshine 2009, 107). What marks the narrative handoff in both scenes is Austen's use of the adverb 'meanwhile' or adverbial phrase 'all this while' to modify the reader's point-of-view, recuperating it as Fanny's own "through which what was said before now has to be re-perceived" (Zunshine 2009, 106).

First, during the scene of introduction of Henry and Mary Crawford to the Bertram household, Zunshine argues that the "narrative is brought to a screeching halt" as the narrator routes the scene through "the mind" of Fanny Price, forcing us "to rethink what came before from the point of view of another person" (106). Here's Zunshine:

> Deep in discussion of Mary's thoughts about Tom, the narrator suddenly stops short and asks, 'And Fanny, what was *she* doing and thinking all this while? and what was *her* opinion of the new-comers? (Zunshine 2009, 106; emphasis Austen's)

This sentence is then compared to a scene appearing much later in the book, depicting Henry Crawford reading a passage from Shakespeare for the Bertrams. Here, the reader initially receives the events through Edmund, constructing him as the position of narrative primacy because "it is through Edmund's subjectivity that other subjectivities are represented" (Zunshine 2009, 107). Yet Zunshine argues that cognitive (and thus, narrative) dominance remains with Fanny Price, as her "mind is superimposed on the scene, using almost the same construction as in the earlier passage" (Zunshine 2009, 107). Yet, while

[35] In discursive terms, what's happening here is the exploitation of a slippage between free and direct discourse, which is retroactively assigned to the protagonist.

previously the narrator *asked* after Fanny, now Fanny's thoughts are given plainly as simply the narrative's point-of-view, thus granting the reader access to Fanny's thoughts on the events observed by Edmund. Zunshine provides us with the decisive sentence in which "we learn that":

> Fanny, *meanwhile*, vexed with herself for not having been as motionless as she was speechless, and grieved to the heart to see Edmund's arrangements, was trying, by everything in the power of her modest gentle nature, to repulse Mr. Crawford. (Zunshine 2009, 106, emphasis Zunshine's)

At issue in this scene is the observation and explanation of Fanny's involuntary shake of her head, which Edmund (and thus the reader) observes to happen, as Zunshine puts it, "in response to something Henry [Crawford] says" to Fanny (Zunshine 2009, 107). This shake prompts Henry to move "closer to her, 'entreating to know [the] meaning' of her gesture," and, in turn, prompts Edmund to withdraw himself from the scene entirely (Zunshine 2009, 107; brackets Zunshine's). A great deal hinges on this moment for Zunshine's analysis, because by staging an interaction between characters while simultaneously denying the Edmund-the-reader, Edmund and the reader, access to that interaction, the narrative splits off from Edmund's position and prepares a point of return to be assumed by Fanny. The cognitive differential between Edmund and Fanny, that Edmund's knowledge of the meaning of Fanny's gesture is deficient compared to her own, becomes narrative dominance: "it is her mind and not Edmund's that gets the last word (so to speak) by being superimposed on top of everybody else's" (Zunshine 2009, 108).

At this point Zunshine's argument has successfully imbued all corners of Austen's novels with the cognitive structure of paranoia by firmly mooring the narrative within 'the mind' of the protagonist and rendering narrative ambiguity into a hierarchy of knowing. The "penetrating" heroine whose "mind" gets to be "on top of everybody else's" is unproblematically put forth as congruent with cognition *as we know it*. What remains highly questionable and potentially damaging is that Zunshine's argument operates by installing a partition within the fictional narrative that authorizes her to interpret [crafted] textual structures

as evolved cognitive strictures, "intuitively" chosen by Austen to "intensify" "the pleasure afforded by following various minds in fictional narratives" (Zunshine 2009, 111, 110).

In a telling footnote to the close of this section Zunshine effectively doubles down on the centrality of 'knowing' for literary interpretation and enjoyment. Acknowledging that the techniques Austen deploys in these "scenes could be broadly described in terms of Aristotelian 'reversal,'" such terminology now appears too blunt, too traditional (perhaps too textual?), as it does not account for every readers' 'known' cognitive mechanisms which "makes literature, as we know it, possible" (Zunshine 2009, 121n5, 110). Instead, due to our modern knowledge/science of cognition, Zunshine's footnote continues, "Austen's reversals are peculiar because they often seem to be concerned with mental competition among her characters: who get the last look and most inclusive view of the scene; which character's perspective we as readers would ultimately adapt as the most interesting and informative." The rich and varied non-cognitive interpretive pathways available in Austen's novels, many of which Zunshine initially draws from the work of other scholars, are thus here factored out to the lowest cognitive denominator.

A Pleasure of One's Own

Retrofitted with an epistemology of cognition, the narrative questions, opacities, uncertainties, tensions and ambiguities of the literary text[36] are rendered thoroughly *familiar* and knowable, they're nothing more than the mundane obstacles of getting to know other people, the commonsense of experience now backed with cognitive science. And this is precisely the desired outcome of the interpretive framework Zunshine deploys, yet once the question of *pleasure* is introduced we see that the cognitive framework depends crucially upon an equally known and sanitary notion of pleasure in order to be applied to literature.

[36] Latent connection here to the terminology of narrative "pressure/pressures" that Sedgwick introduces in "The Weather in Proust" p. 5.

But why pleasure? And, further still what *type* of pleasure and *whose* pleasure? The pertinency of these questions in 'traditional' literary critical interpretive projects underscores a stubborn quality of the text and readers' engagements with it, and revives [a range of] concerns I have been voicing throughout my evaluation of how Cognitive Literary Studies operates and the interpretative results it produces. As we've seen thus far, one objective of the Cognitive Literary Studies we've been looking at has been to argue that certain crucial/critical/significant aspects/properties of the literary text correspond to common cognitive patterns, patterns which have been forged into cognitive 'mechanisms' by evolutionary psychology, themselves modeled on psychological modules of cognitive processes.[37] I have detailed the consequences entailed by this approach to the literary text, while closely tracing how these correspondences are determined and located within the narrative. Certain elements have emerged as having primary importance for this project: the specification of genre and the position of a text in literary history (i.e., the Sentimental Novel, the psychological novel), the formalization of specific, new narrative techniques (i.e., deep intersubjectivity), and the integration of dramaturgical optics as an alternative epistemology to the introspective-confessionary model of prose available in literary discourse (i.e., the peculiarly "hypersensitive: I know that I know…").

To extend the cognitive approach—formulated here as a domesticated, familiarized paranoid knowing—to encompass the complex sphere of reader relations to the text, it remains necessary for Zunshine to establish access to the "sociocognitive pleasures of Jane Austen's novels," answering the question of motivation belied by the explanatory, titular proclamation of Zunshine's book: why we read fiction. Yet, by taking the nexus of literature and pleasure to be simply self-evident, that "[f]iguring out whose mind gets to be on top of other minds is a pleasant enough exercise," Zunshine still needs to account for this pleasure (presumably individuated) in terms of our evolved cognitive adaptations.

The term by which Zunshine brings pleasure into her analytic framework, and into relation with both

---

[37] For a useful discussion of this, see William Bechtel's article, "Modules, brain parts, and evolutionary psychology." Evolutionary Psychology. Springer, Boston, MA, 2003. 211-227.

the reader and the text is 'sociocognitive.' A relative new-comer to the array of cognitive terminology she has deployed thus far, this term is introduced by Zunshine in the process of explaining the instinct-like activity of "[o]ur cognitive adaptations for 'mind-reading'" (Zunshine 2009, 110). Citing an essay by Jess M. Bearing, "The Existential Theory of Mind," Zunshine provides the following linkage between the sphere of evolution and the sphere of literature in order to establish the evolutionary basis for the pleasure generated by the text:

> Our cognitive adaptations for "mind-reading" (another term for theory of mind) are promiscuous, voracious, and proactive; their very condition of being is a constant stimulation delivered either by direct interactions with other people or by imaginary approximations of such interactions. So important is the mind-reading ability for our species, and so ready is our theory of mind to jump into action and to subject every behavior to 'intense sociocognitive scrutiny' (Bering 12), that at least on some level we do not distinguish between attributing states of mind to real people an attributing them to fictional characters. (Zunshine 2009, 110)

This doubly footnoted passage serves as a buttress of sorts, fortifying Zunshine's argument for the [fundamental] centrality of theory of mind to the literary experience[38] by locating these cognitive processes on the species level and thereby removing them from willful control of the individual reader. Depicted here as "promiscuous, voracious, and proactive," these cognitive adaptations are presented as automatic, instinct-like mechanisms of response for 'reading' behavior, real or fictional. And yet not much has been said about pleasure

On this exact point, things begin to get shaky as the evolutionary 'why' is rendered concretely in literary

[38] cf. "While the question of whether our enjoyment of fiction is an adaptation of a byproduct of other adaptations (such as theory of mind) remains open, the centrality of mind reading to our engagement with fictional narratives continues to be borne out by ongoing studies in cognitive psychology" (111).

terms. Following the passage given above, Zunshine elaborates as follows: "Figuring out what Henry Tilney is thinking as he holds forth on muslins feels almost as important as figuring out what a real-life attractive stranger is thinking as she looks us in the eye and holds forth on how she enjoyed reading the book that we currently have in our hands" (110). Leaving the veritable lynch-pin of her argument implicit, it would appear that Zunshine contends that the pleasure motivating our literary engagements is analogous to the pleasure of being hit on, albeit done here in eighteenth-century pomp. The 'sociocognitive pleasure' of reading is that "almost as important" *feeling* we presumably have when discerning a pick-up line from casual conversation; if Catherine Morland is prospected, so too is the reader.

Or not quite. Zunshine qualifies this implicit claim by explicitly routing the readers' pleasure through a more sanitary level of textual engagement, whereby our sociocognitive pleasure in reading Jane Austen is experienced as the "satisfying confirmation that we remain competent players in the social game that is our life" (110). What remains problematic is that Zunshine's explication of the 'sociocognitive pleasure' that supposedly drives our engagement with literature teeters between a form of sexual excitation and a self-satisfied competence, producing an account of literary pleasure that has an unreassuringly ambiguous evolutionary standing: "The experience is pleasurable, but the adaptive reasons for this pleasure are completely under the radar of our consciousness" (Zunshine 2009, 110).

The "narrative's 'mind'"

Surprisingly, after introducing 'pleasure' to serve as a vehicle suitably flexible enough to gather the divergent strands of her argument, a 'sociocognitive pleasure' that comprises and amalgamates the social and cognitive varieties under a single term, Zunshine's analysis shifts quickly away from evolutionary theory to consider the more strictly narratological implications of the "sociocognitive dynamics" of a narrative scene, which she will now hypostatize as "the *mindfulness* of social situations" (111). Used loosely as a sort of diagnostic criterion, a scene's 'mindfulness' roughly correlates with the number of characters of which it is composed because each character increases the potential for "reflection" within the scene by

a factor of one.[39] Zunshine proposes that Austen and other writers "(intuitively, of course)" "intensify the pleasure of their readers" "by adding an extra mind to a seemingly completed scene." However, for this use of 'pleasure' Zunshine offers a different analog than the implied erotics alluded to before, substituting an "exercise metaphor" carried over from her book *Why We Read Fiction*. The addition, after the fact, of an "extra mind to a seemingly completed scene" now affects the reader much like an extra weight suddenly lobbed onto the weight bar: "As we adjust to a new perspective on the interplay of subjectivities that we have just witnessed," Zunshine writes, "...we 'work-out' in a focused way our theory of mind, stretching and feeling (so to speak) our powers of social reasoning" (Zunshine 2009, 111).

Yet this "exercise metaphor," in evoking a weight room setting, confuses the challenge of this activity with the experience of surprise which Zunshine is attempting to specify and capture as the effect of our cognitive constraints.[40] I argue that without venturing use of the term herself, what Zunshine designates as the reader's pleasure, is more accurately understood in terms of the interrelated experiences of *surprise* and *confirmation*. If this is the case, the more relevant questions to investigate are how the narrative is capable of generating the effect of surprise, and, further, why a cognitive analysis might not be equipped to identify it as such. Before proceeding to these concerns, we can attempt to clarify how 'pleasure' becomes an operative term.

How can we understand Zunshine's use of the term pleasure, despite the problematic aspects of its deployment we've already discussed? In other words, what is it in Zunshine's view that's pleasing to the reader about the *narrative*? The beginning of an answer is available in her consideration of Alex Woloch's work on character in *Pride and Prejudice*, which Zunshine draws from his study, *The One Vs. The Many: Minor Character and the Space of the Protagonist in the Novel.*

---

[39] This is assuming the characters to be 'Austenian,' that is, non-sentimental.

[40] In other words, what I'm saying is that the image of weight-lifting is rhetorically successful for 'explaining' the activity of ToM which remains invisible, because weight-lifting is available as a highly familiar, routinized and regular form of exercise. But because I'm saying that what Zunshine is talking about is 'surprise,' the weight-lifting metaphor falsifies the experience she's actually talking about by representing it with a 'programmatic' form of exercise; i.e., there's little that's unexpected in weight-lifting.

Woloch argues that Austen constructs Elizabeth Bennet as the novel's protagonist in two steps: first, Austen distinguishes Elizabeth as having a superior "ability to reflect other people's states of mind," and second, Austen "slowly integrates Elizabeth's thoughts with the narrative's own point of view" (Woloch 78, quoted in Zunshine 112). The import of Woloch's argument regarding Elizabeth and *Pride and Prejudice* is that Zunshine sees it to be congruent with her interpretation of Fanny's role as protagonist in *Mansfield Park*. Both characters are shown to represent minds of other characters, and these qualitatively different "thoughts" of the protagonists are integrated, argues Zunshine piggy-backing on Woloch's phrasing, "with the narrative's own point of view" (112). This comparison of Fanny and Elizabeth is not given as confirmation of Woloch's argument *per se*, as Zunshine is at pains to demonstrate that her "sociocognitive approach to Austen does not replace or contradict other established literary-critical approaches" but is instead "complementary" (Zunshine 2009, 112).

Yet, this complementary relation is complicated and undermined when Zunshine contends that "the cognitive perspective clarifies how the process of the integration of the protagonist with the omniscient narrator, described by Woloch, works" (Zunshine 2009, 112). Zunshine's introduction of the "omniscient narrator" into the present discussion occurs almost by fiat, flatly claiming the authority for this transition on the basis of Woloch's work: "if we can speak of the 'narrative's own point of view,' as Woloch does," she writes, "than we can also speak of the narrative's 'mind' and say that it is the mind capable of representing, or *embedding*...all other complex mind-embeddings in the novel" (112). The rhetorical payoff of personifying "the narrative's point of view" (Woloch) and reformulating it as "the narrative's mind" is that Zunshine can now speak about the narrative *as if* it were a character's mind, structurally analogous to the all minds 'embedded' within it.[41] This enables Zunshine to specify that the process of integration

[41] cf. Page 117, where Zunshine will work same equation of mind and character in the opposite direction, driving an "implied character" from the narrative's "state of mind," rather than deriving "mind" from "point of view": "In other words, a fictional narrative may conjure up a state of mind that cannot be traced to *any* character, embedding that state of mind with a shared mind of several physically present characters and with that of an implied character, such as the narrator."

"described by Woloch"[42] works—works, that is, by means of cognition: "The protagonist, such as Fanny or Elizabeth, may at times come closer than anybody else in the novel to being able to embed as many embedded mental states as the omniscient narrator is able to embed" (Zunshine 2009, 112).

What emerges in Zunshine's description is a certain so-to-speak 'cognitive privilege' which the protagonist enjoys, and which, in turn, differentiates this character position from the other characters in the novel. What is slightly obscured by figuring these character positions as persons is that Zunshine's gloss of Woloch's work installs a precise structure of relation between the Woloch's protagonist and Zunshine's omniscient narrator. Viewed from a textual perspective, the relation described is one of an asymptotic dissymmetry knowledge between the protagonist and the narrative itself; in other words, in terms of knowing, the protagonist is delimited only by the narrative as a whole, to which only the reader has full access. Braided throughout the layers of Zunshine's cognitive embeddings is hierarchy of 'sameness,' the protagonist is closest to the 'omniscient narrator,' with whom the reader is most closely identified. Zunshine leaves implicit the cognitive threshold differentiating the reader from the 'omniscient narrator,' because by rendering these positions inside and outside the text 'cognitively' indistinguishable, the text's resistance to being read—to being *known*—is eliminated. It is necessary for Zunshine's argument that the reader be positioned *as* the omniscient narrator, because it is only from this all-knowing position that the reader is enabled to experience the pleasure (i.e., the sociocognitive pleasure) that Zunshine contends drives the reading of fiction.

In different ways the complicated nexus of pleasure and cognition has been at the root of Zunshine's cognitive framework and has remained unexamined because it has been disguised in ways that render its operation familiar and commonsensical: represented obliquely as a cognitive mechanism, legitimated through the lenses of evolutionary processes, gestured to seemingly without need of further explanation,

[42] I'm quoting Zunshine saying this because it's unclear to me to what extent she's projecting this claim onto Woloch, given the fact that Zunshine initiates the transition from "narrative's point of view" to "narrative's mind" it seems to be overstating things to assert these processes to be identical/the same.

invoked and evoked, connoted and denoted: pleasure remains the motive force that powers the machinery of Zunshine's literary fMRI. Seen from this angle, Zunshine's choice to coordinate her cognitive approach with work by Butte and Woloch (and Vermeule) appears less complementary and rather installs the cognitive perspective in the position of *ex post facto* confirmation. From charactological to narratological to epistemological, Zunshine's analysis traces over a multivalent notion of pleasure with cognitive terms of awareness and attention, formalizing this pleasure as an object of perception (for both characters and readers), and thereby producing a narrative 'pleasure' that is as hygenic, clean, and empirical as the science of cognition that underpins it. A final example nicely demonstrates the comforting, seemingly explanatory effect this process yields while alienating any substantial analytic purchase.

Immediately following her distillation of Woloch's work into cognitive particulates, Zunshine proposes an alternative consideration of the well-worn categorical "distinction between 'flat' and 'round' characters," inspired by Blakey Vermeule's "closely related exploration of narrative consciousness" (Zunshine 2009, 112). In the text referenced, Vermeule's own concern is located at a more general level,[43] with the narrative effects produced by the formal and psychological *disparity* between these character categories and the interactions that result because of this disparity. Combining Vermeule's comments with her gloss of Woloch's observations regarding Austen's protagonists, Zunshine appears to isolate an interrelation of pleasure and cognition throughout Austen's novels which marks "[g]enuinely compatible romantic partners" (114).

Pace Vermeule, Zunshine contends that the development of a "cognitive parity" between characters is "one way in which Austen makes her readers feel that some characters are particularly compatible as romantic partners or friends, not at all compatible, or not-yet-compatible" (113). The survey-esque phrasing notwithstanding, the argument being made here is that Austen's characters are shown to be in love—that is, genuinely "belong together," despite whatever marital arrangements actually happen in the

[43] "In literary narratives from ancient to modern times,…" (Vermeule, quoted in Zunshine 2009, 113)

novels—by means of a mutually comparable "ability to entertain multiple-embedded mental states" (114, 113). Zunshine proffers the observation that Austen's characters who are genuinely compatible with one another "exhibit more parity in their capacity for representing the complex mental states of others" than those characters "who do not belong together" (Zunshine 2009, 114, 113). On this view, the reader is supposed to recognize this relation (intuitively, perhaps?) beyond the events of the text, so that "even if [the incompatible characters] do end up married, *we know* that they can never be truly happy together" (113, my emphasis). Thus Zunshine argues that we know—that the reader *knows* true romantic love as a recognizable cognitive parity, and that it is this knowing that is *pleasurable*. By staging this 'knowing relation' between characters, "Austen *makes us feel*" their compatibility, and it pleases us to know their love is *real*, that is, to feel its reality despite the text: "Still, if toward the end of the story, we are convinced that Anne and Wentworth belong together, their emergent cognitive parity is one of several psychological strategies that Austen intuitively employs to *make us feel that conviction*" (Zunshine 2009, 113 and 114; my emphasis).

What is it, precisely, that Zunshine is showing us here? "Genuinely compatible romantic partners," or should we rather say genuine, heterosexual love? By inserting the word 'genuine' into the discussion it is clear that Zunshine uses this word to demarcate a sphere that stands outside the narrative artifice. The reader's expected recognition of "[g]enuinely compatible romantic partners" as different from "not at all compatible" romantic partners indicates a quality that somehow resists textual manipulation and narrative encasement: even if these characters do marry, we know they will never be happy because their love is not genuine and *we feel it*. Initially seemingly mundane, the dependence of Zunshine's cognitive framework upon pleasure is revealed to be no small matter when it is recalled that "Jane Austen's Novels" are widely considered to trade almost exclusively in depictions of romantic, heterosexual courtship; and that the particular *novelistic form* given to these depictions by Austen, as Zunshine herself has previously argued, inaugurated a change within the history of the novel so different that "We May Need Cognitive Science to See It." If we accept that 'pleasure' drives cognition and thus reading to the degree that Zunshine's system requires, we must consider the particularity of that pleasure: is all reading pleasure the same? Is pleasure

the only drive of the cognitive reading process?

Alongside Zunshine's cognitive approach to literaturem which we've show to rely heavily upon a notion of knowing pleasure entailed by the readers' evolutionarily adapted cognitive mechanisms, I propose an alternative model for thinking through the cognitive and affective pressures generated by the text upon the reader developed from the work of Eve Kosofsky Sedgwick. As we have seen, Zunshine's cognitive approach mobilizes cognition and the cognitive processes of knowing by supplementing these processes with pleasure, and her choice of Austen's texts as exemplifying these cognitive processes enacts and disguises a circularity whereby the text's staging of heterosexual, romantic pleasure is 'discovered' to be—is *recognized* as—the very model of reading-pleasure. What remains highly problematic in this account is the multivalent purposes served by a singular notion of 'pleasure,' which is said to be produced by the reader's cognitive consumption of the narrative and, simultaneously, to be the very thing consumed. The evolutionary basis of Zunshine's analysis implicates the nexus of pleasure and cognition as the evolutionary outcome of the species level struggle for survival, thus locating the cognitive processes and mechanisms of both readers and writers beyond the text.

Yet, if pleasure is constitutive of the cognitive processes of reading *and* produced in the reader by these processes of cognition, Zunshine's use of pleasure in fact withholds explanation rather than granting the promised clarity of its familiarity. This problem—which is at the heart of cognitive literary reading models—is not an unintended 'side effect' but instead results precisely *because of* the analogical foundation of these cognitive projects that equates 'real life' with the literary text.[44] So structured, the cognitive approach Zunshine elaborates in fact stifles exploration of textual pleasures and affects to the

[44] Anticipating criticism on this point, Zunshine acknowledges that "[p]rofessors of literature continue to find it troubling that their students treat literary characters as real people," but then bypasses serious consideration of her colleagues' concerns, and gesturing instead to an implicit consensus among sociologists studying reading practices as if to concede the issue settled, she continues: "whereas sociologists who study practices of reading outside the academic setting continue to find evidence that such a 'misidentification' of fictional characters is a crucial aspect of any reading experience. As Elizabeth Long observes in her study of women's book clubs, readers' 'knowledge about literary characters can have the same certainty as their experiential knowledge of other people' (156)" (121n7). What remains overlooked in this observation is the degree to which "experiential knowledge of other people" can be equally faulty as poor readings of literary works.

corresponding degree that its analytic purchase depends upon unitary, self-evident and familiar image of pleasure.

Indeed, all the passages from Austen's novels selected by Zunshine to develop her analysis of fiction writ-large operate by trafficking a presumed familiarity of real-life, interpersonal interactions into the literary text via a cognitive 'bottleneck'. On this view, the characters populating a narrative are figured and sorted according to their cognitive 'abilities' and, like tiny reservoirs of 'awareness,' the reader consumes their knowledge as he consumes the narrative itself. As this knowledge accumulates, it produces pleasure in the reader, satisfied with his epistemological superiority, relishing in his omniscience, protected from reprisals for his penetrating voyeurism and vicarious appropriation of another's pleasure as his own.

The reliance of Zunshine's model upon a presumed familiarity of pleasure, of its easily recognizable quality and self-affirming feedback effect hypothesized as taking place within the reader, offers a compelling analytical elegance. And yet acknowledging this also lays bare the still open question of how unfamiliar pleasures circulate and organize a text. Differently put, if the cognitive approach must be primed by a minimal familiarity of pleasure in order to kick-start its interpretive operation (or prior even to interpretation, the act of reading), such a model is unable to discover or even discern pleasures not already known in advance. Zunshine's transposition of a familiar pleasure, namely heterosexual courtship, into the evolutionary-cognitive register of the human species falsely naturalizes and consolidates the multiple and disparate pleasures in the text. If, as seems likely, other forms of pleasure operate beyond the grasp of a cognitive analysis, how might a fuller examination of this possibility complicate, or even destabilize the architectonic congruence of the cognitive to the literary, for which Zunshine argues?

In responding to these questions, I write to challenge the seamless incorporation of Austen as an emblem of cognitive reading protocols and to elaborate an alternative interpretive approach, a mode of 'reading-otherwise'. I develop this account from two of Sedgwick's essays, "Jane Austen and the Masturbating Girl" and "The Weather in Proust," in order to trace an analytic trajectory implicit in Sedgwick's queer readings of canonical texts that articulate a shift away from 'known' models of pleasure

and sex.[45] My interest in these essays is fixed upon Sedgwick's double readings in each instance; that is, her ability to precipitate critical questions regarding traditional or habituated interpretive protocols through readings grounded in literary texts. By enlarging our view of the strangeness of a literary text, Sedgwick's analyses amplify literary-interpretive questions about the operations of pleasures/affects between characters—and between reader and text—necessitating that different critical tools be brought to bear. Her use of object-relations psychoanalysis and the theory of affects formulated by Silvan Tomkins, I argue, open up these resources for wider use in revising and reformulating critical theoretical reading practices.

In her notorious essay, "Jane Austen and the Masturbating Girl," first published in 1991 in *Critical Inquiry,* Eve Kosofsky Sedgwick advances a highly sophisticated reading of Jane Austen's *Sense and Sensibility* that raises significant objections to uncritical uses of pleasure in literary interpretive projects. I will follow Sedgwick's intervention here closely along two specific lines, namely her discussion of masturbation's peculiar resistance to historicization and her use of passages from a clinical case study from 1881 for what it implies about the specificities of literary writing, 'writing in the general sense' as distinct from real world interactions.[46]

Circulated by journalists critical of the curriculum of the Humanities as "an index of depravity in academe," Sedgwick's essay initiates an exploration of the interconnections between "literary pleasure, critical self-scrutiny, and autoeroticism" in order to approach what she describes as the vast, "indeed foundational, open secret about how hard it is to circumscribe the vibrations of the highly relational but,

---

[45] These two essays are most relevant in the context of the issues brought forth by Zunshine's arguments, but the set of Sedgwick's essays I see as participating in developing this line of thought also includes "The Privilege of Unknowing: Diderot's *The Nun*", "Tales of the Avunculate: *The Importance of Being Earnest*", "Is the Rectum Straight?: Identification and Identity in *The Wings of the Dove*", "Shame, Theatricality, and Queer Performativity: Henry James's *The Art of the Novel*". These 'demonstrations' or 'articulations' of modes of reading-otherwise are, to my mind, supplemented by the more critical-theoretically aimed essays that braid together Silvan Tomkins's affect theory with paradigms drawn from the object relations psychoanalysis developed by Melanie Klein, such as, "Shame in the Cybernetic Fold: Reading Silvan Tomkins", "Paranoid Reading, Reparative Reading; or, You're So Paranoid, You Probably Think This Essay Is About You", and "Melanie Klein and the Difference Affect Makes".

[46] In other words, what I'm trying to outline here is that Sedgwick's analyses demonstrate or attest to the intricacies of writing, specifically with regard to the circulatory paths of pleasure and affect within a text and the turbulence generated by them on the structure of the written text.

in practical terms, solitary pleasure and adventure of writing itself" (820). More than simple window dressing, Sedgwick cites the prodigal itinerary of her essay's title though the various accounts of "the degeneracy of academic discourse in the humanities" in order to demonstrate that the performative force of the title as an indictment of the Humanities depends in no small part on its self-evident status.[47] Journalists have merely to mention this phrase as evidence of degeneracy itself. This present day usage is quite telling for Sedgwick's purposes: that the phrase "Jane Austen and the Masturbating Girl" can stand independently from the essay it titles as signally representative of the Humanities-gone-bad testifies to the endurance of "[t]he narrative link between masturbation and degeneracy" into our modern era. In drawing our attention to the peculiarly "self-evident" performative charge of this "narrative link" between masturbation and degeneracy, Sedgwick notes that this link is constituted on the basis of "pre-1920s medical and racial science" that "no longer has any respectable currency" (819). How then are we to account for masturbation's encore role as self-evident indictment?

What sustains "the self-evidence of a polemical link" between autoeroticism and narratives of wholesale degeneracy" is the new position accorded to masturbation in the twentieth century (819). Now "conclusively subsumed under that normalizing developmental model," autoeroticism is widely understood as "representing a relatively innocuous way station on the road to a 'full,' that is, alloerotic, adult genitality defined almost exclusively by gender of object choice" (825). Albeit differently valued than in pre-1920s medical literature, but no less demeaning for it, this new "body of expertise" continues to support the use of 'masturbation' as a figure that both trivializes, infantilizes, and renders comic whatever activity to which it is applied. It is within this multiple semantic field, Sedgwick notes, that "the literal-minded and (at least by intention) censorious metaphor 'mental masturbation'" functions to describe "any criticism one doesn't like, or doesn't understand" (820).

---

[47] See "Against Epistemology," particularly when reflecting on the essay, Sedgwick writes: "Most of my energy in 'Jane Austen and the Masturbating Girl' was devoted to experimenting with structures and rhetorics that I thought might suggest some new sight lines and relations around the performativity of sex, of literature, of theory" (134).

With a methodological gesture analogous to her more extensive analyses of literary texts seen in *Epistemology of the Closet*, Sedgwick's purpose here is not to adjudicate but rather to identify the logic of an 'open secret,' by the operation of contradictory definitional structures. Once dangerous, but now infantile, trivial, laughable, irrelevant, and edging solipsistic; masturbation designates a diverse and sometimes contradictory semantic field. Such is the rhetorical calculus employed by the Op-ed authors to knowingly sneer—from within the safety of public sentiment—at the prospect that 'masturbation' should be given place at the convention for the Modern Language Association. Yet as we've learned from *Epistemology of the Closet*, what this 'open secret' structure marks is the potentially significant connection between "literary pleasure, critical self-scrutiny, and autoeroticism," and it is this rich, overlapping space of multiple meanings that, Sedgwick argues, belies just "how hard it is to circumscribe the vibrations of the highly relational but, in practical terms, solitary pleasure and adventure of writing itself" (820).

Subtly shifting her weight here, Sedgwick's articulation makes clear that the scope of her critique not only includes "writing itself" as the very words on the page, but also targets the established/traditional modes of relation to the text, and the critical or personal/solitary pleasures produced. I contend that this element of Sedgwick's thinking—the relation to writing, to text—has been overlooked in the wider reception of her critical interventions, but nonetheless deploys an incisive analysis of 'writing in the general sense' as elaborated in the work Jacques Derrida.[48] Although Sedgwick signals her intention to focus critical attention on writing early in the piece, the exposition of it arrives late and is given in an anecdotal tone which muffles its theoretic import. The essay's architecture is highly pertinent because it performs (or re-enacts) the textual effects Sedgwick experienced, textual effects which now animate her analysis performing them upon Sedgwick's readers. Sedgwick constructs her analysis of *Sense and Sensibility* by placing Austen's text alongside passages quoted from a contemporaneous document, described as, "a

[48] "Signature, Event, Context" provides a uniquely compact, if dense, outline of the structural aspects of writing, of written marks in relation to speech, that, I argue, Sedgwick mobilizes in this essay through her juxtaposition of Austen's *Sense and Sensibility* and Demetrius Zambaco's "Onanism and Nervous Disorders in Two Little Girls".

narrative structured as a case history of 'Onanism and Nervous Disorders in Two Little Girls' and dated 1881".

What this yields is a form of 'critical collage' or experimental writing, interlacing these passages in her essay with little more than voice-over narration jump-cuts, such as "More from the manuscript dated 1881:" (827, 829).[49] On one level, Sedgwick utilizes the document as evidence of the forgotten complexes that once comprised the now-obsolete medicalized typology of The Masturbator; and on another, uses it to demonstrate a highly compelling overlap between the behaviors documented and those of Marianne Dashwood. Yet in the final pages of the essay Sedgwick includes a brief account of her motivations for teaching *Sense and Sensibility* alongside this document in her graduate class, and thereby conveys a crucial insight about textuality that, I argue, deploys/manipulates the iterative structures/effects of text detailed by the Derridian deconstruction of writing. Distancing herself from "the new historicist's positivist alibi for perpetuating and disseminating the shock of the violent narratives in which they trade," by placing historical documents alongside fictional works, what emerges as particularly determinate for Sedgwick is the specifics of the 1881 document's format; that is, the published text itself:

> But the pretext of the real was austerely withheld by the informal, perhaps only superficially, sensationalistic *Semiotext(e)* format, which refused to proffer the legitimating scholarly apparatus that would give any reader the assurance of "knowing" whether the original of this document was to be looked for in an actual nineteenth-century psychiatric archive or, alternatively and every bit as credibly, in a manuscript of pornographic fiction dating from any time—any time including the present—in the intervening century. (835)

The "the informal, perhaps only superficially, sensationalistic *Semiotext(e)* format," Sedgwick identifies here is elucidated in the footnote accompanying her initial quotation from "Case History". "In quoting from

[49] cf. "To quote again from the document dated 1881:" (828); "More from the manuscript dated 1881:" (829); "Back to 1881:"

this piece," she writes, "I have silently corrected some obvious typographical errors; since this issue of *Semiotext(e)* is printed entirely in capital letters, and with commas and periods of indistinguishable shape, I have also had to make some guesses about sentence division and punctuation" (827, n16). Considered together the peculiarly uniform typographical format and the absence of bibliographic or contextual information withholds "the pretext of the real" from the reader, and thereby distinguishes the shock-effect Sedgwick is attending to from that of New Historicism she previously identified. Indeed, crucial to her argument is *that* the document titled "Case History" (as a piece of writing shorn of all referential supports of time and place) continues to operate *as if*—not simply as if it were real, but *simultaneously* as if it were fiction. On the textual level, "Case History" emerges as no less pornographic than the other pieces included alongside it in the *Semiotext(e)* volume, one staging of sexuality and discipline among others. "And wasn't that part of the shock?," Sedgwick writes, further specifying the point under scrutiny, "—the total plausibility either way of the same masturbatory narrative, the same pruriently cool clinical gaze at it and violating hands and instruments on it, even (one might add) further along the chain, the same assimilability of it to the pseudo-distantiating relish of sophisticated contemporary projects of critique." (825).

To fully grasp the logic underlying Sedgwick's argument, and the implications her insight holds for questions of literary interpretation and critical habits, we must first take into account *how* the two components of her analysis are co-assembled. Specifically, Sedgwick identifies a structural congruence between "Case History" and *Sense and Sensibility,* according to which the character of Marianne Dashwood and the characters of the anonymous girls presented in "Case History" are aligned by common passions, symptoms, and behaviors. "Marianne Dashwood," Sedgwick writes, "though highly intelligent, exhibits the classic consciousness-symptoms noted by Samuel Tissot in 1758, including 'the impairment of memory and the sense,' 'inability to confine the attention,' and 'an air of distraction, embarrassment and stupidity'" (827).

In identifying these points of contact, Sedgwick's purpose is to re-introduce contemporary readers of *Sense and Sensibility* to the elements by which the "sexual identity" of The Masturbator was constructed

prior to its disappearance *as* an identity. This re-introduction of a forgotten 'sexual identity' that circulated widely during Austen's time destabilizes and interrupts the critical consensus around *Sense and Sensibility* which fixes the novel as a mundane heterosexual marriage plot, "a dryly static tableau of discrete moralized portraits, poised antitheses, and exemplarily, deplorable, or regrettably necessary punishments, in an ascetic heterosexualizing context" (835).[50] Before explaining in more detail why Sedgwick understands the reintroduction to necessarily reshape readings of *Sense and Sensibility* (and literary texts more generally), I will first elaborate how the modern concept of 'sexual identity' imposes a significantly problematic influence on the novel, much as it does for our contemporary understandings of sexual identities and the relations between them.

Sexual Identities, Familiarity, Character Behaviors

With a nod to work in gay and lesbian studies on the "historicization of sexuality following from the work of Michel Foucault," Sedgwick notes the increasing involvement of these projects with "issues of representation" and that "different varieties of sexual experience and identity are being discovered to possess a diachronic history...and to be entangled in particularly indicative ways with aspects of epistemology and of literary creation and reception" (820). Yet within this field of study autoeroticism, unlike heterosexuality and homosexuality, offers a distinct "challenge to the historicizing impulse" because "it escapes both the narrative of reproduction and (when practiced solo) even the creation of any interpersonal trace," and thus appears to "have an affinity with amnesia, repetition or the repetition-compulsion, and ahistorical or history-rupturing rhetorics of sublimity" (820).

In the present day context of hierarchically oppressive relations between genders and between sexualities, Sedgwick points out that beyond being particularly well suited as an analogy to writing,

[50] "Austen criticism is notable mostly not just for its timidity and banality but for its unresting exaction of the spectacle of a Girl Being Taught a Lesson—for the vengefulness it vents on the heroines whom it purports to love, and whom, perhaps, it does" (833).

masturbation "can seem to offer...a reservoir of potentially utopian metaphors and energies for independence, self-possession, and a rapture that may owe relatively little to political or interpersonal abjection" (821). For projects of sexual activism, masturbation and the historical vicissitudes of masturbation phobia present huge value as a tool of critical leverage for queer people "against the crushing mass of legitimated discourses showing us to be moribund, mutant, pathetic, virulent, or impossible" (821). Indeed because this "long-execrated form of sexuality, intimately and invaluably entangled with the physical, emotional, and intellectual adventures of many, many people,...today completely *fails* to constitute anything remotely like a minority identity," the use of autoeroticism for projects of literary interpretation would be resistant to the contemporary politically charged rhetorics of sexual liberation.

Selecting one of the very few bedroom scenes from Jane Austen's novels overall, this "particularly devastating" scene from *Sense and Sensibility* opens when Elinor Dashwood awakens to find her sister, Marianne, writing a letter to Willoughby, her might-have-been suitor, that is, when her uncontrollable sobbing and agitation permit her a steady hand. Alarmed by her sister's condition, Elinor's attempted inquiry into the cause of Marianne's distress is immediately rebuffed, while she struggles to complete her letter against the "frequent bursts of grief which still obliged her, at intervals, to withhold her pen" from the paper (822-3).[51] Noting that the absent but named Willoughby marks this "depicted bedroom space of same-sex tenderness, secrecy, longing, and frustration," Sedgwick contends that the novel must be considered as both hetero- and homoerotic: "No doubt it must be said to be both, if love is vectored toward an object and Elinor's here flies toward Marianne, Marianne's in turn toward Willoughby" (823). Readers familiar with the larger narrative of *Sense and Sensibility* will recognize Sedgwick's observations are undeniably true, but posing this question—pursuing this mode of reading—is soon exposed to be highly problematic. Immediately Sedgwick asks the following question: "But what, if love is defined only by its

[51] The alternation of sobbing with the writing activity itself seems to figure the interpretive 'hunch' that Sedgwick is pursuing here, that the metonimic association of writing, masturbation, and autoeroticism may offer "metaphors and energies for independence, self-possession, and a rapture that may owe relatively little to political or interpersonal abjection" (821).

gender of object-choice, are we to make of Marianne's terrible isolation in this scene; of her unstaunchable emission, convulsive and intransitive; and of the writing activity with which it wrenchingly alternates?" (823).

Before suggesting an answer, Sedgwick doubles back, uncomfortable with the anachronism of ascribing either hetero or homosexual identities (certainly female ones) to the Dashwoods, which she rightly points out "were not supposed to have had a broad circulation until later in the nineteenth century" (823). Beyond the chronological issue of availability, Foucault's work further demonstrates that the "conceptual amalgam represented in the very term *sexual identity*, [as] the cementing of every issue of individuality, filiation, truth, and utterance *to* some representational metonymy of the genital," post dates Austen by a half- or three-quarters century" (823). What this means, Sedgwick writes, "is that the genital implication in either 'homosexual' *or* 'heterosexual,' to the degree that it differs from a plot of the procreative or dynastic...may mark also the possibility of an anachronistic gap" (823).

Differently put, in 1811 what it can mean to identify either character as homo or heterosexual is not bound to the contemporary dominance of the genital, but instead is radically expansive with alternative relational possibilities. Inspired by Paula Bennett's work on Emily Dickinson (Austen's near contemporary), Sedgwick explains that Bennett's accomplishment is to have offered "a model for understanding the bedrock, quotidian, sometimes very sexually fraught, female homosocial networks in relation to the more visible and spectacularized, more narratable, but less intimate, heterosexual plots of pre-twentieth century Anglo-American culture" (824). What is tremendously exciting about the proposition of sexual identity with radically expansive meaning are the implications this insight holds for our contemporary moment. Once we have recognized the historical character of the concept 'sexual identity,' we are enabled to think critically about the "emergence and cultural entailments of 'sexual identity' itself during this period of the incipience of 'sexual identity' in its (still incompletely interrogated) modern senses" (824). Under what set of circumstances and which pressures did this concept, so fundamental to the organization of present-day life, crystallize and precipitate?

> Indeed, one of the motives for this project is to denaturalize any presumptive understanding of the relation of 'hetero' to 'homo' as modern sexual identities—the presumption, for instance, of their symmetry, their mutual impermeability, or even of their both functioning as 'sexual identities' in the same sense; the presumption, as well, that 'hetero' and 'homo,' even with the possible addition of 'bi,' do efficiently and additively divide up the universe of sexual orientation. It seems likely to me that in Austen's time *as in our own*, the specification of any distinct 'sexual identity' magnetized and reoriented in new ways the heterogeneous erotic and epistemological energies of everyone in its social vicinity, without at the same time either adequating or descriptively exhausting those energies. (824)

Outlined here is the anesthetizing and damaging effect that 'familiarity' exercises upon critical thought, and the thrust of Sedgwick's critique precisely targets analyses such as Zunshine's, which are unconcerned with how pleasure is identified, ascribed, and circulates through literary artifacts. Zunshine's contention that processes of cognition such as Theory of Mind "make possible literature as we know it" now appears newly problematic precisely for its knowing confidence.

Although Austen did not have homo- and heterosexual identities yet available to her, the 'sexual identity' that did exist as an identity was that of the onanist, 'the masturbator,' which, Sedgwick notes, was "already bringing a specific genital practice into dense compaction with issues of consciousness, truth, pedagogy, and confession" (825). As just one of the many sexual dimensions overridden by the "world-historical homo/hetero cleavage" in the last century, the distinction between alloerotic and the autoerotic identities no longer holds credibility as a persisting minority identity in the form of 'the masturbator.' This choice has been consolidated within the broader normalizing developmental model, so that today masturbation appears as an otherwise mundane precursor to "a 'full,' that is, alloerotic, adult genitality defined almost exclusively by gender of object-choice" (825). Alongside the many other forms of proscribed and regulated sexual identities that had developed by the end of the nineteenth century,

Sedgwick points out that "even the most canonical late-Victorian art and literature are full of sadomasochistic, pederastic and pedophilic, necrophilic, as well as autoerotic images and preoccupations" (826). Admittedly one of many identities to be "subsumed, erased, or overridden in this triumph of the heterosexist homo/hetero calculus," the masturbator's status, Sedgwick argues, was in fact "uniquely formative" (826). As, she writes:

> One of the earliest embodiments of 'sexual identity' in the period of progressive epistemological overloading of sexuality, the masturbator may have been at the cynosural center of a remapping of individual identity, will, attention, and privacy along modern lines that the reign of 'sexuality' and its generic concomitant in the novel and in the novelistic point-of-view, now lead us to take for granted. (826)

With this in mind, Sedgwick's worrying of *Sense and Sensibility* with which we began appears trace a fundamentally different problem. All but invisible to contemporary readers, the question of the autoerotic in *Sense and Sensibility* as a text that is "at once germinal and abjected, in the Austen canon and hence in the 'history of the novel" takes on a renewed importance—as does the need for formulating modes of reading otherwise. Here's Sedgwick: "[I]n our reimaginings of the history of sexuality 'as' (we vainly imagine) 'we know it,' through readings of classic texts, the dropping out of sight of the autoerotic term is also part of what falsely naturalizes the heterosexist imposition of these books, disguising both the rich, conflictual erotic complication of a homoerotic matrix not yet crystallized in terms of 'sexual identity,' and the violence of the heterosexist definition finally carved out of these plots" (826).

With an abrupt jump-cut Sedgwick neatly returns us to the bedroom scene of *Sense and Sensibility* but not before laying out what will be the first of many citations to a what she terms "a narrative structured as a case history," ascribed to Demetrius Zambaco and published by *Semiotext(e)* in 1981, itself dated 1881, titled "Onanism and Nervous Disorders in Two Little Girls". Here Sedgwick makes a crucial intervention that not only will challenge the mundane and widely accepted reading of the Dashwoods' central role in

the novel's marriage plot, but also thereby will challenge familiar and domesticated understandings of sexual identity on the basis of the ascription and legibility of pleasure within this (or any) literary narrative.

This intervention proceeds in double fashion—feeding-back and feeding-forward simultaneously—by targeting the contemporary concept of 'sexual identity' as a presumptively allo-erotic relational structure necessarily organized around genital pleasure.[52] As Sedgwick's consideration of the question of Marianne Dashwood's 'sexual identity' reveals an unaccounted for auto-erotic element, the figuration and deployment of the auto-erotic element within Austen's narrative is proven to be illegible within the conceptual framework that underpins the contemporary taxonomy of possible sexual identities as exclusively allo-erotic and genitally organized. Feeding-forward, the conceptual instability of 'sexual identity' that Sedgwick reveals to be at play in the narrative of *Sense and Sensibility* now destabilizes the sense of familiarity promised by that very concept; that is, the familiarity of knowing *how* the relations between self and others are organized around pleasure, according to a genital rubric.

The conclusion that emerges is that auto-erotic, non-genitally organized pleasure is a central component of the novel's plot as figured through the character of Marianne Dashwood, and Sedgwick argues that the reader's access to the vicissitudes of this pleasure is granted via the frame of discipline and the implantation of desire granted by Elinor's tutelary role in the novel. Through the sometimes-anguished, sometimes-loving but always constant gaze of Elinor and her consistent reflection upon her sister Marianne's situation, Sedgwick demonstrates that Austen's text deploys and constructs a narrative-cum-

---

[52] cf. Specifying that the link between masturbation and degeneracy exists in "pre-1920s medical and racial science" Sedgwick goes on to demonstrate that the case is not yet fully settled: "To the contrary: modern views of masturbation tend to place it firmly in the framework of optimistic, hygienic narratives of all-too-normative individual development" (819). Sedgwick's reference in the opening section of the essay to Jane E. Brody's recent "Personal Health" column in *The New York Times* is highly pertinent. In this column Brody "reassures her readers that, according to experts, it is actually possible for people to be healthy *without* masturbating;" thereby disproving the implied concern that masturbation is an *essential* developmental stage to allo-erotic adult sexuality. Noting the slippage now bestowed upon masturbation's developmental role—once essential, now superfluous—Sedgwick outlines the resulting status of masturbation: "But [masturbation] now draws on this body of expertise under the more acceptable gloss of the modern, trivializing, hygienic-developmental discourse, according to which autoeroticism not only is funny...but also must be relegated to the inarticulable space of a (barely superseded) infantility" (819).

case-history for the reader's complicit but invisible delectation.[53] Fundamental to this insight is the position of 'the reader' herself, which as I've shown above, Sedgwick reenacts and underlines throughout the Austen essay, the reader to whom the "Case-History" is undecidably (and thus simultaneously) a disciplinary pornographic fiction and a written report of clinical observations.

Genital Implication, Or the Lack thereof

At this point, an easy objection to raise against Sedgwick's argument turns around the question of the "genital implication" that forms the core of "the conceptual amalgam represented in the very term *sexual identity*," an implication which no shortage of Austen's contemporary readers would find wanting for Marianne (823). 'How is it,' critics may ask, 'that Marianne Dashwood can be seen—can be *read*—as the masturbating girl when she is never shown engaging her genitals in masturbation?' Differently put but perhaps more relevant for our purposes, we might ask what underwrites the connection Sedgwick draws/sees between more-and-less explicitly pornographic "Case-History" and *Sense and Sensibility*? Yet it is precisely on this question of sexual identity—that is, of genital implication—that Sedgwick's analysis turns to the Dashwood sisters' bedroom scene and considers the implications that a (mis)interpretation of this scene holds for novel as a whole, and literary criticism generally. After first noting how rarefied bedroom scenes are in Austen's novels, Sedgwick quotes *Sense and Sensibility* at length, which I reproduce here:

> Before the house-maid had lit their fire the next day, or the sun gained any power over a cold, gloomy morning in January, Marianne, only half-dressed, was kneeling against one of the window-seats for the sake of all the little light she could command from it, and

[53] Her recognition of this structure is consonant with her analysis in *Epistemology of the Closet* of the narrative optics in both "Billy Budd" as well as her powerful analysis of the construction of the sexuality of Proust's narrator in "Proust and the Spectacle of the Closet".

> writing as fast as a continual flow of tears would permit her. In this situation, Elinor, roused from sleep by her agitation and sobs, first perceived her; and after observing her for a few moments with silent anxiety, said, in a tone of the most considerate gentleness,
>
> "Marianne, may I ask?—"
>
> "No, Elinor," she replied, "ask nothing; you will soon know all." The sort of desperate calmness with which this was said, lasted no longer than while she spoke, and was immediately followed by a re- turn of the same excessive affliction. It was some minutes before she could go on with her letter, and the frequent bursts of grief which still obliged her, at intervals, to withhold her pen, were proofs enough of her feeling how more than probable it was that she was writing for the last time to Willoughby.[54]

Taking account of "who is in this *bedroom*," Sedgwick underlines the obvious fact that it is populated by "two women." "They are Elinor and Marianne Dashwood," she continues, "they are sisters, and the passion and perturbation of their love for each other is, at the very least, the backbone of this powerful novel" (823). Signalling an analytic shift to the narrative register and her return to the questions of sexuality and sexual identity, Sedgwick prompts her readers with two further questions: "But who is in this *bedroom scene*? And, to put it vulgarly, what's their scene?" (emphasis original, 823). Here, Sedgwick's use of italics to underscore the distinction at issue between *'bedroom'* and '*bedroom scene*' indicates that much more than the number of people in the room is at stake. With the analytic weight now concentrated by the additional specification of the narrative 'scene,' Sedgwick's response to her own rhetorical question is marked by its attention to the deconstructive, destabilizing effects of writing.

That is, this bedroom scene is also a *scene of writing*, and it is through taking account of this fact that her reading develops: "It is the naming of a man, the absent Willoughby, that *both* marks this as an unmistakably

---

[54] Jane Austen, *Sense and Sensibility* (Harmondsworth, 1969), p. 193, quoted in Sedgwick (CI) pp. 822-23.

sexual scene, and by the same gesture seems to displace its 'sexuality' from the depicted bedroom space of same-sex tenderness, secrecy, longing, and frustration" (my emphasis, 823). What Sedgwick observes here is the overlapping of three scenes, a "sexual scene" and a "bedroom space," joined in Austen's text by the inclusion of Marianne's activity of letter *writing*, which simultaneously *marks and displaces* the question of the scene's sexuality. The consequence of this overlapping of scenes is that it disrupts one's ability to cleanly answer the very question raised by this "bedroom scene": "Is this, then, a hetero- or homoerotic novel (or moment in a novel)?" To which Sedgwick responds accordingly: "No doubt, it must be said to be *both*, if love is vectored toward an object and Elinor's here flies toward Marianne, Marianne's in turn toward Willoughby" (my emphasis, 823).

Yet asserting the polysemy of this bedroom scene fails to address its most pivotal aspect, the scene of writing around which the the narrative is wrapped. While a cursory reading of Sedgwick's analysis may give the impression that her purpose here, as in many other queer literary theorists' projects, is to reveal a homosexual love story at work, if encrypted, within a canonical text of Western Literature, more careful attention to her thinking demonstrates her to be in pursuit of a very different goal.[55] I contend that in answering "both" to the question of the bedroom scene's sexuality is not Sedgwick's verification or attempt to set the historical record straight, instead she gives this answer in order to set the question aside and prepare to address a different problem. Instead, the problem at issue is the emerging insufficiency of the notion of love as "defined only by its gender of object-choice" to account for the composition of the scene at hand: "But what, if love is defined only by its gender of object-choice, are we to make," Sedgwick asks, "of Marianne's terrible isolation in this scene; of her unstanchable emission, convulsive and intransitive;

---

[55] Sedgwick announces this goal explicitly at the outset of the essay "The three participants in "The Muse of Masturbation," like most of the other scholars I know of who think and write about masturbation, have been active in lesbian and gay as well as in feminist studies. This makes sense because thinking about autoeroticism is beginning to seem a productive and necessary switchpoint in thinking about the relations- historical as well as intrapsychic- between homo- and heteroeroticism: a project that has not seemed engaging or necessary to scholars who do not register the anti-heterosexist pressure of gay and lesbian interrogation. Additionally, it is through gay and lesbian studies that the skills for a project of historicizing any sexuality have developed; along with a tradition of valuing nonprocreative forms of creativity and pleasure; a history of being suspicious of the tendentious functioning of open secrets; and a politically urgent tropism toward the gaily and, if necessary, the defiantly explicit. (821)"

and of the writing activity with which it wrenchingly alternates?" (823). Striking a tone that seems slightly facetious, Sedgwick's question is in fact quite earnest and even rigorously self-critical, a preliminary outline of the elements of Austen's bedroom scene that remain unaccounted for.

Backing off from the queer literary-critical impulse to adjudicate the 'sexual identity' of these and other characters is of a piece with what we might call Sedgwick's critical 'wide-view' of the entanglement of queer theory and literary criticism, because the ascription of a 'sexual identity' to these characters is both necessarily anachronistic as well as inadvertently complicit in furthering a "presumptive understanding of the relation of 'hetero' to 'homo' as modern sexual identities" (824). Posing the "homo/hetero question" to Austen's text asks it to respond in the yet-to-be formulated terminology of 'modern sexual identities,' or we might rather say, *sexualized* identities which Sedgwick characterizes as a "conceptual amalgam" (823). With reference to Foucault's history of discursive structures that congealed to produce this new concept in *History of Sexuality, Vol I*, Sedgwick notes that the modern concept of 'sexual identity' is composed by the "cementing every issue of individuality, filiation, truth, and utterance *to* some representational metonymy of the genital" (emphasis original, 823). The primacy of this 'genital implication' to the modern taxonomy of sexual identities ('homo/hetero') thus renders the question itself an ill fit for Austen's characters.

Given this, if we consider again Austen's bedroom scene, Sedgwick's accomplishment is to re-center attention on to the narrative as an alternative to literalizing the characters as people: "to the degree that [the genital implication in either 'homosexual' or 'heterosexual'" desire] differs from a plot of the procreative or dynastic," this difference marks the possibility of an anachronistic gap" (823). Refusing a shift in register by sticking to narrative question of plot, Sedgwick avoids rendering these characters as humans-who-have-a-sexual-identity, and addresses them as textual constructs responding to historically 'local' conditions. The incoherent erotics shown circulating around the writing activity of the Bedroom Scene mark the plot differing from the genital implication of homo- or hetero-sexual identity. The assertion that Marianne's character bears the 'sexual identity' of The Masturbator, is not genitally figured but is

shown instead through all the disorders of attention and autoconsumption, associated with this degenerate identity.

If what "defines 'sexual identity' is the impaction of epistemological issues around the core of a genital possibility" Sedgwick here moves to specify that what is critical about "Case History," and anti-onanist discourse generally, is their compulsive interest in "disorders *of* attention" (827). Reiterated with reference to Samuel Tissot, physician and author of the eighteenth century medical account of onanism, Sedgwick discerns that Marianne Dashwood indeed displays all of the "classic consciousness symptoms" he tabulates in 1758; in fact, Marianne's attention or inattention comprises a "surprising amount of the narrative tension of *Sense and Sensibility"* (827). Detailing Marianne's symptoms, Sedgwick's success is to demonstrate that the figuration of Marianne's autoerotic sexual identity proceeds primarily through non-genital terms. Nonetheless, Marianne's autoerotic sexuality exercises a powerful attraction upon almost every character in the novel: "The vision of a certain autoerotic closure, absentation, self-sufficiency in Marianne is radiantly attractive to almost everyone, female and male, who views her;" and yet this 'sexuality' while obscure to us is simultaneously legible to the other characters in the novel "through contemporaneous discourses as a horrifying staging of autoconsumption" (829).

The paradoxical outcome of Marianne's sexual identity for the narrative's structure is that her undisciplined and uncontrollable sensibility becomes the focal point, what in the wake of Barthes we might call the *punctum*, of the novel's controlling character, her sister Elinor, whose "unwavering but difficult love" for her sister forms the novel's erotic axis (824). As it is Elinor's subjectivity that is made most consistently available to the reader, it is her "self-imposed obligation to offer social countenance to the restless, insulting, magnetic, and dangerous abstraction of her sister [that] constitutes most of the plot of the novel" (830). More than this, Elinor's love "creates both the consciousness and the privacy of the novel," and her constant concern for Marianne, Sedgwick writes, "hollows out a subjectivity for Elinor and the novel that might best be described in the 1980s jargon of 'co-dependency' were not the pathologizing stigma of that term belied by the fact that, at least as far as this novel is concerned, the co-

dependent subjectivity simply *is subjectivity*" (830).

This destabilization of the status 'written-oral' has
not just always been an organizing theme for me,
but first of all—and these things are indissociable—
the very element of my work.
—Jacques Derrida, "Paper or Me, You Know…"

Chapter 3

Genette and Fodor, Derrida and Sedgwick

This chapter braids together the strands of argument I've been developing in the previous sections by tracking a thematic or conceptual pair of critic practices in Sedgwick's work, which I've referred to as 'reading otherwise' and 'thinking otherwise.' The interconnection between these practices is the organizing force of Sedgwick's feminist and gay affirmative/anti-homophobic analyses, and I now propose to follow her lead in order to articulate an alternative approach the question of literary writing and thinking. My position is informed by the deconstruction of writing elaborated by Jacques Derrida and suggests that the critical reading practices Sedgwick formulates around the theory of affect help clarify how to read literary writing and the 'destabilizing effects of writing' that operate within it. I am arguing against the ahistorical, timeless status assigned to literature by Cognitive Literary Theorists whose methodologies rely on bringing all forms of reading back to the foundation of the evolution of human cognition. With Sedgwick and Derrida, I propose a method of 'reading otherwise' that acknowledges the situatedness of 'literature' within a variety of historical and institutional contexts that do not determine what literature is, but instead open up a plurality of ways of reading that cannot be known or predicted in advance of the reading of reading.

As a framework for literary criticism and analysis, the strategies Sedgwick develops and deploys for 'reading otherwise' share a common motivation, namely, to multiply the interpretive possibilities and meanings in a literary text by carefully reading for articulations of difference. Such multiplication is the result of Sedgwick's deliberate return to canonical texts, reading them with a conceptual framework that has been either dismissed, forgotten, or thought to be rendered obsolete by the slow advance of history.

As we have seen in her reading of Jane Austen, these interpretive strategies respond to aspects of the text, including its historical time of production, the history of its critical, scholarly and popular reception, as well as current critical interpretations, that have been left unaddressed or unnoticed. Such critical blindness is characteristic of the limitations to Cognitive Literary Studies, and through Sedgwick and Derrida I will advance an outline of 'reading otherwise' that gives serious consideration of the textual problems Cognitive Literary Studies claims to resolve. All of these aspects of a text are eliminated, neglected, or ignored by the interpretive framework of cognitive literary studies. The interpretive framework cognitive literary studies deploys deliberately functions to consolidate interpretive possibilities by restricting allowable interpretations through the application of models of cognition drawn from cognitive science and psychology. Indeed, this is the explicitly stated mission of Cognitive Literary Theorists, to correct or move past interpretive techniques and concepts developed by structuralists and deconstruction. In positioning themselves in opposition to these critical and interpretive strategies, Cognitive Literary Studies seeks to arrogate to itself the status of the correct or true interpretive model by demonstrating that the questions posed and problems developed by deconstruction do not properly adhere in literary texts. These questions are, according to the account of the proponents of Cognitive Literary Studies, rather the invention of the critics themselves and the product of structuralist, post-structuralist methodologies, but are resolved or rendered moot through recourse to cognitive science.[56]

After identifying the component strategies assembled under the umbrella of Cognitive Literary Studies and by comparison to the reading strategies advanced by Sedgwick in her reading of Jane Austen, we can discern an additional premise, a presupposition of a shared consensus on the status of writing and of the relation of writing to speech. This relation, which privileges speech and deprecates writing, is the same relation identified and elaborated by Derrida throughout the body of his work. The suppression of writing

[56] Characteristic and representative of this attitude, see Alan Richardson's "Literature and the Cognitive Revolution" in which he writes: "A spreading dissatisfaction with the more bleakly relativistic and antihumanist strands of poststructuralism has given a new urgency to the groundbreaking efforts of these and other literary critics to forge a 'new interdisciplinarity'" (2).

and concomitant prioritization of speech are familiar to almost anyone acquainted with the critical and philosophical activities grouped under and around the heading of Deconstruction. With the arrival of Cognitive Literary Studies, we are able to discern a new layer to this deprecation of writing, that is, the explicit use of autism as a negative definition of the cognitive mechanisms that are alleged by Zunshine to make literature possible. Specifically in Derrida's work, from his earliest critiques of phenomenology to later analyses of speech act theory, the questions set in motion by the sustained attention to the effects of writing have been shown to destabilize and mount a significant challenge to philosophical knowledges.

Derrida's thinking on literature and fiction, never formalized in like fashion as other schools of literary criticism (psychoanalytic, new historicist, etc.) and packaged as an interpretive 'method' or 'technique', has routinely been received with confusion and frustration, leading quickly to disregard and ridicule by those scholars and literary critics who initially saw Derrida's 'deconstruction' of the history of philosophy to be extremely attractive for its critiques and analytic insight. Literary scholars critical of Derrida's readings of literary texts frequently point to the fact that many of the texts he examines all works of high modernism, implying (or outright arguing) that Derrida's identification of the destabilizing effects of writing point to nothing other than a deliberate choice by these authors, a stylistic feature common to modernist texts. In so doing, these critics attempt to walk a middle line, declining to fully dispute Derrida's readings but segregating these concerns to a specific genre and period of literary history. Sure, the question of authorship and authorial intention may be relevantly challenged when reading Mallarmé, such critics concede, but these questions are not applicable to more 'conventional' works of literary realism by authors such as Jane Austen or Daniel Defoe. We all *know* what these novels are about, they would contend, these authors are not up to any of those tricks played by modern authors, there is no need to over-complicate these straightforward stories.

The main fulcrum around which the analytic lever of Cognitive Literary Studies proposes to turn is the foundational assumption that language and the linguistic models of syntax that form the basis of cognitive psychology provide the basis on which literary narratives are governed by the same structures.

Throughout the readings proposed, the ready alliance of language and cognition is underwritten by the computational theory of mental representation that, as Fodor will elaborate further ahead, leverages a linguistic model of representation to resolve the philosophical problems that have plagued *philosophical* psychology. The theory of computation advances cognitive psychology in precisely this regard.

Prepared for in the introduction, Fodor's reading of Hume's cognitive psychology is valuable for two reasons, first, for Fodor's analysis of the irresolvable philosophical problems posed by Hume's "image theory" of mental representation, and second, for the analytic pressure Fodor's elaboration of the computational theory of mind puts on philosophical accounts of the mind in the "Wittgenstein tradition," in which in he includes "Dewey, Quine, and Ryle." As we will see below in more detail, Fodor's preoccupation with these philosophical challenges to the computational theory of cognition cluster around Wittgenstein's figure, unified by a common principle. "I take the idea that *content is interpretability in context*," Fodor writes, "to be paradigmatically (neo)Wittgensteinian" (my emphasis, Fodor 2003, 97).

Fodor's rejection of these "(neo)Wittgensteinian" critiques of the computational theory of cognition on the basis of "the idea that content is interpretability in context," responds to the implication that the meaning of a linguistic unit is dependent on its context, thus destabilizing the syntactic laws governing computational processing. Fodor's concern is to maintain the integrity of his cognitive psychology by demonstrating that the equivocating effect of context *in language*, in fact, does not apply to cognition. At the end of a long analysis, Fodor concludes, that if ever we are backed into an unresolvable question of language: "Context typically serves to resolve the equivocation by making clear which it is that he has in mind. (And, if the context doesn't, you can always just ask him.)" (Fodor 2003, 104).

Fodor's elimination of the problem of equivocation through the instantiation of an always available cognitive context may serve to resolve the effect of equivocation at the level of cognition, but its dependence on the presence of intention does not resolve the effect of equivocation at the level of writing. As Derrida has shown throughout his career, but most usefully in his essay "Signature Event Context," the destabilizing effect of writing (specifically fictional writing) is due to the structural relation of writing

to "the absence of every empirically determinable subject" (Derrida 1972, 315). Taking this into account shows that the linguistic model that differentiates Fodor's computational psychology, supposedly insulating it from the "essential drifting" of meaning, is not possible "due to writing as an interactive structure cut off from all absolute responsibility, from *consciousness* as the authority of the last analysis" (Derrida 1972, 316).

Now a touchstone in the history of Derrida's work, the essay "Signature Event Context," and the critique of "speech act" theory developed through a reading of J.L. Austin's *How to Do Things With Words*, gives unique access to an overlapping conceptual map, a 'hinge' by which cognitive psychology is connected to the philosophical movement associated with "ordinary language" and closely associated to Wittgenstein, that proposes to derive a theory of language from "ordinary" or "everyday" uses of language, a theory of language as communication, and organized by the representational model of communication or communicating a message.[57] Delivered in 1971, the title page of the essay notes that it is "A communication to the Congrès international des Sociétés de philosophie de langue française, Montreal, August 1971. The theme of the colloquium was "Communication" (Derrida 1972, 307). The significance of this time period and theme, in connection with my own work undertaken here, is its close proximity to the explosive developments in technologies of electronic communication, beginning with Claude E. Shannon's establishment of information theory in 1948 with the publication of "A Mathematical Theory of Communication" in which he articulated a formal theory of communication that is widely accepted as establishing the fundamental basis on which electronic computer systems are based[58]. In the opening

[57] In the opening sentences of the essay, it appears that Derrida is making an allusion to certain concepts and premises found in information theory when he raises this obstacle to the communication theory of language: "Is it certain that there corresponds to the word *communication* a unique, univocal concept, a concept that can be rigorously grasped and transmitted: a communicable concept?" (309)

[58] Shannon's influence on the development of the electronic age cannot be understated, beginning as far back as his graduate thesis from MIT, his work elaborated solutions to a fundamental layer of engineering and theoretical questions in proposing how to utilize electronic circuits to solve problems articulated in Boolean logic. His contributions are pivotal in the history of the computer, in decisively demonstrating how electronic circuitry and digital computers could be built to implement computational processes.

sentences of Derrida's essay, the allusion to certain concepts and premises of information theory resonate in his articulation of the obstacles Derrida sees in the communication theory of language:

> Is it certain that there corresponds to the word *communication* a unique, univocal concept, a concept that can be rigorously grasped and *transmitted* [my emphasis]: a communicable concept? Following a strange figure of discourse, one first must ask whether the word or signifier "communication" communicates a *determined content* [my emphasis], an identifiable meaning, a *describable value* [my emphasis]. (Derrida 1972, 309)

The use of "transmitted," "determined content," and "describable value" each invoke a strong connection to the vocabulary of technical specifications for quantifying information as well as the standards of the system of communication theorized and developed by Shannon.[59] What may not be so immediately clear from the opening lines of Derrida's own communication, is that the issue he raises regarding the meaning of the word 'communication' is not simply semantics or philosophical game-playing. The concept of communication that organizes and governs ordinary language philosophy derives from a highly specific, highly technical formulation, one that foregrounds precisely, as the requirements of the its system, that the message be of "a determined content, an identifiable meaning, a describable value" (309). If it can be shown that "*communication* has several meanings, and if this plurality could not be reduced," as is the message in the technical theory of communication, "then from the outset it would not be justified to define communication *itself* as the transmission of a meaning" (309). Plainly stated, the word *communication*

---

[59] Further confirmation of this can be read later on, when under the subheading of "Writing and Telecommunication" Derrida will describe what in Shannon's theory is identified as the 'channel,' medium through which the message is transmitted: "To say that writing extends the field and the powers of locutionary or gestural communication presupposes, does it not, a sort of homogeneous space of communication? Of course the compass of voice or of gesture would encounter therein a factual limit, an empir-ical boundary of space and of time; while writing, in the same time and in the same space, would be capable of relaxing those limits and of opening the same field to a very much larger scope. The meaning or contents of the semantic mes-sage would thus be transmitted, communicated, by different means, by more powerful technical mediations, over a far greater distance, but still within a medi-um that remains fundamentally continuous and self-identical, a homogeneous element through which the unity and wholeness of meaning would not be affect-ed in its essence. Any alteration would therefore be accidental" (Derrida, *Limited* 3).

does not conform to such limitations, but rather

> opens a semantic field which precisely is not limited to semantics, semiotics, and even less to linguistics. To the semantic field of the word *communication* belongs the fact that it also designates nonsemantic movements. Here at least provisional recourse to ordinary language and to the equivocalities of natural language teaches us that one may, for example, *communicate a movement*, or that a tremor, a shock, a displacement of *force* can be communicated—that is, propagated, transmitted (Derrida 1972, 309).

Again invoking reference to ordinary language philosophy and preparing the later discussion of J. L. Austin's, *How to Do Things with Words*, and the "speech act theory" he elaborates, Derrida has also foregrounded the issue of "the equivocalities of natural language" that will pose a considerable challenge to Fodor in his account of cognition. As we'll see in more detail when examining Fodor's proposed solution, Derrida notes here that the concept of context will play a powerfully determinate role in "reduc[ing] massively" "the field of equivocality covered by the word *communication*" (310). Rapidly reciting the series of interlocking questions surrounding 'context' activated as soon as such broad regulatory powers over the field of communication are granted to it, Derrida delivers an extremely efficient formulation of the problem: "But are the prerequisites of a context ever absolutely determinable? Fundamentally, this is the most general question I would like to attempt to elaborate. Is there a rigorous and scientific concept of the *context*? Does not the notion of context harbor, behind a certain confusion, very determined philosophical presuppositions?" (310).

The dependency of communication on the concept of context will only grow more troublesome when the discussion turns to consider "writing in its usually accepted sense…as a *means of communication*" (311). With this turn, the philosophical ground of the discussion begins to shift rapidly as Derrida, having introduced Condillac as an index of the "properly philosophical interpretation of writing" as a means to communicate a semantic message, will move quickly to isolate the decisive difference that enable writing

the greater range of extension than claimed by "locutionary and gestural communication" (311). This difference is "the notion of absence" that organizes and determines the value of writing as a means of communication.

The notion of absence is specific to writing in two senses, first, "it is the absence of the addressee," a quite ordinary, everyday use to which writing is put: "One writes in order to communicate something to those who are absent," Derrida acknowledges readily (Derrida 1972, 313). But from this premise, however, arrives the implication as to how this purpose is served, that is, absent the sender of the message, how is it that the message is transmitted by the means of writing? Derrida elaborates an answer that expands to pertain to "all language in general," and this will be main point that we establish now in view of the following sections on Genette and Fodor. "The absence of the sender," he writes, "the addressor, from the marks that he abandons, which are cut off from him and continue to produce effects beyond his presence and beyond the present actuality of his meaning, that is, beyond his life itself, this absence, which however belongs to the structure of all writing—and I will add, to all language in general" (Derrida 1972, 313).

If the structure of writing is to function in the absence of the sender, "beyond his life itself" and then the Derridian figure of survival and Sedgwick's reincarnation have surged into new relevance: "To write is to produce a mark that will constitute a kind of machine that is in turn productive, that my future disappearance in principle will not prevent from functioning and from yielding, and yielding itself to, reading and rewriting" (Derrida 1972, 316). Fodor's reliance on the availability of intention cannot resolve this, for it explicitly includes "the nonpresence of my meaning, of my intention-to-signify, of my wanting-to-communicate-this, from the emission or production of the mark" (Derrida 1972, 316).

Not wholly dissimilar from the structural analysis that Fodor will insist on to articulate the 'neutral structure' of the cognitive architecture, Derrida's demonstration of structural necessity of writing's connection to death establishes the need for different theory of literary narrative such as that developed by Genette, and along with it an analytic reading sensitive to the affective turbulence that this close

association of writing to death generates.

> All writing, therefore, in order to be what it is, must be able to function in the radical absence of every empirically determined addressee in general. And this absence is not a continuous modification of presence; it is a break in presence, "death," or the possibility of the "death" of the addressee, in the structure of the mark… (Derrida 1972, 316)

Section 1

Sedgwick's and Derrida's 'effects of writing'

Sedgwick's interpretive reading strategies are responsive to what Derrida broadly designates as the 'effects of writing' operating within a literary text: those that dis-organize, disrupt, deconstruct signification and meaning. Sedgwick's frequent identification of 'surprise' and of the capacity to be 'surprised' as crucially necessary to literary reading practices demonstrates a substantial connection to Derrida's analyses of literature, fictionality and 'that writing called literary.' More specifically, Sedgwick's attention to the affect of surprise, and of the capacity to be surprised by literature and critical theory—by reading otherwise—can be brought into closer connection with Derrida's lexicon when reframed in terms of knowledge, so that the possibility or capacity to be surprised is understood as a necessary prerequisite state or condition of 'non-knowledge'.

Reformulating Sedgwick's insight into Derridian phrasing, 'surprise' designates the general precondition antecedent to knowledge, the condition of possibility of knowledge. By re-articulating this 'knowledge deficit' in terms of 'surprise', the significance of this capacity beyond literary confines becomes easier to grasp, particularly when Sedgwick questions the 'routinization' of critical assumptions about biology, language, the effect of exposure, and the fetishization of binary conceptual pairs, that circulate

within theory.[60] This critical posture has been stated outright, as a preliminary interpretive axiom in the "Axiomatics" section of *Epistemology of the Closet* and is deployed as a critical reading strategy in "Jane Austen and the Masturbating Girl". In "Paranoid Reading and Reparative Reading, or, You're so Paranoid You Probably Think This Essay Is About You" Sedgwick questions the sprawl of a methodology of suspicion throughout critical theory, what "Paul Ricoeur memorably called the 'hermeneutics of suspicion,'" and plainly summarizes her concern with the habits resulting from the wide adoption of them, observing that "to apply a hermeneutics of suspicion is, I believe, widely understood as a mandatory injunction rather than *a possibility among other possibilities*" (my emphasis, Sedgwick 2011, 125).

Sedgwick's discomfort with the interrelation between a type of critical practice and the cognitive and affective habits that practice promotes is obscure at first glance, yet quickly becomes a compelling insight as she sketches how the "methodological centrality of suspicion to current critical practice has involved a concomitant privileging of paranoia" (125). While attention to the interrelation between a critical practice

---

[60] Cf. "Here are a few things theory knows today or to phrase it more fairly, here are a few broad assumptions that shape the heuristic habits and positing procedures of theory today (theory not in the primary texts, but in the routinizing critical projects of 'applied theory'; theory as a broad project that now spans the humanities and extends into history and anthropology; theory after Foucault Greenblatt, after Freud and Lacan, after Lévi-Strauss, after Derrida, after Feminism) when it offers any account of human beings or cultures:

> 1. The distance of any such account from a biological basis is assumed to correlate near precisely with its potential for doing justice to a difference (individual, historical, and cross-cultural), to contingency, to performance force, and to the possibility of change.
>
> 2. Human language is assumed to offer the most productive, if not the only possible, model for understanding representation.
>
> 3. The bipolar, transitive relations of subject and object, self to other, and active to passive, and the physical sense (sight) understood to correspond most closely to these relations are dominant organizing tropes to the extent that their dismantling as such is framed as both an urgent and an interminable task. This preoccupation extends to such processes as subjectification, self-fashioning, objectification, and Othering: to the gaze; to the core of selfhood whether considered as a developmental telos or as a dangerous illusion requiring vigilant deconstruction.
>
> 4. Correspondingly, the structuralist reliance on symbolization through binary pairings of elements, defined in a diacritical relation to one another and no more than arbitrarily associated with the things symbolized, has not only survived the structuralist moment but, if anything, has been propagated ever more broadly through varied and unresting critique—critique that reproduces and popularizes the *structure*, even as it may complicate an understanding of the *workings*, of the binarisms mentioned above along with such other as presence/absence, lack/plentitude, nature/culture, repression/liberation, and subversive/hegemonic." (Sedgwick and Frank, "Shame in the Cybernetic Fold: Reading Silvan Tomkins" 94-95).

and cognitive and affective habits fueled much of her analytic posture in *The Epistemology of the Closet*, "Paranoid Reading and Reparative Reading" marks a fundamental shift in her own thinking made possible by the influence of Silvan Tomkins's affect theory and non-Freudian psychoanalysis of Melanie Klein. The structural flexibility afforded by Klein's elaboration of the paranoid-schizoid and reparative positions, extended through Tomkins's models of affective and cognitive profiles, provides her with alternative conceptual resources in order to disrupt the recirculation of dominant critical assumptions.

The identification of paranoia as an operative pattern is important specifically because of her attention to the 'privileging' effect that is active within critical theory. This 'effect of privileging' is not an overt goal, yet nor is it unwanted or perceived as a negative, because in moonlighting as curiosity, the acute suspicion underlying paranoia is closely linked with the feeling of analytic precision and critical rigor. The purchase that paranoid structures of knowledge hold within critical theory (and especially within feminist and gay and lesbian theory) operates a positive feedback dynamic, whereby the analytic modes of suspicion produce ample confirmation of their own premises and thus accrue more 'strength' as a theory.[61] Initial hypotheses of suspicion find confirmation, supported and reinforced by the abundant supply of "evidence of systems oppression," signal 'success' and thus steadily legitimates a previously provisional line of inquiry. Thus the privilege afforded to paranoia is validated constantly; "to theorize out of anything *but* a paranoid critical stance has come to seem naive, pious, or complaisant" (Sedgwick 2003b, 126).

This observation registers the conundrum of Foucault's 'repressive hypothesis' and Sedgwick carefully outlines why she is motivated to push back against the sprawl of the paranoid critical stance: "I myself have no wish to return to the use of 'paranoid' as a pathologizing diagnosis, but it seems to me a great loss when paranoid inquiry comes to seem entirely coextensive with critical theoretical inquiry rather than being viewed as one kind of cognitive/affective theoretical practice among other, alternative kinds" (Sedgwick

---

[61] This is a placeholder footnote for more explanation of Tomkins's usage of the terms 'weak' and 'strong' as diagnostic terms to describe the behavior of different theories. Weak theories have a circumscribed, limited purview, only enabling the connection of a few data points, whereas strong theories connect and 'network' a greater (and increasingly greater) number of data points.

2003b, 126). To borrow a phrase from psychoanalysis, what Sedgwick is prompting us to consider is that critical theoretical inquiry is or has become 'overdetermined' as paranoid inquiry, and the problem that this overdetermination presents for us (those writing and thinking about critical and interpretive 'tools') is that we may believe ourselves to be executing rigorous critical theoretical inquiry when we are 'just being paranoid.' The implication lurking within the double sense of the phrase '*just* being paranoid,' that one is *only* being paranoid (as opposed to thinking critically) or that paranoid thinking is so 'easy' to engage in (as opposed to critical theory) that it always threatens to cloud one's judgment or present faulty, false-positive conclusions—this implication has the same logical structure of the 'open secret' that Sedgwick explores in *The Epistemology of the Closet*.

Indeed, she herself identifies much of her own previous work as fully participating in the interpretive machinery of paranoid inquiry. What draws me to "Paranoid Reading, Reparative Reading" is that the combination of Tomkins and Klein has enabled her to perform a certain sort of 'self-analysis', a critical position that allows her to put into question her own premises, as well as the wider collective of gay and lesbian critical theory. To what extent have the critical and theoretical practices of suspicion and the exhilaration of 'overturning' of critical assumptions come to structure and shape the cognitive and affective habits we identify with knowledge and critical insight? And how can we approach this question with an alternative set of critical assumptions?

In pursuit of an answer to these questions Sedgwick turns to the work of Melanie Klein and Silvan Tomkins for models of differently structured cognitive/affective theoretical practice, ones which conceive of 'surprise' and paranoid knowledge differently, and thus can provide useful leverage for her critique of the pervasiveness of the hermeneutics of suspicion. Klein's explorations of the paranoid position and her subsequent expansion of these preliminary observations into a broader dynamics of 'positions' instructs Sedgwick's effort to examine her own critical procedures and account for their effects. My argument is that Sedgwick's thinking on 'surprise' and Derrida's attention to the destabilizing 'effects of writing' are each analyses of interrelated problems. I stand Sedgwick and Derrida together as an alternative to the theory

of literature developed by Cognitive Literary Studies, because the interpretive frameworks they develop responds directly to questions of writing, questions that destabilize the foundational integrity that Cognitive Literary Studies places on cognition as the basis of literature. The Derridian analysis of the destabilizing "effects of writing" elaborated throughout his work, from the analysis of writing in *Of Grammatology*, to the critique of Husserl's phenomenology in his Introduction to "The Origin of Geometry," and the reading of Plato's *Phaedrus* and Mallarmé presented in *Dissemination*, demonstrates the that writing opens a vast range of questions clustered around authorship, signification, and meaning. In each case, the problems that writing activates are shown to fundamentally threaten the integrity of philosophical meaning, even as philosophy (in the case of phenomenology) claims to bypass these issues through an enthusiastic embrace of 'the things themselves'.

Although these texts are primarily focused on works of philosophy, the problems they develop and explore are coimbricated with literary writing (indeed, a tired centerpiece of the responses to Derrida's analysis is that it is "too literary"). In this chapter, I will draw specifically from Derrida's work that confronts the question of literary writing directly, because it outlines a theory of literature and literary writing that directly challenges the assumptions that structure Cognitive Literary Studies. The literary critics and theorists adapting cognitive science as tools for literary analysis and reading accomplish their task by ignoring, denying, or disavowing the effects of writing as an active component of the question "What is Literature?" Derrida and Sedgwick both demonstrate serious problems with the project proposed by Cognitive Literary Studies, and also indicate that writing and literature 'collaborate' or "co-operate" in ways that challenge and disrupt an easy identification of writing with speech, speech with cognition, and writing with cognition.

The question at hand is: Do the 'destabilizing effects of writing' Derrida identifies as constitutive of literature persist in the account of literature proposed by Cognitive Literary Studies? Without question, they do. It is Derrida's work that shows us the degree to which the theory of literature developed by Cognitive Literary Studies leaves the question of writing unaddressed and it is the work of Eve Kosofsky

Sedgwick that demonstrates a method of reading otherwise by which the 'destabilizing effects of writing' can be seen as resources that nurture readers and allow writing to survive.

Section 2
Genette and Motivation

In his short essay "*Vraisemblance* and Motivation" Gérard Genette advances a critique of a literary theory that takes the concept of *vraisemblance* to be the key term of its analysis of literary realism. Genette opens his essay with a discussion of two canonical works of French literature, written in the seventeenth century. *Vraisemblance*, as the introductory footnote given by translator David Gorman, helpfully informs us, "[f]rom the title onward" is "the key term in this essay" (254). Gorman's note continues:

> The appropriate translation of this word as used in a seventeenth century context is 'probability,' whereas in a nineteenth century context it would be 'plausibility.' Unfortunately, Genette's whole case involves *comparing* the two kinds of *vraisemblance*. The continuity between them can be expressed elegantly in French, but not in English. Therefore I have left the term in French to avoid some ungainly substitute such as 'verisimilar.'" (254)

Hearing this "unfortunately" as a remark on his own—the translator's—dilemma when confronted by linguistic obstacles inherent in any translation projection, Gorman's note provides us with two additional insights specific to the translation of a contemporary text (Genette's) which examines historical fiction. First, that there has been a historical, diachronic shift of the "appropriate translation" of the word '*vraisemblance*' between the seventeenth and nineteenth century; and second, that Genette's interest in '*vraisemblance*' as a point of analysis is anchored in the *continuity* maintained by the French word, the smoothing effect of a single signifier that masks a shift in signification. We gain a sense of the import of this continuity by observing Gorman's own decision to maintain the use of the original French in his own

English translation, an indication that there is more to this thread of historical continuity than merely a literary critic's indulgent play on words.

The term *vraisemblance*, writes Genette, is found at the center of "two great debates" in seventeenth-century France, the first involving Pierre Corneille's "tragicomedy" *Le Cid*, and the second surrounding what Genette describes as a work of prose (and that is usually considered to be a novel), *La Princesse de Clèves*. Although neither text is itself directly concerned with the concept of *vraisemblance*, Genette's focus is on how the term is deployed by contemporary seventeenth-century commentators in their critical writing on *Le Cid* and *La Princesse de Clèves*, as well as how the concept circulates through popular public responses to these literary works. Comparing the reception of these two works, Genette focuses on the overlapping usage of *vraisemblance* as a crucial term in the writings of different scholars and succinctly outlines the conjunction that orients his own analysis: "[i]n both cases, in fact, critical examination of a work is reduced essentially to a discussion of the *vraisemblance* of one of the constitutive actions of its story" (239). In *Le Cid*, this action is "the conduct of Chimène toward Rodrigue after the death of the Count," (i.e., of her falling in love with Rodrigue after he has killed her father) and in *La Princesse de Clèves*, it is "the confession that Madame de Clèves makes to her husband" that she loves another man than him (239). The critical discussions that Genette cites demonstrate a ready, familiar use of *vraisemblance* as a common critical term, as a self-explanatory standard against which literary works should be measured.

It is this aspect of *vraisemblance* that I emphasize here as it connects to the critical consensus deployed by Cognitive Literary Studies, that is, the consensus that literary narratives are matter-of-factly understood to be written descriptions of individuals' actions, behaviors, and thoughts; and, that the degree to which a work of literature can be shown to correspond to 'real life' is indicative of its status as literature.[62] On the view, characters are people, people with thoughts and motivations that derive from their 'real lives,' lives

[62] An additional and not insignificant connection is observable between CLS's position that novels are best written when the writing itself disappears as writing, becoming a transparent medium for cognition and Genette's observation of the vraisemblable 'costs' of the intrusion of discourse upon narrative.

lived either contemporaneously to the reader's—or not; as the universal, a-historical status arrogated to cognitive mechanisms means that their applicability is not limited to any given historical period.[63] The strength of this connection between how *vraisemblance* is deployed as a criterion of literary criticism and how Cognitive Literary Studies deploys its argument for the cognitive foundation of literary works intensifies as Genette's analysis advances and we learn in more detail how *vraisemblance* functions as an anchoring principle of literary interpretation.

Continuing on, Genette writes, "[i]n both cases [of critical writing], finally, the close connection, or better, the amalgamation of the notions of *vraisemblance* and propriety is especially plainly marked, an amalgamation perfectly represented by the familiar ambiguity (between 'obligation' and 'likelihood') of the verb 'should' [*devoir*]" (239). Recall the earlier attention we paid to the footnote from David Gorman noting that the meaning of *vraisemblance*, in the seventeenth-century context under examination here, is 'probability', yet what Genette identifies here is an initial flexibility to the critical purposes toward which *vraisemblance* is put, the amalgamation of 'probability' and 'propriety'. This 'interpretive subjunctive' by which critiques of what a character *did* are given in terms of what a character *should have done*, affords critics with room for judgments of literary works, that, in turn, provide us with further insight into the composition of *vraisemblance* as an interpretive principle.

Closely reading the critiques supports Genette's suggestion that, from the manner by which characters' actions are judged to be *invraisemblable*, the positive form can be extrapolated with little effort, those actions that are "contrary to good manners, and, at the same time,...contrary to all reasonable foresight," are deemed to be "infraction and accident" thereby moving them beyond the reach of intelligible intention. Genette again cites the critics on *Le Cid* and *La Princesse de Clèves*, "the subject of *Le Cid* being bad because Chimène 'should not' receive Rodrigue after the fatal duel, or desire his victory over Don Sanche, or accept,

---

[63] Cf. Kramnick, "Against Literary Darwinism". Here I am referencing the self-validating composition of 'cognition' in its use by cognitive literary theorists. These literary theorists take up the idea of cognitive 'mechanisms' that are discerned or theorized by cognitive science and psychologists as being 'foundational' to or 'prior to' consciousness, and thereby contend a 'cognitive continuity' to bridge historical distance.

even tacitly, the prospect of marriage, etc.; the action of *La Princesse de Clèves* being bad because Madame de Clèves 'should not' take her husband as a confidant" (240). In other words, the critics articulate their objection to a characters' actions in terms of propriety, a transgression of an obligation, and thus judge these works of literature *invraisemblable* because the transgressive quality of these characters' actions, behaviors, and motivations disappoints or disrupts the expectations of readers. Not applied to accidental or unintentional events or actions, these judgments are reserved specifically for depictions of deliberate decisions made by the characters in the narrative.

Genette's attention to this moment in French literary history is valuable for our purposes insofar as it provides an alternative perspective from which to evaluate the critical standard of vraisemblance (as Genette terms it—of the 'realist illusion' in literature), and to outline a structural account of how vraisemblance underwrites the genre of the 'modern' 'psychological novel'. For cognitive literary studies, literary realism serves much like a lowest common denominator that enables complex questions to be reduced, the baseline certainty that the observed coincidence or correlation between cognitive mechanisms and literary fiction confirms and finalizes the condition of possibility for literature itself. This effect of simplification allows critics to bypass more complicated questions posed by the 'destabilizing effects of writing.' In like fashion, the critical usage of *vraisemblance* simplifies the terms by which literary writing is evaluated, foreclosing further exploration of the effects of writing, of the deferral, delay, and disruption that writing operates within and alongside of a literary narrative.

As a critical concept the *vraisemblable* is reserved exclusively for literary fiction, absent from assessments of historical[64] or particular truth (neither is it a critical term used to discuss written histories or records of past events, nor is it applied when discussing the unique, unexpected traits of a particular character) as demonstrated in the peculiarly palpable mismatch between historical fact and probable action. It operates altogether differently in the context of fiction: the realm of 'the possible' is bent and steadily inflected to

[64] Barthes, "The Reality Effect" would be an example of this differentiation between literary and historical writing shifting.

mean 'plausible'. "It has been understood since Aristotle," Genette writes, "that the subject of theater—and, generally, of all fiction—is neither the true nor the possible but rather the *vraisemblable*, yet the tendency has been to identify the *vraisemblable* more and more closely with the 'should-have-been'" (240). Having demonstrated the link between *vraisemblance* and propriety that animates the usage by seventeenth century critics, Genette takes his formulation one degree further and brings this term into full communication with the stakes of my critique of Cognitive Literary Studies. He writes, "In fact, *vraisemblance* and propriety are joined together under a single criterion, namely, 'whatever is conformable to public opinion' (Rabin 114: this is his definition of the *vraisemblable*). The 'opinion,' real or supposed, is almost exactly what would be called today an ideology—in other words, a body of maxims and presuppositions that constitutes, simultaneously, a vision of the world and a system of values" (240).

Cast in this way, the interpretive principle of *vraisemblance* indexes a degree of coordination between the narrative text and the world, a relationship which solicits the judgment and evaluation of its readers. The actions that take place within the narrative are discussed precisely, one would almost say 'naturally' within the frame of real life, offering critics the opportunity to "produce judgments of *invraisemblance* in an ethical form…, or in a logical form" (240). The narratives whose characters' actions fall beyond the reasonable expectations of their readers are accordingly judged to be *invraisemblable*, that is, depicting behaviors and actions that are "ultimately possible and conceivable, but as an accident" (241). These 'accidental' sequences of action within the narrative render the question of intention unanswerable insofar as the *invraisemblable* designation relieves the characters of any responsibility for intention by being classified as a mistake. Genette summarizes the principle as follows:

> Such, roughly described, is the cast of mind upon which the classical theory of the *vraisemblable* rests explicitly, as, implicitly, all the systems of *vraisemblance* still active among popular genres like the mystery novel, the sentimental romance, the Western, etc. From one period to another, from one genre to another, the content of the system, in other words the tenor of the norms or the *essential judgments* that constitute it, may vary wholly

> or in part…; what remains, and what defines the *vraisemblable*, is the formal principle of respect for the norm, in other words the existence of a relation of implication between the particular behavior attributed to a given character and a general maxim. (241)[65]

The key point of Genette's analysis that I will emphasize here is what he identifies as the "relation of implication between the particular behavior attributed to a given character and a general maxim" (241). Most importantly, it is this relation of implication that enables the actions and behaviors to be 'displayed' by the narrative without support or explanation, and thereby further strengthens the illusion that readers are merely observing 'real life' to the very extent that they find these behaviors to fit with the organization of life in a given historical time and place. The *vraisemblable* quality of the narrative camouflages its fictional status, allowing the narrative to appear merely to 'report' back on behaviors and events happening outside of the text and that are intelligible as such without a more detailed account: "[t]he relation of implication also functions as a principle of *explanation*—the generality determines and thus explains the particular, so that (for example) to understand a character's behavior is to be able to refer to it as a received maxim" (241). When action and behaviors are obscure to readers, the relation of implication is reversed: "inversely, an action is incomprehensible, or 'extravagant,' when no accepted maxim can account for it" (241). Here, we are able to connect the system of 'vraisemblance' to the implicit framework upon which Cognitive Literary Studies draws its own explanatory strength. The degree to which behaviors and actions are *understood and accepted* is indicative of the fit between the narrative and the "system of vraisemblance to which it is attached" (242).

On this point I would like to expand the discussion slightly, in order to more fully map out the connection I am drawing between Cognitive Literary Studies and Genette's analysis of *vraisemblance*. Proponents of Cognitive Literary Studies may argue that their application of cognitive principles to literary

---

[65] Sedgwick's *Between Men* is an example of how this system of *vraisemblance* can operate to conceal or encrypt certain behaviors as well as to capture and display them.

interpretation has the effect of (scientifically) resolving the structure of implication and explanation that Genette identifies to animate the *vraisemblable*. However, the close examination of cognitive literary reading strategies undertaken thus far has suggested a more limited scope of purchase, curtailed by the historical frame of the twentieth century within which cognitive science was developed. Literary theorists have responded to this limitation by themselves turning to evolutionarily inflected cognitive science, claiming to resolve this limitation by postulating a line of continuous cognitive structure that extends to include pre-historical scenes of human life.

In establishing a theoretical basis for a diachronic cognitive 'continuity,' Cognitive Literary Studies re-establishes the relation of implication between any narrative text and the decidedly modern understanding of cognition and cognitive mechanisms. The introduction of autism as a crude form of a controlled variable inversely creates the implication of a discernible, synchronic 'cognitive uniformity' to species of cognition, those termed 'neurotypical' and those with autism. The observable differences of preferred social and physical interactions by those with autism is quickly appropriated, along with a presumed dislike for fictional stories, and packaged together as evidence of the lack of the cognitive structures and mechanisms which "make literature as we know it possible" (Zunshine).

Genette began by introducing critics who complained of the works' *invraisemblance* rather than those who praised the works' *vraisemblance*, and this negative formulation of their critique instructs us as to yet another aspect of the system of *vraisemblance*. The 'relation of implication' that connects the particular behavior of a character to "a body of maxims and presuppositions" is, usually, *silent*: the *vraisemblable* narrative is marked by the absence of the intrusion of explanatory discourse or authorial justification. The silence of this relation of implication is equally at work, yet now invisible[66], in the persuasiveness of Cognitive Literary Studies. Genette provides us with an excellent summary of the stakes of the *vraisemblable*:

---

[66] The point I'm making here is that there's an escalating movement of erasure, the 'relation of implication' is silent, but the cognitive relation is arguing for its invisibility, along the lines that these mental structures are 'always-already' invisible insofar as in always being there prior to the act of writing, Cognitive Literary Studies' line of argumentation is that there was never a choice.

> A *vraisemblable* narrative is thus a narrative where the actions answer, as so many applications of particular cases, to a body of maxims accepted as true by the public to which the narrative is addressed; but these maxims, due to the very fact that they are accepted, most often remain implicit. The relationship between the *vraisemblable* narrative and the system of *vraisemblance* to which it is attached is thus essentially silent: generic conventions function as a system of natural forces and constraints, which the narrative follows as if without perceiving and, *a fortiori*, without naming. (242)

Alongside Genette's identification of the narrative of implicit *vraisemblance* the "opposite extreme" is, as he terms is, a "liberated *invraisemblable* narrative," i.e. "works most emancipated from any allegiance to 'public opinion'" (243, 242). In these narratives, citing Stendhal's *Le Rouge et le Noir* and *Vanina Vanini* as two examples, the "tacit contract between a work and its public" is dispensed with altogether. Narrative events or actions that 'shouldn't have happened' stand alone in these works, unfathomable by readers and unaddressed by the author. Genette's description of these works as 'liberated' and 'emancipated' indicate that narratives of this type no longer trouble themselves "to respect a system of general verities, but depend only upon some general truth or a profound imagination" (242). These liberated *invraisemblable* narratives, while rejecting the "servility of the *vraisemblable*" to a current ideology, nonetheless share a common feature: "which is the identical erasure of commentaries and justifications" (242).

Genette observes that the actions that take place in Stendhal's narrative do so without the auxiliary commentary and explanatory structure in place to communicate the 'reasons' or 'motivations' behind a characters' actions. The actions of the narrative "are not, in themselves, more 'incomprehensible' than many others" for lacking this support, but rather the withholding of it creates an alternative "enigmatic" narrative effect. Genette writes:

> but we will say that Stendhal deliberately chose to conserve in them, or maybe to confer on them, by his rejection of any explanation, this uncivilized individuality that makes great

> actions unpredictable—and great works too. The truthful accent, a thousand miles from any sort of realism, is not separable from the violent feeling of a wholly accepted arbitrariness, which does not bother to justify itself. (242)

In considering these two examples as opposite ends of, as Genette will soon call it, "a 'gradation' in the scale of narratives" the distinction between them can be paradoxically hard to discern as on both ends of this scale we'll find "the extreme reserve with respect to commentary and almost complete absence of general maxims" (243). However, it is this aspect of the *vraisemblable* and *invraisemblable* interpretive 'rubric' that bears the most important connection to my critique of the reading strategies of cognitive literary studies. As will be made clearer through further reading of Genette's analysis, the "two 'extremes' represented here" in the *vraisemblable* and *invraisemblable* narrative describe formally identical narratives. The distinction they describe is more accurately understood as the degree to which narratives disguise themselves as artistic creations, and, further still, as writing, as words exposed to interpretative reading. Cognitive Literary Studies, in the broadly discernible scope of its theoretical project, assumes possession of this *vraisemblable* narrative spectrum, the "'gradation' in the scales of narratives," and seeks to cunningly transform the "valorizing connotation" of *vraisemblance* into a sign of cognitive processes and mechanisms which govern and regulate its function, and determine its purpose.

Having taken up the critical use of the word *vraisemblance* and subsequently analyzed the structure of its binary privative, Genette observes that this critical criterion "leads naturally to the hypothesis of a 'gradation' in the scale of narratives' a spectrum like structure along which narratives can be charted according to their degree of vraisemblance (243). The introduction of this diagnostic range to vraisemblance thus liberates narrative from the coarse and rigid critical binary, and prepares the way for Genette's identification of an alternative narrative type, thus far obscured from clear delineation or recognition. Between the *vraisemblable* narrative and the *invraisemblable* narrative, which Genette now casually reterms "the enigmatic narrative," he describes or deduces the attributes that this alternative narrative type requires:

> ...then it would remain to locate the narrative type corresponding to the *in-between* position, of the *half-artful* narrative, in other words: emerging from the natural silence of the *vraisemblable* and having yet to attain the deep silence of what we might call...the *improbable*. (243)

This "middle region" of narrative demarcates a story "too different from the conventions of the *vraisemblable* to base itself upon the consensus of vulgar opinion, but at the same time too attached to the consent of opinion to impose upon itself without commentary any actions the reasons for which would then run the danger of escaping it" (243). Genette's situation of the question here is immediately eminently logical, and in fact a faithful follow through regarding the theoretical basis of the *vraisemblable* scale. Is it possible to discern a narrative "still too original (perhaps too 'true') to be transparent to its public, but still too timid, or too compliant, to assume its opacity"? Without recourse to the ideology of *vraisemblance*, this type of narrative would

> seek to give itself the transparency that it lacks by multiplying its explications, by supplying for every purpose the maxims, unfamiliar to the public, capable of accounting for the actions of its characters and the interconnections of its plot, in short of inventing its own conventions and in simulating in every work and for the needs of its purpose an 'artificial *vraisemblance*' that would be the theory—this time, and perforce, explicit and declared—of its own practice. (243).

This type of narrative is, Genette asserts, the "Balzacian narrative" which is marked by the profusion of "instructional clauses that introduce, with a powerful clumsiness, those explanatory flashbacks in the *Comédie humaine*: 'Here is why…'; 'To understand what follows, some explanations are perhaps necessary…'; 'This calls for an explanation…'; 'It is necessary to go into some explanation...'; "It is necessary, for an understanding of this narrative…;" (243). Genette advances further still, and observes that these explanatory clauses do more than gather and knit together the connections among various

narrative facts, but extend to serve the same purpose as the silent 'relation of implication' supplied by *vraisemblance*. The "most frequent and characteristic manifestation" of the Balzacian narrative's "artificial *vraisemblance*" is "rather the justification of a particular fact by a general law presumed to be unfamiliar to, or perhaps forgotten by, the reader, and which the narrator must teach him or recall for him" (244).

In the course of Balzac's fictional project, the additions of general laws steadily accumulate and ultimately precipitate "a veritable anthropology of the French province, with its social structures," its characters, its professional categories, its manners, its intellectual traits, and its passions. All of these passages function, Genette writes, as "so many formulas that, along with many others, compose as it were the ideological 'background' [English in original] 'necessary for an understanding of' a large part of the *Comédie humaine*" (244). Acknowledging that Balzac "has 'theories about everything'" in a small citation of Claude Roy, Genette ensures that this fact about Balzac's writing is not brushed aside as a casual observation, but is recognized for the true role these theories play: "these theories are not there for the sole pleasure of theorizing; they exist first in the service of the narrative: they serve it at every moment as warnings, as justifications, and *captatio benevolentiae* [request for sympathy], they seal its cracks, they mark all its intersections" (244).[67]

Genette now begins tracking ever more closely to the path carved out by 'cognitive' reading strategies, and we see this "middle region" of narrative more fully mapped out. Among the interventions made within the Balzacian narrative there is a discernable trend, Genette writes: "Less evident but more numerous and in the end more important are the interventions that bear on the determination of behavior, individual and collective, and that reveal the author's desire to direct the action in one direction and not in some other" (244-245). The identification of this common trait of "artificial vraisemblance" of the Balzacian

---

[67] The overlap between this description of how Balzac's narrative is composed, i.e., with many different 'theories' echoes the core idea of Silvan Tomkins's affect theory. In Tomkins's thinking, theories of affects are made and constantly maintained by every individual (lay person and learned) unconsciously, which operate as interpretive structures used by individuals to process their daily experiences. See "Shame and the Cybernetic Fold" for a sustained discussion.

narrative, Genette identifies that the production of this "artificial vraisemblance" leads the narrative interventions to concentrate or cluster around "a determination of behavior" in the service of clarifying the stakes of a behavior "individual or collective" that is ultimately the author's desire to direct narrative action, this claim raises new questions as to the purchase, or necessity of cognitive reading strategies. One of their premises, after all, is that the behavior of characters within a novel is governed by, and is uniquely interpretable by recourse to, cognitive laws that ostensibly govern our own cognitive processes. Genette's insight here has the virtue of acknowledging the tendency of narrative to specifically address character's behavior, as well as unraveling the necessity that this behavior is governed by cognitive laws instead of the "author's desire" (245).

Why then does this development matter so greatly? It does appear, after all, to be a rather commonplace observation, and even one that cognitive literary critics would readily admit. Thus the sleek, recursive elegance and convenience of the cognitive theoretical basis: certainly the author's desire to direct the action is at work, but these desires are structured by the cognitive laws and processes of the author's mind. The significance of what Genette observes and now makes explicit is that this fact about the Balzacian narrative, this possibility (or even further still this condition of the possibility of *vraisemblance* in this case) derives from an infinite flexibility of the narrative, the "great sequences of pure intrigue" in Balzac's novels "are full of those decisive actions the consequences of which could have just as well have been entirely different" (245). The author's determination of characters' behavior is wholly free, the narrative does not restrict authorial choice.

Yet the narrative is indeterminate, and the actions that take place within its scene bring consequences that *are not* entailed: these consequences are entirely artificial and yet in being nested among other possible consequences, they accrue the sense of a determination issuing from out of a field of indeterminacy. Genette: "When a character in Balzac is on the path to success, all his actions pay off; when he is on the slope toward defeat, all this actions—*even precisely the same ones*—conspire in his fall: there is no better illustration of the uncertainty and irreversibility of human affairs" (245, my emphasis).

Here the coimbrication of cognitive interpretive strategies and those structured around *vraisemblance* pull yet closer together, and it is Balzac himself this time who articulates exactly why the authorial interventions Genette is tracking work to synthesize a quality of indeterminacy all their own insofar as they are connected with an expectation that 'life is stranger than fiction' and how the resolution of this indeterminacy by a sequence of authorial decisions accrues additional value:

> Quite often, particular actions of human life seem, *literally speaking, invraisemblables*, though true. But isn't this because we almost always neglect to cast a kind of psychological light on our spontaneous decisions, failing to explain the mysteriously contrived reasons that necessitated them?...Many people prefer to deny the outcomes that measure the force of the links, knots, and bonds that secretly fuse one fact with another in the moral order. (Balzac, quoted 245, Genette's emphasis)

Balzac's forthright admission regarding the specific utility of "psychological" interventions for the narrative project is immediately amplified by Genette:

> We see how "psychological light" indeed has the function here of exorcising the *invraisemblable* while revealing—or presupposing—the "links," "knots," and "bonds" that guarantee, more or less, the coherence of what Balzac calls the moral order. (245)

At this juncture every indication seems to confirm the cognitive premise, here referred to "a kind of psychological light," that narrative action and behavior are best understood with recourse to psychology, but this would mistake a crucial detail. As Genette will now demonstrate, the utility of this "psychological light" is its ready acceptance as an explanatory basis or, to use a term Genette will introduce later on, as an authorial "alibi" which acts to abrogate the author of responsibility over characters' actions and behaviors by "exorcising the *invraisemblable*" quality, that "uncertainty and irreversibility of human affairs" (245). That is, more simply put, that the 'psychological light' is uniquely well suited to establish and enforce

the relation of implication between "the *vraisemblable* narrative and system of *vraisemblance*" thereby to "give to itself the transparency that it lacks" (242, 245).

At this point the tension between the *vraisemblable* and *invraisemblable* indeed appears to have heightened further still, as the "uncertainty and reversibility of human affairs" emerges as the very condition of possibility which the Balzacian narrative interventions ceaselessly, casually, work to eliminate by deft interpositions of 'psychological' "theoretical justifications" for the actions taken by characters in its story (245). And yet, perhaps, not so deftly, as these 'psychological' justifications begin to contradict one another, as if stepping on each other's toes, within the same narrative and even concerning the same character: "[i]t happens moreover that the same event leads successively to two opposite consequences, just a few lines apart" (245). Genette cites for us four passages from Balzac's *Le Curé de Tours* in which the narrator is discussing the Abbé Birotteau, providing us with two pairs of the narrator's interventions each of which deploy psychological theories as connective tissue tying together a narrative sequence of event and consequence. In both sets the same psychological 'theory' is used to justify mutually conflictual narrative events. I reproduce Genette's citation in full:

> It happens moreover that the same event leads successively to two opposite consequences, just a few lines apart. 'Since the nature of limited minds helps them to grasp minutiae, he suddenly indulged himself in some very grand reflections on these four events imperceptible to anyone else' (*Le Curé de Tours*); but 'The Vicar came to recognize the truth a little later, the signs of a concealed persecution...the evil intentions of which would doubtless have been much more quickly divined by a more intelligent man.' Or again: 'With that interrogative sagacity that priests develop through the habit of directing consciences and digging out trifles in the confessional, Abbé Birotteau….'; but: 'Abbé Birotteau, … who had no experience of the world and its ways, and who lived between mass and confession, grandly occupied in deciding the cases of the gentlest consciences, in his role as confessor to the town boarding-schools and some fair souls who appreciated

him, this Abbé Birotteau could be considered just a big child'. (246)

These oversights and inconsistencies are not being scrutinized here in order to discredit Balzac or expose him as a fraud, nor is Genette keen to disillusion Balzac's readers by disrupting the presupposed basis for their heretofore high appraisal of *Le Curé de Tours*. Instead, Genette simply acknowledges that although "there is an element of negligence in these little contradictions" these are not unresolvable conflicts nor would their coordination entail a grand architectonic reworking of the narrative, they remain issues that "Balzac would hardly have had trouble eliminating had he been told about them" (246).

Recall that one of Genette's primary observations about the *vraisemblable* narrative, distilled from his analysis of criticism which utilized *vraisemblance* as a critical standard, was that the "relation between the *vraisemblable* narrative and the system of *vraisemblance* to which it is attached is...essentially silent" (242). In the *vraisemblable* narrative, "actions answer...to a body of maxims accepted as true by the public to which the narrative is addressed" and it is quality of implication, this 'implicit support,' which the Balzacian narrative gives to itself whereby, "generic conventions function as a system of natural forces and constraints, which the narrative follows as if without perceiving" (242). What is thus additionally remarkable about the selections from *Le Curé de Tours* which Genette pulls forth is, in fact, just how unproblematic and insignificant they are to the narrative; a demonstrable conflict when read closely and examined side-by-side, and yet that which "the narrative follows as if without perceiving" (242). This connection is not explicitly drawn by Genette but it bears particular import in challenging the coherence of the cognitive conventions alleged to underwrite narrative itself.

Genette sees these passages as confirming, with uncanny accuracy, the stakes of the boastful provocation embedded in a passage from *Eugénie Grandet*, the sly suggestion that *invraisemblance* emerges from our "neglect to cast a kind of psychological light on our spontaneous decisions" (245). "Such lapses," he writes, referring to the passages from *Le Curé de Tours*, "also reveal deep ambivalences, which the 'logic' of the narrative can reduce only superficially" (246). The marked use of "'logic'" appropriately captures both the artificial quality of *vraisemblance* these narrative interventions achieve while simultaneously

underscoring the fundamental limitations on returns of such a project. This reconciliation of Balzac's narrative logic only more fully conceals the essentially arbitrary quality of any narrative and the concomitant 'liberty' of the author: "Such motivational ambivalences thus grant a complete liberty to the novelist, with the responsibility for insisting, by way of epiphrasis, at one time one value, at another on another value" (246).

The terrain outlined here of the "narrative type corresponding to the *in-between* position," between the *vraisemblable* and the *invraisemblable* narrative unexpectedly gives considerable ground from which to reassess this pair of oppositional terms. The exploration Genette has undertaken of this "*half-artful* narrative" steadily disclosed the means by which the *vraisemblable* narrative is sustained. In so doing, Genette's reading identified for us the common focus of such interventions, "the determination of behavior," as well as the peculiar flexibility of psychological justifications which the narrative constantly affords itself. It is this flexibility, or better put, the condition of possibility that underlies this narrative structure, that Genette names uncertainty, indeterminacy, and finally "*the arbitrariness of the narrative*" (247). The function these interventions serve, their effect, is to disguise themselves *as* determinations concealing the narrative's artificial *vraisemblance* by calling into question the very terms by which *vraisemblance* establishes itself.

Smuggling itself past readers under the cover of *vraisemblance*, the narrative of the middle range deploys its "ideological background" through "the justification of a particular fact by a general law presumed to be unfamiliar to, or perhaps for gotten by, the reader, and which the narrator must teach him or recall for him" (244). Balzac's taunt describes for us the inverse 'relation of implication' persisting between *invraisemblance* and the "psychological light" explaining "the mysteriously contrived reasons" necessitating our "spontaneous decisions". This middle range narrative keeps a comfortable distance from the "infallible sequence that one ascribes to [the narrative] out of faith in its assurance and…its 'apparent rigor'" and in this space demonstrates the complete liberty of writing, the arbitrariness of the narrative, that through the endless recombinations of action and consequence by means of psychological explanation: "Any sentiment whatever can for that matter, at the level of novelistic psychology, justify any sort of behavior,

the determinations here being nearly always pseudo-determinations," and pseudo-determinations affect a pseudo-*vraisemblance* (247). Conforming to the implicative structure that makes a *vraisemblable narrative* possible by unspoken reference to a *vraisemblable* system to which it is attached, psychological justification and explanation emerge as the all-too well suited, silent, invisible means to a—indeed, any—narrative end. Genette continues:

> [A]nd everything happens as if Balzac, aware yet troubled by this compromising liberty, had tried to dissimulate it [that narrative liberty] by multiplying almost at random the use of 'because,' 'since,' 'thus,' and all those motivations that one wants to call *pseudo-subjective* (as Leo Spitzer called 'pseudo-objective" the motivations attributed by Charles-Louis Philippe to his characters), and the suspicious abundance of which can only emphasize for us, in the end, what they wanted to disguise: *the arbitrariness of the narrative*. (247)

What results from Genette's analysis of this "*in-between* position, of the *half-artful narrative*" is a transformation of the conceptual framework of *vraisemblance*. This transformation bears a significant connection to the analysis of critical terms accomplished by Sedgwick in the introductory "Axiomatics" section of *The Epistemology of the Closet*, where she proposes an alternative diagnostic of "universalizing" and "minoritizing" as conceptual terms to describe, evaluate, and work with theories of sexuality that would otherwise recirculate the nature/nurture analytic impasse. It is perhaps not a coincidence that in "Vraisemblance and Motivation" Genette draws upon his previous essay "Valéry et l'axiomatique litteréraire", translated by Gorman as "Paul Valéry: Literature as Such."

The *invraisemblable/vraisemblable* binary judgment, critical or popular, of a literary narrative is shown to designate the quality of implication the work sustains for itself, and the two ends of Genette's roughly sketched 'scale of narrative' no longer carry a 'possible' vs. 'impossible', real vs fictional. Balzac demonstrates this narrative function to be independent from the "public opinion", the "body of maxims and presuppositions" that sustains the *vraisemblable* narrative's "highly perceptible charm" (242). Amidst

this general arbitrariness of the narrative field, all narratives function under the same conditions, the "realist illusion" of the *vraisemblable* is a produced effect, and so too is the *invraisemblable*.

The functional description of literature Genette offers by the end of his essay presents us, in turn, with a proposition on how to reconcile the contradiction between recognizing the condition of a general arbitrariness of the narrative and simultaneously acknowledging that the narrative "appears to the reader as so many mechanical determinations" that have "not been produced by the narrator" (251). Determined by the author's decision, the narrative elements presented "in the most unilinear story" can also exist alongside "story forms the finality of which is exercised not by linear sequencing, but by a cluster of determinations" (252). How then to understand the connection Genette is developing between the 'arbitrariness of the narrative' and language itself?[68] As set in motion by the Russian Formalists, "the term 'motivation'" now designates "the way in which the functionality of a narrative's elements is dissimulated under a mask of causal determinations: so that the 'content' can only be a motivation" (253). Genette fleshes out the implication as follows:

> Motivation then is the causalist appearance and alibi that is given to the finalist determination that is the rule of fiction: the *because* appointed to make one forget the *why?*—and so to naturalize, or realize (in the sense of: to make pass for real) fiction while dissimulating what has been 'pre-arranged' in it,..., in other words what is artificial—fictive, in short. (253)

We have seen that the *vraisemblable* narrative effect is sustained through fictive means, so how does this fact impact the broader argument against the interpretive strategies of Cognitive Literary Studies? In a pair of footnotes to the passages cited above, Genette warns us against the easy mistake: "We should not confuse function and intention" when gauging or evaluating the standing of a narrative element, that is, given the

---

[68] "Thus we have called the arbitrariness of narrative here its functionality, which may rightly seem a badly chosen name; its *rasion d'être* is to connote a certain situational parallelism between narrative and language" (253).

general arbitrariness of the narrative the narrative elements have a "*functional overdetermination* that is always possible (and always desirable)" (256, 252). A narrative "function may be to a large extent be involuntary," dictated by the author's choice in a different part of the narrative sequence, a deliberate calculation, or even 'appreciate' a functional value by way of a kind of narrative 'interest': "and an intention may be unfulfilled, or exceeded, by the actuality of the work" (256 n22).

Genette embraces economic terminology, and following this path we can see a kind of motivational 'inflation' that cause a narrative element's value to 'depreciate' to the degree it is made to account for itself. And 'accounting for itself' is precisely what Balzac's narrative does so efficiently; distinct from that, the *vraisemblable* narrative behaves as a type of narrative derivative that makes nothing itself but accrues value albeit unpredictably. This derived value is the motivational "alibi" invoked by the narrative in the transformation of determination into "a (natural) relation of cause and effect" (253). Again, continuing valuably in the footnote: "The importance of the alibi is obviously variable. It is at its maximum, it appears, in the realist novel at the end of the nineteenth century. In the earlier periods (antiquity and the Middle Ages, for example), the cruder or more aristocratic state of narrative hardly disguised its functions" (256n25). This 'implicit contract' embedded in the narrative persists in the written element and yet does not predictably determine its value. Citing the expert on Russian Formalism Tzvetan Todorov: "This certitude as to the fulfillment of the foretold events profoundly affects the notion of plot. The *Odyssey* contains no surprises; everything is recounted in advance, and everything which is recounted occurs….[W]hat does 'our' plot of causality have in common with that plot of predestination which belongs to the *Odyssey*?" (256n25).

The question Genette raises regarding the relation of narrative 'plot' to narrative surprise shows a remarkable structural disconnect, that is, that there cannot be an easy equation of narrative surprise with the reader's awareness of the narrative events. The value of surprise is generated by different means.

"'Are you not really a behaviourist in disguise? Aren't

you at bottom really saying that everything except

human behaviour is a fiction?'—If I do speak of a fiction, then it is of a *grammatical* fiction."

— Wittgenstein, *Philosophical Investigations*, § 307

Section 3
Fodor

Reflecting on his choice to write a book about the work of David Hume, cognitive scientist Jerry Fodor explains his purposes to his readers in the opening section of book as follows:

> So, then, what kind of book does this purport to be? Well, from the beginning the main reason I've cared about Hume's account of the mind was that it seems, in a number of respects, to anticipate the one that informs current work in cognitive science. And the reason I care about cognitive science is that the theory of the mind that it proposes (perhaps I should say the family of theories of the mind that it proposes) is, I think, the best cognitive psychology that anybody has thought of so far. (Fodor 2)

Plainly, Fodor's manifest purpose here is to strengthen the connection between Hume's work on cognition and the 'current work in cognitive science,' in which Fodor is fully invested. Fodor continues:

> It could even be that parts of it are true. At the least, I take it to be a much more subtle theory than it's often said to be; its polemical resources are considerably richer than the sorts of objections that philosophers have brought against it might suggest. I've spent most of my professional career trying to understand how this type of theory works, and what kinds of things it can do, and what type of things it can't. It seems to me that thinking seriously about our theory of mind in relation to Hume's might help with this project." (Fodor 2)

Citing Fodor's description of his own project here efficiently demonstrates the principal concerns guiding his analysis of Hume's thinking on cognition, while also providing some indication of the context[69] and questions to which Fodor himself is responding. First, Fodor outlines a *propositional* relationship between cognitive science and cognitive psychology, that it is on the basis of the former that the latter is constructed.[70] Second, and perhaps more difficult to discern in this passage alone, Fodor's interest in Hume specifically stems in part from Hume's particularly peculiar standing in the History of Philosophy, partially celebrated by contemporary philosophers for his commitment to an empirical methodology of philosophical inquiry and partially disavowed for his "unshakable attachment to the Theory of Ideas" (Fodor 8). Third,[71] Fodor asserts that Hume's account of the mind "seems, in a number of respects, to anticipate the one that informs current work in cognitive science," and thus suggests the possibility that a return to Hume will have a corrective value, that is, that "thinking seriously about our theory of mind in relation to Hume's might help" those of us "trying to understand how this type of theory works, and what kinds of things it can do, and what types of things it can't" (2). On this account, Hume offers both defensive and polemical resources with which to push back against philosophical challenges to cognitive psychology as well as the means to question contemporary or rival schools of cognitive psychology such as connectionist models of the mind.

The second point mentioned above sketches out one of the primary angles by which Fodor's critique of contemporary philosophy of mind will proceed, and lays the groundwork for how we are to/should understand his intervention into—or correction of—the developmental path along which philosophical

---

[70] "And the reason that I care about cognitive science is that the theory of the mind *that it proposes* (perhaps I should say the family of theories of the mind *that it proposes*) is, I think, the best cognitive psychology that anybody has thought of so far" (my emphasis, Fodor 2003, 2).

[71] "Back (*way* back) when I was a boy in short pants and graduate school, there was a substantial philosophical consensus about how to read Hume; or, more precisely, about how much of Hume is worth the bother of reading. According to the understanding that then prevailed, the historical Hume had been subject to a misapprehension, characteristic of his time (come to think of it, of all times but our own) as to the nature of the philosophical enterprise" (Fodor 2003, 5).

theory of mind has progressed from its beginnings, Fodor asserts, with Hume and with Descartes before him.[72] What makes Fodor's work of particular interest is his exploitation of Hume and his subsequent transformation of Hume's theory of cognition, as an argument against contemporary critics of cognitive psychology, which he indexes by way of a quick reference to Wittgenstein and Ryle. Of the many issues Fodor identifies in "the Wittgenstein / Ryle tradition" of philosophy of mind is "its inability even to make sense of such notions as that of a *mental process*" (emphasis Fodor's, 10).[73] The Theory of Ideas, Fodor argues, provides Hume "with a framework not only for raising diachronic, 'genetic' questions about where our ideas come from, but also etiological questions about how our "perceptions" interact in the course of synchronic psychological processes like perceiving and thinking" (10). The theory of 'mental processes' is fundamentally crucial, and in reassessing its value for Hume, Fodor makes the wider import of this loss explicit:

> When [the Theory of Ideas] went out of fashion, psychology and philosophy ceased to offer theories of mental processes. This was a historical watershed. Dewey, Quine, and Ryle (for example) are in various ways modern heirs of Hume's empiricism, but not of his psychology. Hume was interested in thinking, but Dewey, Quine, and Ryle weren't. (141n4)

---

[72] cf. Fodor 2n1

[73] Wittgenstein persists throughout Fodor's book as the touchstone figure of the problem that, per Fodor, contemporary philosophy has created for itself, that of "[t]rying to run a theory of the mind (or, anyhow, of discourse about the mind) that does without the Theory of Ideas was the defining project of such mid-twentieth-century philosophers as Wittgenstein and Ryle. In my view, they made a shambles from which philosophy has yet fully to recover" (Fodor 9). Further on in the introductory chapter (titled, *Introduction: Hume's Cartesian Naturalism*) Fodor parses this 'misdirection' a bit further: "On the other hand, Hume's Cartesianism is, on the face of it, incompatible with the pragmatism about concepts that analytic philosophy learned not just from Wittgenstein and Ryle, but also from Sellars and Dummett (to say nothing of Dewey and Pierce), according to which, as we've seen, concepts are individuated by their *function* in some proprietary sense of that notion" (16). In the footnote to this comment he extends the claim to include continental philosophy as well: "I suspect that much the same view is held by such Continental icons as Heidegger. But finding out for sure would require reading them, which I intend to continue assiduously avoiding" (16n11).

Cf "It's the main thesis of this book that Hume was well advised not to have been Wittgenstein. Hume's representational theory of mind, though it needs to be purged of his empiricism, is much nearer to being right about the mind than Wittgenstein's pragmatism (than anybody's pragmatism, come to think of it)" (84).

Against the inability or unwillingness of this 'Wittgenstein tradition' to address the question of mental processes, Hume's approach has, Fodor writes, "a diagnosis on offer" for the underlying issue. "The constituents of mental processes," begins Fodor's Hume, "are causal interactions among the very sorts of things whose existence Wittgenstein, Ryle, (and Stroud) are committed to denying, namely, causal interaction among the 'ideas' that The Theory of the Ideas purports to be the theory of" (Fodor 10). According to Fodor, Hume figures a point prior to the Wittgensteinian contamination of philosophy, where the influence of Locke and Descartes have led him to a "methodological commitment to a scientific theory of mind" (9). Further elaboration of this sentiment begins slightly earlier on the page, where Fodor describes Hume's project as to construct "an empirically adequate cognitive psychology" or later on that same page, as "an independently warranted empirical psychology"; the subtext here is that the empiricist turn with which Hume is famously identified is in close communication with the turn to the 'scientific method' of inquiry, opposed (beginning) in this historical moment to philosophical inquiry and thus free of the metaphysical commitments that would distort his analysis.

Fodor volunteers that his interest in Hume rests primarily in the fact that Hume "holds a fairly rudimentary and straightforward version of the sort of cognitive psychology that interests [him, Fodor]," but this aspect of Hume's work, "the basic structure of this kind of theory has gotten increasingly hard to see" as a result of the reception of Hume's work over the intervening "couple of hundred years since he wrote the *Treatise*" (2-3). What further distinguishes Hume in Fodor's reading of his account of cognition, is that Hume is, Fodor asserts, "often remarkably perceptive about what would nowadays be called the 'architecture' of psychological theories of cognition" (Fodor 3). This point will be expanded in the following sections, but it is worth emphasizing now as it highlights the lines of my own (and Fodor's, for that matter) critique of Cognitive Literary Studies. Fodor helps give the phrase some heft, and defines it for us in a short note that also foreshadows Hume's distinctive contribution:

> The phrase 'cognitive architecture' is evocative," Fodor writes, "but not particularly well defined. It means something like: the census of entities and properties that a theory of

> cognition postulates (explicitly or otherwise) in the explanation is affords. Pylyshyn [himself a frequent collaborator with Fodor] says that the architecture of a cognitive system 'includes the basic operations provided by the biological substrate...as well as the basic resources and constraints of the system, as a limited memory. It also includes...the 'control structure'. (Fodor 3n2)

Before closing his Prologue, Fodor makes one last note which is significant for properly understanding the resources that he sees in Hume's work, the resources that have, as he put it earlier, "gotten increasingly hard to see" since being first introduced. Fodor notes that he will "have nothing much to say about Hume's skeptical epistemology" in the pages that follow, and thus initiates a move that will come to be decisive for Fodor. He continues "I think (and I think Hume did too) that, insofar as it's about the analysis of justification and the like, epistemology hasn't really got much to do with psychology" (Fodor 4). Fodor announces this at the end of his "catalogue of caveats" without much additional fanfare, but this exception will ultimately prove to be a crucial discrimination, allowing Fodor to retrieve from Hume the theoretical resources on cognition without also having to reconcile with the aspects of Hume's work that have proven the most problematic in its historical reception, especially by the twentieth-century philosophers such as Wittgenstein, Fodor takes as emblematic of obscure and incoherent arguments against the pillars of cognitive psychology.

Fodor's continual return to the figure of Wittgenstein indicates more than a personal distaste. As will become more clear in the closing chapter of Fodor's text, 'Wittgenstein', 'Wittgensteinians', and '(neo-)Wittgensteinians' serve to index two major objections to the theory of mind Fodor cares about; first, as to the ontology of ideas and mental processes; second, as to the theory of language and grammar that underwrites the "representational" or "computational" theory of mind assembled by cognitive psychology. Largely, Fodor will avoid a direct engagement with Wittgenstein's writing and instead houses these objections under the subject headings of "atomism" and "pragmatism" respectively.

These are, however, no small stakes, as what is at issue here is the broader question of 'concept

acquisition,' and specifically of the status of complex concepts, which opens onto further questions regarding the process by which Hume theorizes how "impressions" relate to simple concepts, a process, so Fodor claims, that results from Hume's epistemological commitments. The center of Fodor's intervention is found precisely at this point and he does not shy away from it: his project is "to abstract from the aspects of Hume's theory of mind that are dictated primarily by his epistemology" through the substitution of a different mechanism or model of representation, graphic for linguistic, "Old lamps for new" (Fodor 33).[74] Thus is the fine line Fodor has to walk:

> I take it that Hume is both a Methodological Naturalist (in Stroud's sense of someone who is committed to developing an empirically defensible theory of the mind) and a Cartesian Representationalist (he holds that concepts are mental particulars that serve to represent things in thought; and that having a concept is being able to think about whatever it's the concept of). My view is that, so regarded, Hume is remarkably preceptive and remarkably prescient about the architecture of such theories; in particular, he's exceptionally good on *what else* you have to do if you want to run Cartesian Representationalism as an empirical option in cognitive psychology (26-27).

In the pages that follow we will see Fodor's steady reshaping of the empiricism that shall not speak its name, as he works Hume into shape around a linguistic model of representation compatible with the syntactic operations of a computational theory of mind.

---

[74] The prologue to Fodor's book carries the subtitle "Old Lamps for New" and is a reference to a short article by Charles Dickens titled "Old Lamps for New Ones" published in *Household Words* in 1850, in which Dickens responds to the painting "Christ in the House of His Parents" by John Everett Millais, one of the founding members of the Pre-Raphaelite Brotherhood. The title of Dickens's article is drawn from the story of Aladdin in *The Arabian Nights*, but reverse the syntax of the magician's call ("New Lamps for Old ones") when he sought to obtain possession of the magical lamp.

"So, this is Whig history if it's any kind of history at all."

— Fodor, *Hume Variations*

Whig History

As Fodor deploys Hume against the philosophers who have revived him,[75] his analysis also enacts an interpretive transformation of Hume's argument, and, as a result, a transformation of the history of cognitive psychology and its philosophical precursors. This analytic tactic is congruent with Fodor's broader interpretive strategy carried throughout the text, which is to find philosophical friends of cognitive psychology in the history of philosophy so as to establish and strengthen an almost 'genealogical' relation to philosophy (of the mind). Thus Fodor announces in the conclusion: "Hume's *Treatise* is the foundational document of cognitive science: it made explicit, for the first time, the project of constructing an empirical psychology on the basis of a representational theory of mind; in effect, on the basis of the Theory of Ideas" (134). Fodor's maneuver is not, by any means, covert, and the payoff for annexing Hume to the historical development of cognitive psychology is that it secures for its proponents a fund of critical resources with which to defend themselves against the objections of contemporary philosophers and philosophical accounts of the mind.

Thus the 'nativist' tradition in psychology which, broadly speaking, theorizes that humans are born with 'cognitive modules' that enable them to learn and acquire complex skills such as language, and with which Fodor's work is strongly identified along with the work of Noam Chomsky and Steven Pinker, finds itself in a surprising relation to one of history's most famous empiricists. This is, Fodor acknowledges, "Whig History if it's any kind of history at all" (3).[76] Why 'Whig History'? The answer is that Fodor's move

---

[75] "I'll concentrate mostly on [Barry] Stroud's reading of Hume in this introductory chapter. Both for better and for worse, it seems to have been pivotal in prompting the revival of philosophical interest in Hume's theory of mind" (Fodor 6n1). The work to which Fodor is referring is Stroud's 1977 volume, simply titled *Hume*.

[76] Fodor's use of this phrase is provocative, insofar as the term carries a pejorative connotation and identifies a approach to historiography that presents the past as an inevitable progression towards ever greater liberty and enlightenment, culminating in modern forms of liberal democracy and constitutional monarchy. David Hume himself proves to be a crucial figure in Whig History itself, as his work *The History of England* is considered to be one

here—and which he carries on through his text on Hume—highlights the stakes of an internal rift within so-called nativist cognitive psychology between what, to borrow Tomkins's useful diagnostic distinctions, may be described as a 'weak' and 'strong' nativism. Within the field of cognitive science and cognitive psychology, psychological nativism, to recall from earlier discussion in the introduction, and particularly (Fodor's) 'modern' psychological nativism, is itself much of a historical artifact of the twentieth century, insofar as its 'contemporary' lineage is shaped by a response to and critique of Behaviorism that began in the mid 1950s, enabled by developments in electronic computer technology and newly developed theories of grammatical syntax.[77] The invocation of 'Whig History' here carries with itself a set of complex implications, as Hume's role within Whig History is significant for the fact that he broke with prior Whig historians in acknowledging the fact of the revolution which others had attempted to excise from the historical record. Much, that is, in the same way as Fodor makes use of Hume to legitimate current cognitive psychological theories of mind.

As Fodor will corroborate explicitly in the concluding chapters, ("Imagination" and "Conclusion: Hume's Program (and Ours)") the question lurking in the wings of his revival of Hume's *Treatise* is how to reconcile the apparent incompatibility of his claim that Hume anticipates the field of cognitive science and psychology with the aspects of Hume's theory of the mind that are considered fundamentally untenable by those theoretical heirs. Fodor has spent much of the text thus far giving careful consideration of them: associationism, the copy theory, and his position on innate ideas.[78] Fodor boils down the issue by framing the question about Hume's problems through "the argument between nativism and empiricism [which] is about where ideas come from" (112). This framing proves decisive, preparing for Fodor the

---

of the defining examples of this historiography, which he wrote with the intent to legitimate the Glorious Revolution of 1688 from which the modern constitutional government of England was established.

[77] The use of the term 'constituents' in Fodor's breakdown of the units of mental language is yet more telling. In a preceding passage from the same article, Fodor makes reference to Noam Chomsky's review of B. F. Skinner's *Verbal Behavior* as initiating the "'paradigm shift' in theories of the cognitive mind," now known as the 'cognitive revolution' of the 1950s (Fodor 2006); Chomsky's *Verbal Behavior* (1957), George Miller's "The Magical Number Seven, Plus or Minus Two" (1956).

[78] "I take it, for example, that Hume was seriously wrong about innateness" (112).

means by which he can execute a neat separation of Hume's empiricist impulses, and thereby to strip away his psychological theory from his epistemological obligations that have become so problematic. This light touch of Whig History is nonetheless extremely fortuitous, as it enables Fodor to bracket the steady accumulation of questions as to the compatibility between Hume's empiricist commitments and the psychological nativism that underpins "our current cognitive science" (113). Here is Fodor's synopsis:

> Theories of cognitive architecture are primarily about the *synchronic* structure of the mind, so they can often be more or less neutral on questions of ontogenesis[79]; once the stuff gets in, it doesn't much matter how it got there. One could thus imagine plugging some variety of cognitive nativism into Hume's representational realism, leaving much of the rest of it intact. In effect, our current cognitive science does so. (112-113)

The casually diagnostic tone that Fodor strikes here while describing the "more or less neutral" mechanics of psychological theories is meant to underscore how "Hume's rejection of innate ideas" can be made compatible with cognitive nativism. The separation of the epistemological question as to where ideas come from, from the question of how those ideas interact in "the *synchronic* structure of the mind" is crucial for Fodor,[80] and it is a process he has had in motion since he began with Hume's account of impressions[81] in the second chapter.

> "Hume thinks that there are two kinds of mental particulars, 'impressions' (roughly = sensations) and 'ideas' (roughly = concepts). This sensation/concept distinction does a lot of work for Hume. For example, it both explicates and underwrites his empiricism. Hume

---

[79] cf. 85, "Hume thinks that what bounds the population of simple concepts is their etiology

[80] It is also crucial for CLS as well, the way that their argument is streamlined requires shedding the responsibility for answering these epistemological questions.

[81] Hume maintains that ideas (for Fodor "perceptions") are *copied as pictures* from impressions (for Fodor, 'raw sensory data'), but Fodor argues that Julesz's experiments demonstrate "that there are aspects of sensory representations that carry unconceptualized information" (48).

> holds that simple ideas come from impressions, and that complex ideas reduce without residue to the simple ones that are their constituents. The claim that the concept/impression distinction is exhaustive thus implies that there is *nothing at all* in the (cognitive) mind except sensations and what is 'derived' from them" (28).

This is a pivotal moment for Fodor insofar as Fodor wants to maintain a healthy degree of skepticism regarding where the structure of concepts comes from, because this keeps alive the possibility that Hume can provide a philosophical pedigree for Fodor's nativism[82] and the computational laws of mental processes. "In short, it's important to Hume's epistemology whether ideas are just copies of impressions. But the overall architecture of his cognitive psychology needn't really much care. If the structure of concepts doesn't copy the structure of sensations, then it doesn't" (42). The prospect of accepting this problem in Hume's theory, has quite a strong appeal for Fodor himself, allowing him to keep his house in good order:

> Ideas and impressions can still be mental particulars, and perception can still be a process that starts by receiving impressions and ends by applying concepts. Take away Hume's empiricism, and his motivation for the copy theory goes too. Take away the empiricism *and* the copy theory, and what's left is a perfectly stand Representational Theory of Mind, one that's compatible with as much (or as little) nativism as the facts turn out to require. (42)

Yet the stakes at issue here will come to be decisive when it comes to the question of linguistic ambiguity and equivocation, both of which challenge Fodor's account of the mind and its cognitive architecture and derive from Wittgenstein's shadowy influence. In severing the empiricist tie that bound

---

[82] "But if the structure of ideas isn't copied from the structure of impression, where *does* it come from? Why, for example, mightn't it be innate?" (Fodor 42)

Hume's psychology to his copy theory epistemology, Fodor has overplayed his hand. As will be apparent, Fodor's psychology remains structurally dependent on a metaphysics of presence to resolve the problems of writing that continually warp the linguistic model of representation that he proposes instead of the picture-copy theory.

The interpretive transformation Fodor has underway of Hume's work picks up pace as the argument approaches the challenges introduced by writing, as Fodor begins to extract from Hume's theory the aspects of it that are most compatible with a linguistic model of representation. As we have seen in the last section, Fodor's impulse is motivated by the broader goal to eliminate the image-picture-copy theory of representation from Hume's account of the mind but to otherwise retain the 'architectural structure' through which the account itself is built. The reason for this split purpose is that Fodor recognizes a structural *composition* within Hume's theory that anticipates the syntactic requirements of the Representational Theory of Mind corresponding to current models of cognitive processes grounded by the computational model.[83]

Fodor parses the stakes of this connection well, and in doing so clearly reveals *how* and *why* the 'architectural structure' makes Hume's account of the mind appear to be a good candidate as the philosophical precursor: "Part and parcel of Hume's theory of mind," Fodor writes, "is that some concepts are complex, and some are simple, and the simple ones are the (ultimate) constituents of the complex ones. And it's part and parcel of Hume's empiricism that concepts that share the *structural* property of being Simple also share the *etiological* property of being copied from corresponding impressions and the *semantic* property of representing whatever it is that the corresponding impressions do" (56). Here we can see Fodor reverse engineering the syntax/semantic relationship required by the computational model, deploying it as an interpretive strategy by which to read Hume's account of the mind. The exception Fodor takes to the parity of the taxonomic relational structure that Hume outlines for impressions—that there

[83] Composition and Composite mingled here in a advantageous way.

are both simple and complex impressions each of which mint[84] simple and complex ideas—is foregrounded earlier in the text, in the opening paragraphs of chapter two "*Impressions*" and is itself not without caveats:

> The origin of simple ideas is unproblematic; they are almost always[1] copies of simple impressions. Likewise, according to Hume it's unproblematic where the structure of *some* of our complex ideas come from; it's copied from the structure of complex impressions. This can't, however, be the general case. For example, the structure of the concept UNICORN couldn't be copied from the structure of an impression of a unicorn; since there are no unicorns, there are no such impressions. There are, in short, complex ideas for which corresponding complex impressions are lacking. Where do they come from?[2] (29).

The issue Fodor draws attention to here cuts to the heart of the issue that Hume's theory is attempting to account for: the apparent imbalance of constituent parts for the number of complex mental representations on hand. Fodor's notes to this paragraph corroborate that the main concern lies with complex concepts; note one reads: "The caveat is on account of the notorious 'missing shade of blue' (see I.I.I, 53) Hume says this execution isn't serious enough to bother about. For present purposes I'll assume so too" (29n1). Whereas note two gives the following: "Hume is aware that, whereas one's experience is finite, there are indefinitely many complex ideas. The productivity of complex ideas, all by itself, requires that some of them must not derive from impressions" (29n2). The issue for Hume is the breakdown of the relay by which impressions must warrant concepts, but Fodor understands the lopsided ratio of simple and complex constituents differently, influenced by Chomsky's theory of generative grammar. The barrier to this transposition is, however, the photographic model of representation.

In explicating the property of a concept "being Simple", Fodor's interest is guided by the what he terms its "*semantic* property of representing whatever it is the corresponding impressions do" (56).[85] Once

---

[84] Or one might say, 'token' these ideas, in a reversal of the direction that Fodor authorizes later in the text.

[85] It is not insignificant that Fodor collapse here the nested structure of representation by which simple concepts

the conversation advances to the relation between simple concepts and complex concepts, the property of "being Simple" is more clearly defined as a constituent or composite relation: "simple [concepts] are the (ultimate) constituents of the complex [concepts]" (56). Here is Fodor's first gloss of this property which signals the shift toward a linguistic model of representation:

> Being simple is just *not having semantically evaluable parts*. It's analogous to, say, the property that two words share if both are monomorphemic. Now, we don't, of course, expect structurally simple *words* to have anything much in common except their simplicity of structure…. By contrast, it's pretty generally assumed that structurally simple *concepts* are homogeneous in all sorts of important ways. Hume, of course, does assume that; it's required by his empiricist semantics, according to which simple concepts must all be sensory. (57)

Looking ahead to complex concepts, and to the syntactic requirements they have, Fodor is advancing here the means by which it can be palatable to separate the synchronic, "architectural structure" of mental particulars from their 'contents'[86]; around this jointure of the "not having semantically evaluable parts" Fodor distills a computational grammar from out of the taxonomy of mental particulars. Doing so enables Fodor to proceed through the impasse for which Hume required the intervention of the imagination:

> "In short, there's no end to the things that one can think of. But since the population of simple concepts is fixed, there is an end to the things one can think of by thinking *them*. So the concepts that are productive (i.e. the ones of which there are infinitely many) mustn't be simple; which is to say that they must be complex. Fine so far" (85).

represent simple impressions which represent "whatever" they do.

[86] The path Fodor is following here is analogous in a way to the one first cleared by Saussure, in that he is attempt to fix or stabilize a 'unit' of cognition according the model of a 'word'. The similarity ends quickly as Fodor does not see simple concepts to be composed of constituent parts in the way that words are composed of letters.

Fodor proposes that Hume's use of the faculty of 'imagination' or 'fancy' to address "the untenable assumption that ideas that aren't 'derived from' from impressions can't be bona fide" is the result of a (axiomatic?) limitation inherent to associationism (112). This limitation is that associationism is, per Fodor, unable to parse the difference between "intentional generalizations" about belief/behavior and the "taxonomy of intentional states"; that is, equipped only with the law of association, Hume's theory of mind is unable "to distinguish *the intentional relations among the contents of thoughts*, from *the causal relations among the thoughts themselves*" and thus "Hume is forever counting on the imagination ('fancy') to get him out of [this problem]" (113-114). This problem is resolved in current cognitive science and psychology, thanks to Turing's discovery and theoretical elaboration of computation:

> We, however, are better off. We don't need imagination because we don't need association. And we don't need association because Turing showed us how to replace it with computation. And computation is able to operate, not just on the associative relations among thoughts, but also on the mental representations that specify their intentional contents" (115-116).

The issue that Fodor believes to be resolved by the swap of imagination for computation is the question of the 'upper limit,' that is, the ability of the mind to generate an infinite number of complex ideas with a finite number of sensory impressions. The concern here is in part driven by a 'general' pre-occupation on Fodor's part with the question of complexity in a general sense. And this is in contrast to Hume, whose concern remains on the 'lower limit', the empirical level, with, as Fodor terms it, the etiological question of where simple concepts come from:

> "It bears emphasis that Hume's tactic of arguing from the productivity of conceptual repertoires to the postulation of complex ideas does *not* require his empiricism. To be sure, Hume thinks that what bounds the population of simple concepts is their etiology; they have to be derived, one by one, from experience, since "we cannot form to ourselves a

> just idea of the taste of the pineapple, without having actually tasted it" (I.I.I, 53). But this argument could dispense with its etiological premise if it were so inclined. Suppose, for example, that some, many, or all of our simple concepts are innate. Still our minds are finite, so there must be an upper bound on how many primitive [simple] concepts we can entertain. So, if our concepts are productive, they can't all be primitive. So some of them must be complex" (85-86).

And dispensing with the empirical, that is, *etiological* premise is exactly what Fodor proposes to do, but only insofar as complex concepts are concerned. The issue of conceptual productivity that is emerging (or re-emerging) here as the perpetual topic at hand, goes together with the question of etiology because it necessitates a break in the semantic transparency required by Hume's epistemological empiricist commitments. Conceptual productivity, complexity, Generativity of conceptual repertoires, and semantic productivity, all of these processes amount to a building challenge to the empiricism etiology underwriting the atomistic, individuated mental particulars composing or constituting complexity in a general sense.

This is the element (the *elemental* unit) that Fodor faithfully retains from an otherwise enthusiastically supportive survey in "Simple Concepts": "Still," Fodor writes, "there's at least one apriori connection between the structure of concepts and their ontogeny[87]: by definition, unstructured concepts have no constituents; a fortiori, they can't be learned by assembling them from concepts that were previously available" (82). That is the rub, "a concept can't be *both* simple *and* composed" that is, cannot be composed of constituent concepts because this compositional structure is precisely the specification of complexity, *not* simplicity. What results from this analysis is that Fodor has cleared a certain amount of space to serve as a foothold for the nativists to share with the empiricists. Declining to attempt to adjudicate the question of the etiology of simple concepts, Fodor instead suggests that Hume's account of the mind, despite his

[87] This may be the first time in the text that Fodor articulates this as 'ontogeny' as opposed to etiology, as he does when describing Hume's concern. His use of 'ontogeny' begins a few pages earlier.

declared allegiance to the legacy of Lockean empiricism[88], remains not only fully compatible with the possibility that some ideas are *innate*. Such a feat is prepared for, and the redefinition of 'innateness' is support by Fodor in a rare excursion from Hume's *Treatise*, what troubles Hume's account of simple concepts likewise troubles his account of simple impressions:

> Notice that, from Hume's point of view, much the same question arises about where simple *impressions* come from (see Chapter 2 above); by definition, they can't be composed and, on pain of regress, they can't be derived from other impressions. Hume says: "The examination of our sensations belongs more to anatomists and natural philosophers than to moral; and therefore shall not at present be entered upon" (I.I.2, 55). He also says, in the *Enquiry*, that "understanding by *innate*, what is original or copied from no precedent perception, then we may assert that all our impressions are innate" (Hume 1994: 68). And so say I. (82)

As I indicated at the outset, the reason that I've chosen to give Fodor's work such a close examination is that Fodor's theoretical work in cognitive psychology is frequently relied upon by Cognitive Literary Studies. This reliance is, however, *is not* to be understood as authorized by Fodor's work itself, but rather results from an exploitation of exactly that question of etiology that Fodor (and Hume, for that matter) specifically excludes from the broader examination of Hume. Fodor's theorization of innate cognitive modules or mechanisms becomes the avenue by which Cognitive Literary Studies presses its claims forward. Fodor, for his part, makes clear that the innate components that figure in his cognitive architecture *does not* authorize *nor* legitimate such adventures. "So, perhaps surprisingly, even if the lack of conceptual structure is epistemologically or semantically neutral," in other words, that simple mental particulars (either concepts or impressions) which lack a composite, *complex* structure must not "also be homogenous in other

[88] "Lockean empiricism (roughly, the thesis that there are no innate ideas)" (Fodor 83).

theoretically important respects" (81, 83). The status of this 'neutrality' is precisely the opening exploited by Cognitive Literary Studies, and Fodor's rebuke is worthy of quoting in full here:

> So, perhaps surprisingly, even if the lack of conceptual structure is epistemologically or semantically neutral, it's said to be rife with *phylogenetic* consequences. The rationale for such claims is often that the innate ideas can include only what would have been good for our ancestors when they hunted and gathered back on the primordial savannah; a pop-Darwinist scenario for which, however, there exists nothing to speak of by way of evidence. There may be—perhaps, indeed, there must be—biologically interesting constraints on what concepts human minds can have innately. But if there are, none of them are known as of this writing. "CARBURETOR can't be innate" may, for all I know, be true; but, evidentially speaking, it's ethology by mere fiat. (83)

The other aspect of psychological nativism's umbrella that Cognitive Literary Studies takes coverture under, and which is of particular importance here, is the 'computational' theory of mind to which Fodor's contributions are many. The connection that Fodor is so interested in maintaining between Hume's work and current psychological theory consists in its 'architectural' structure, but also as we've seen, in the rules of its operation that Hume outlines. This characterization of the structure of complex concepts as 'architectural' describes an operational logic of computation as a grammatical syntax. It is Fodor's use of linguistic theoretical resources as the model for mental representation that establishes a conceptual vocabulary that is 'overly fit' to the critical program of Cognitive Literary Studies, and thus is easily and readily exploited beyond the very limits that he has loudly insisted on.[89]

The reading of Fodor that we have been tracing thus far has returned us, at this point, to the larger questions of the dissertation; that is, broadly put, to what extent the theoretical resources of cognitive

[89] Cf. Fodor, Jerry, and Massimo Piattelli-Palmarini. *What Darwin got wrong*. Profile books, 2011.

science and psychology as they are deployed by cognitive literary theorists as reading strategies for literary interpretation, to what extent these theories address the questions raised by writing, to what extent they address the questions raised by *fictional* writing, and to indicate how and why the reading strategies elaborated by Cognitive Literary Studies are of limited value. Following Fodor's reading of Hume has enabled us to trace and to outline how nativist cognitive psychology is related to empirical cognitive science, and what this revealed, indeed by Fodor's own analytic light, is that the claims of empirical warrant made by Cognitive Literary Studies are not supported by the cognitive psychology they invoke.

As I've signaled many times above, Fodor's purpose in his text on Hume has never been to mount a defense of empiricism but was instead to "[plug] some variety of cognitive nativism into Hume's representational realism [while] leaving much of the rest of it intact" (113). How this transformation is completed, Fodor proposes in his penultimate chapter, is by updating the image model of representation available to Hume with the linguistic model of representation that is congruent with the composite structure of concepts: "Even so, Hume was right about his most fundamental architectural claim: there must be simple concepts and there much be mechanisms (for Hume, association and imagination; see Chapter 5 below) that are able to construct complex concept from them" (83). For Fodor, these mechanisms would be replaced by syntax and computation, what Fodor will term 'compositionality' as a term that describes the way concepts "are distinguished either by what constituents they contain or by how their constituents are arranged, *or both* (my emphasis, 87).

With the compositionality thesis freshly in hand, Fodor's main concern now returns to two specific obstacles; first, the challenges articulated under the figure of Wittgenstein as "lexical ambiguity" and "equivocation"; second, and bound up with the first in a crucial yet distinctly different way, is the explanatory structure by which Fodor proposes to dispense with these 'Neo-Wittgensteinian' objections. As regards 'ambiguity' Fodor notes that "there is a (neo-)Wittgensteinian recently in evidence according to which the content of all representation is intrinsically context-dependent" (96). Such a proposition spells trouble for the careful exit from the discussion of etiology that Fodor has just completed, so it is of

no small importance to address this issue as quickly and efficiently as possible. Indeed, given that this question was supposed to have been sidelined, it's of not insignificant interest exactly *how* this question became a question again at all? Fodor continues on a bit further:

> Since, according to this view, all content is content-in-context, the content of simple concepts (and the content of monomorphemic words) is too. But the compositionality thesis says that the content of complex concepts (and complex linguistic forms) is inherited from the context-*independent* meaning of their constituents. So either content isn't inherently contextualized or there's something wrong with the compositionality thesis. (96)

Slowly creeping back into the frame of Fodor's argument is a differentiation between types linguistic elements, "monomorphemic words" and "complex linguistic forms", that was first given as rough articulation of the spectrum of conceptual complexity previously carved up by Hume into 'simple' and 'complex' segments. The problem that this poses for Fodor is that indivisible atomic unit of the "monomorphemic words" which are taken without further consideration to be unproblematically—that is—*transparently* representative of "whatever it is they represent" out in the world, are *the* model for simple concepts—and simple concepts, by definition, are not composites: "Hume's psychology defense of his empiricist epistemology consists of the claim that the content of *simple* concepts is empiricistic (they just copy experiences), together with the assumption that compositional processes are semantically transparent (they add nothing to the content of simple concepts when they join them together into complex ones)" (95). These monomorphemic words/simple concepts are supposed to be, *required* to be "context-*independent*" but with the substitution of the linguistic model of representation for the graphic model required by the computational model, the question of context has been smuggled back on to the scene. The analytic bracketing of the question of semantic content, supposed to have been executed by the rigorous specification of syntactic structure, is challenged by the "pragmatically (neo-) Wittgensteinian idea" that "content is interpretability in context" (97).

The reason why this moment in Fodor's reading is so significant for the purposes of my own analysis of Cognitive Literary Studies, is that the premise now under consideration has the potential to upend the neat analytic separation of conceptual content from conceptual structure that has guided his interpretation of Hume. We have carefully traced Fodor's motivation to purge the empirical commitments from Hume's psychology, and here the equation of mental representation and linguistic representation is almost complete. Responding to the "(neo-) Wittgensteinians" 'pragmatism' Fodor writes: "First, when compositionality is the issue, it's essential to distinguish two quite different things that may be intended by claims that the content of (mental and/or linguistic) representation is *ipso facto* contextualized" (97). In what follows it is important to note the seamless suppression of the graphic resources of typography that Fodor deploys as neutral tools of explanation:

> On the one hand, there's the sort of view sketched out above: that, the exigencies of compositionality notwithstanding, the content of complex representation is metaphysically prior to the content of simple ones. On the other hand, there's the 'externalist' idea that the content of simple representations is (in part or entirely) supervenient on their mode of being in the world. I'm much inclined to suppose that some sort of externalism must be true. That 'cow' means *cow* (and hence that 'brown cow' means *brown cow*) surely has something to do with how 'cow' tokens are situated with respect to cows, and cows *are* things in the world. No cow is a text.[16] We are required to be semantic externalists on pain of otherwise being semantic idealists. (97)

The very heft of the explanatory effect leveraged by Fodor is accomplished through an otherwise unremarked upon configuration of typography, punctuation marks, and monomorphemic words.[90] It's

[90] In the *Prologue*, in a footnote that specifics the edition of Hume's work that he will be citing through the text, Fodor provides the following: "Hume's technical terminology frequently co-opts non-technical expressions. *Caveat emptor*.

worth emphasizing this fact not, much like Genette's observation about Balzac's writing, that the tools of Fodor's explanation in a way tip his hand and thus invalidate the entire enterprise. In turn, what this passage confirms is the logocentric model of linguistic representation anchoring Fodor's project remains entangled with the problems of reference and of writing, and fact that these typographic differences are not interesting enough to be addressed head-on.

But this is the corner that Fodor has been backed into by his own terms. In addressing Wittgenstein (or Frege a few pages later) and the pesky, persistent problems of language their figures evoke, the argumentative tactic he employs is shaped by the one-hand-other-hand alternative sketched out above.[91] 'Sure, sure,' he seems to say, 'We all know that there's *some* degree of contextualization happening; no one is denying *that*' but there's no reason to be concerned because these "exigencies" are nothing like the impression Wittgenstein has given them!' He then promptly moves as if to demonstrate just how matter-of-fact and commonsensical it is by assembling a complex "mental and/ or linguistic representation" from monomorphemic constituent words.[92] Fodor has to address monomorphemic words not for their referent, because doing so would concede the point he is after, but for their structure as the atomic and constituent unit of complex representation[93], and the question at hand is how meaning can be guaranteed if context-

---

I adhere to the standard convention that canonical names of concepts are spelled in full caps. 'DOG' names the concept of a dog" (3n3).

[91] "I'll begin with some ground cleaning" begins Fodor's response.

[92] "…'cow' means *cow* (and hence that 'brown cow' means *brown cow*)…" (97).

[93] A fascinating implication with the weather, plausibility, and the historical period so important to the development of the cognitive sciences of the 1950s, is made by Fodor when, hard pressed at the end of his concluding chapter in a discussion of the question of atomism, that, per Fodor, ranks as one of the primary points of difference between Hume and Wittgenstein (and thus between the computational cognitivists and the (Neo-)Wittgensteinians or pragmatists: "It's pretty widely agreed that Hume's version of the Theory of Ideas is basically atomistic; and, as we saw in the Introduction, it's not unheard of to hold this against him. When Hume is berated for not having been Wittgenstein, it's generally his atomism that's the gravamen of the reproach. What's at issue here is a thesis about relations among the conditions for concept possession: You're an atomist insofar as you hold that the possession conditions for some concepts are independent of the possession conditions for any others; you're not to the extent that you don't" (146-47). As we have seen, the disavowal of atomism is problematic because Fodor needs to preserve atomism-as-conceptual-individuation in a form that serves as an additional buttress for his cognitive psychology as the condition of possibility for *inference*, and thus for the inferential learning and concept acquisition: "Now, you may think that not being compatible with atomism is a virtue in a theory of concept possession (individuation). But I don't, because I think atomism is quite likely true (at very least of nonlogical concepts)" (149). Fodor packages this

dependence indeed threatens the stability of these elemental-like monomorphemic words?

It's clear Fodor knows that it's time to face facts and he addresses the question of 'lexical ambiguity' by reference to a pun given in Mark Twain's *A Connecticut Yankee*:

> Consider lexical ambiguity: there's the 'page$_1$' that (according to my dictionary) means *one side of a piece of paper*, and there's the 'page$_2$' that means *a boy or a young man employed in a hotel to carry messages*.[18] These undergo disambiguation in contexts like 'John wrote half a page'; so one might say that what "page$_{tout\ court}$" [sic] means depends on its context. (99)

The suppression of the explanatory contribution of these diacritical marks is enthusiastically confirmed in a footnote, which elaborates: "The subscripts aren't pronounced; they go without saying" (99n19). This is indeed a bold move to take in the arena so recently and thoroughly cleared of referents, and, as Fodor himself notes, all joking about lexical ambiguity aside, he moves forward to serious matters: "By contrast, what really would undermine the compositionality of English (and, *mutatis mutandis*, of the concepts that English is used to express) is equivocation that can't be resolved" (99). The footnote to this sentence provides the full view of how Fodor understands to get himself out of the situation he's presently in, which is a novel articulation of how to gin up a backstop to 'equivocation'; he writes:

> I speak of "equivocation" rather than "ambiguity" because the right models for ineliminable context-dependence are less plausibly homonyms than (what linguists sometimes call) "polysemous" expressions; ones which can have any of a family of related

---

premise for us neatly: "Call the thesis that some inferences belong to the possession of conditions of (some of) the concepts they deploy, 'inferential atomism' (IA)," but continues on to provide us with the obstacles that, paradoxically, consist in a recourse to the empirical. He writes, "prima facie IA is committed to a variety of empirical claims which seem to be, quite simply false. The situation is egregious when one considers the question of concept *acquisition*" (150). Throwing up his hands when faced with the question of why everybody takes "for granted that IA must be true" he notes the following: "Atomism didn't seem implausible to Hume, or to hosts of philosophers who preceded and followed him (including, by the way, Wittgenstein circa 1920). In fact, as far as I can make out, atomism didn't start to be intuitively implausible on a really big scale until around 1950. What they always say about the weather is true in spades of intuitions of philosophical plausibility. If you don't like what you've got now, just wait until Monday. (151-152).

> meanings depending on their host. The ambiguity / polysemy distinction is arguably important, but not for present purposes. All we care about is whether equivocation goes on forever; i.e. whether there are any *un*equivocal expressions of finite length. (99n20)

Fodor has rightly acknowledged that when importing the model of linguistic representation, at some point, the problem comes down "whether equivocation goes on forever;" and moreover "whether the semantic facts about 'cow' explain the semantic facts about 'brown cow' or vice versa" (99, 97). That is the narrow line Fodor has to walk, his concern is to outline the means for stabilizing meaning in the face of runaway equivocation, but to also allow for just enough equivocative flexibility within language for "inference in language (or concept) *acquisition*" (98). Per Fodor, acquisition works "'top down…whereas, by contrast, compositionality works 'bottom up'" (98).

Admirably, Fodor puts the issue to rest in a manner consonant with the tightly drawn scope of his own intervention, and proposes a resolution that maintains a considerable fidelity to the field of cognitive psychology. This solution, however, does not resolve the problem introduced by writing, and it leaves the question of fiction untouched entirely. Speaking to the example set out by Frege's famous "blue ink" question, Fodor's response proposes to resolve context-equivocation by recourse to the speaker's intention:

> But one might equally suppose that the content of what's said stays the same and what changes is the point that one has in mind to make by saying it. On this latter view, "the ink is blue" means the same in every context….What's up for interpretation is the *speaker*, not the *language*; sometimes he's wanting to call attention to the blueness of the ink on the page, sometimes he's wanting to call attention to the blueness of the ink in the bottle. Context typically serves to resolve the equivocation by making clear which it is that he has in mind. (And, if the context doesn't, you can always just ask him.) (104)

Under this configuration, the linguistic model of representation is duplicated, as Fodor advances the interpretable frontier between context and the unit of representation, "the language," to the intention of

the speaker and the language spoken, "what he has in mind" (104). The effect of this move is to leave wholly intact the problems of writing, and thereby also to bypass or effectively limit the disruptive effects of writing (the runaway effect of context equivocation) through establishing a second linguistic layer of mental representation by which to disambiguate and dis-equivocate language. Dividing or redoubling the system of representation, that is, abstracting the linguistic model of representation *while also* leaving the linguistic model in place, in this way allows Fodor to set the record straight, diagnosing these troubling Wittgensteinian issues as "a confusion between the metaphysics of meaning and the epistemology of interpretation" (105). He continues on to explain how this correction is warranted, and outlines the basis of the boundary lines; Fodor is here in position of the "we":

> We think that what makes a symbol mean what it does has, literally, nothing to do with the considerations that bear on its interpretation in context.[27] In fact, we don't think that there is any such thing as the interpretation of a *symbol* in context; it's the sort of category mistake to speak that way. What needs interpretation is *the symbol's being tokened in the context*. Theories about that belong to the epistemology of communication, not to the metaphysics of meaning. Some day philosophers will stop confusing the epistemology of communication with the metaphysics of meaning. (105)
>
> 27. Barring indexicals and bona fide ambiguities like 'case' or 'flying planes'.

Notice that that what began as the split between "the metaphysics of meaning and the epistemology of interpretation" has been quietly updated to the "epistemology of communication" as Fodor off-loads almost all the issues of linguistic representation from their cognitive implementation.[94] Tracking Fodor's path here does not even attempt to expose a fact that he has otherwise attempted to conceal, that is, that

[94] It's worth mentioning in passing that here, as in many other sentences throughout the text, the "technical" tradition of denotation for mental representation mingles seamlessly with (presumably) stylistic uses of the same typography.

one of his primary contributions to cognitive psychology is the elaboration of thinking as system representative modeled on language, a "language of thought". This theory underlies his text on Hume, and, given the name "Mentalese" to preserve a resemblance to other linguistic forms natural (yet foreign) language, Fodor mentions this linguistic-cognitive 'language' without direct introduction once earlier, in the chapter on simple concepts, while again diagnosing a problem to be a defect in communication that comes near to a claim of a deficiency in natural language. There, Fodor frames the issues along classic lines, language expresses thought, and thus leaves open the question of how *well* it might be suited to the task of translating Mentalese to English: "because you can't say what CAUSE means in English (or, presumably in any other natural language) though you can say it (to yourself) in Mentalese" (67).

If it were not for this unique and privileged access granted to the contents of thought, the suggestion that there may be a linguistic structure to mental representation is not unacceptable, provided it was equipped to address the problems of written language. This is of the most importance for the very stakes of the Cognitive Literary Studies on the table, that a problem in linguistic representation can be resolved by recourse to the 'mentalese' representations of the speaker which have been translated English. The breakdown of this theory for instances of language in which the intention of the author is unavailable or not semantically exhaustive, specifically in the case of fiction and literary language, fundamentally complicates the viability of the connection—supposed or presumed—between cognition and language.

Indeed, the effect of expanding the model of linguistic representation Fodor imports into cognitive theory to *include* the problems of writing, particularly as they are outlined by Derrida in "Freud and the Scene of Writing", is a research path that I intend to pursue elsewhere: How would Fodor's theories of cognitive psychology impacted and reorganize the shape of cognitive reading strategies in the light of Melanie Klein's work? It may prove extremely productive if, following Fodor, we expanded his model of language to account for the sub-atomic letters from which monomorphemic words are assembled combined with Klein's observations: "As concerns the forms of signs, even within phonetic writing, the cathexis of gestures, and of movements, of letters, lines, points, the elements of the writing apparatus

(instrument, surface, substance, etc.)" (Derrida 231). Perhaps there are more resources to the computational theory of mind, maybe nativism is right and some ideas are innate and they institute a generative computational system, but how would computation be deployed differently if the instruction set implemented through its syntactic operations[95] was developed under a Kleinian analytic?

Throughout this project the question is to determine how to evaluate the reading strategies developed by Cognitive Literary Studies and the interpretive claims they make or enable readers to make. The reading of Genette's analysis of *vraisemblance* demonstrated that this quality resemblance that a literary fiction exploits is a carefully composed effect, an effect, moreover, that consists in the "disappearance" of the written text. Genette's analysis of this effect of the text showed how well it conceals itself under the cover of characters' behavior, of narrative action, as architectural structure of written, fictional narrative. The close examination of Fodor's annexation of Hume as the philosophical precursor to current cognitive psychology provided the rare opportunity to observe how cognitive psychology, through the elaboration of computation, proposes to have resolved the problems of representation raised by analytic philosophers.

Observing the challenges and objections that Fodor's text anticipated and responded to indicates an additional aspect of the problem, seen in the analytic energy spent working through the differentiation of simple and complex mental representations; like so much processor waste heat,[96] this energy can tell us

---

[95] When considering how the role played by Hume's 'imagination' is not longer necessary because "once you've got traces, you don't need an independent faculty of imagination to implement inductive principles" (130). "Likewise, *mutatis mutandis*, if you prefer a classical cognitive architecture [as compared to the 'connectionist' architecture considered in the sentence prior]. In that case, records of X/Y coincidences are written in whatever language the mind computes in (Mentalese, say) and are stored in locations in memory (for example, on the tape, assuming that the mind has the sort of architecture that Turing machines do)" (130).

[96] The term 'processor waste heat' describes the "heat that is produced by a machine, or other process that uses energy, as a byproduct of doing work. All such processes give off some waste heat as a fundamental result of the laws of thermodynamics." The term is defined by the OED as "heat produced as the by-product of some process"; in computer technology the term's technical specification is derived from the "transistor junction temperature" which is the upper limit to the operating temperature of the semiconductors in an electronic device." While I am using the term in a metaphorical sense here, as has been mentioned frequently in this chapter, the development of electronic, digital computer technology, was crucial to the expansion and development of cognitive science and cognitive psychology. Beginning in 1958 with the development by Jack Kirby of the 'integrated circuit', composed of electronic circuits set (or 'integrated') into a single piece of semiconductor material such as silicon. The Integrated Circuit is the basis on which modern computer processors are built, and has driven the development of computer technology by providing the technical means to process ever increasing numbers of computations required by

something more diagnostic about how a system is operating.[97] It would appear that Fodor's preoccupation with constructing a viable theoretical model to account for the mind's productivity by recourse to the discovery and implementation of computation and Hume's preoccupation with accounting for simple ideas both share in a common problem: the destabilizing effects of writing.

I argue that Derrida and Sedgwick both provide more durable approaches to this challenge as they both begin their own work with a fundamentally different model of linguistic representation than is available in the philosophical tradition on which cognitive psychology depends. In the second chapter, the detailed analysis of Sedgwick's essay on *Sense and Sensibility*, "Jane Austen and the Masturbating Girl," provided us with a sense of the 'interpretive room' that is preserved by the written text, that is, that the designation of a text as fiction or non-fictional, literary or clinical, does not eliminate the destabilizing effects of writing. Genette's contribution extended this position further by providing additional confirmation that the observed effect is not simply the interpretive distance that New Historicism proposes to address. In that the remaining sections, I will turn to Derrida's reading of *Robinson Crusoe* and Sedgwick's "The Weather in Proust" in order to sketch out an alternative interpretive reading strategy—a mode of 'reading otherwise'—that deploys a different articulation of 'complexity' in the terms of literary fiction. My contention is that a better account of the 'destabilizing effects of writing', that is, of its complexity, can be developed by understanding fictional writing as something that is able to pose questions about its own conditions of possibility.

Derrida's Robinson Crusoe

As demonstrated above, the narrative dynamics Genette's analysis outlines mounts a serious challenge

---

increasingly complex operations.

[97] See also Chapter 1 above; this reading strategy is guided by Sedgwick's Axiom 4, insofar as she observes the deadlock pervading discussions of the origins of sexual preference. This tactic is also heavily indebted to deconstruction, and the reading strategy Derrida develops and refines throughout his career in the on-going deconstruction of the metaphysics of writing.

to the privilege granted to cognitive interpretative reading strategies. Yet the dispute remains insufferably abstract and the theoretical stakes seemingly ambivalent: suggesting the question but without fully articulating it. Genette's contribution has shown us that, first, all written narrative necessarily participates in a process of balancing between its units' function and motivation, and that interpretive frameworks developed around a core concern for a work's *vraisemblance* are blind to the resources that writing marshals for itself to create its narrative disguise. Cognitive Literary Studies, for its part, depends upon a presumed 'empirical' basis underlying literary narrative for its coherency as an interpretive framework. This framework holds few resources for addressing the diversions, delays, and displacements writing and fiction set in motion to fabricate these qualities and disappear into themselves.

How do Derrida's analyses of the destabilizing effects of writing and Sedgwick's formulation of the need to 'read otherwise' challenge the theory of literature developed by Cognitive Literary Studies? And why is this challenge important for our present concern? Stepping back from the 'grammatical' or 'technical' focus of Derrida's work on writing, I want to turn our discussion toward his reading of *Robinson Crusoe* for an example of how the 'destabilizing effects of writing' hitherto associated with Derrida's readings of philosophy, or in certain High Modernist literary texts, are also at work in 'everyday' canonical works of literary fiction. This can be seen in Derrida's reading of *Robinson Crusoe*, in which he raises questions about the text that are not resolvable by recourse to Cognitive Literary Studies. Did Robinson die or did he survive? Cognitive Literary Studies does not assist us in addressing this question nor the implications its possibility (the possibility of asking this question) has for the use of cognitive science in literary theory more widely.

The direction that this takes me in is toward the Derridian analysis of 'writing in the general sense' as well as toward the Derridian conception of literature, specifically as elaborated in "My Chances", "This Strange Institution Called Literature", "*Le Facteur de la Vérité*", "Freud and the Scene of Writing," but most importantly in the reading of Daniel Defoe's *Robinson Crusoe* Derrida gave as part of the *Beast and the Sovereign* seminars in 2003, posthumously published in English translation in 2009 as *The Beast and the*

Sovereign, Vol. II. The inclusion of *Robinson Crusoe* is, to say the least, surprising, not only for the fact that it will be one of the two primary texts to be studied alongside Heidegger's 1929-30 seminar *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude*, but also because, as Michael Naas remarks, it is "an eighteenth century English novel that one is more apt to read in a US high school than in an advanced philosophy seminar at the École des Hautes Études in Paris" (Naas 2014, 64). It is this aspect of Defoe's text (among many more), however, that makes Derrida's interest in is of prime interest to us.

While Naas is correct that *Robinson Crusoe* is a staple of US high school English classes, the novel also holds a strong claim to having been such a fixture of adolescent education many years before US high schools first came to exist. In devoting his seminar *Robinson Crusoe*, "several other figures will make cameo appearance in the seminar, from Rousseau,[98] Marx, Joyce, and Virginia Woolf to Levinas, Lacan, and Deleuze, most of them because of what they had to say about *Robinson Crusoe*" (Naas 2014, 42). Indeed, Defoe's novel also contends with Austen's *Sense and Sensibility*, as one of the texts "that many claim to be the first English novel," Naas mentions in passing, which situates *Robinson Crusoe* nicely for our consideration.[99] For Derrida, as well as for the other writer and philosophers who have written about it, the "unparalleled 'fascination'"[100] that the text exercise on its readers, Derrida states, "will survive for a

---

[98] Derrida notes that Rousseau accords *Robinson Crusoe* an unique role in the education of the child, a role which returns strong impressions of a Humean motivation: "…we would have followed Rousseau's advice, the advice given in *Emile*. The first book, 'the first that my Emile will read,' the 'only one' that 'for a long time will compose his entire library,' will be *Robinson Crusoe*. There too, there's a sort of *tabula rasa*, the islands as desert, the phenomenological deconstruction of all prejudices and socio-cultural stratifications, and a naive, native, natural originally return to the things themselves before all the historical perversions of taste, and the social an inegalitarian dissimulations and simulacra, everything Rousseau here call 'prejudices'" (Derrida 18).

[99] Suggestive of the common ground these novels may tread, Derrida himself appears to give a nod to certain masturbatory pleasure that *Robinson Crusoe* gives him, and he hopes, the seminar attendees too. Preparing to quote from Rousseau, Derrida writes: "In the long passage I am going to read, the preceptor begins by saying 'I hate books,' which means that the exception made for *Robinson Crusoe* will consist in holding this book to be *both* the first and only book worthy of the name, *and* a non-book. As, on the other hand, among all virtues of this book, there will be that of serving for, I quote, 'both amusement and instruction,' you can always conclude that I have chose this text for this above all because it amuses me and I hope will amuse us, and I find it even more amusing, *even if some may find this in dubious taste, to read it with one hand*,…" (my emphasis, Derrida 18).

[100] Naas does not appear to draw the connection between Derrida's use of the word "fascination" here in the *Beast and the Sovereign* seminar with Derrida's earlier use of the word in *The Death Penalty* seminars during his discussion of how the 'hold' that fiction exercises is constituted through reference to the figure of Christ and the *bandelettes* that Derrida introduces somewhat cryptically: "The bandages envelop, attach, they tie but also become detached: they

long time" and works on "not only Joyce or Woolf but on every child and adult the world over" because it "is in large part because *Robinson Crusoe* is a novel that essentially creates or recreates the world itself, a novel whose central character resembles either the first man, abandoned to his solitude while being granted dominion over all the creatures of his world, or the very last man at the ends of history or at the end of the world" (Naas 42, Derrida 24, Naas 42).

Naas's summarization of why *Robinson Crusoe* is important is accurate, but it is crucial to note for the present project that the written text of *Robinson Crusoe* is largely constructed as a 'found text' written on the island, it is "Robinson's *Journal*" written by Robinson Crusoe himself: "an English book, a book of fiction that pretends to present itself as realistic and non-fictional memoirs" (Derrida 4, 56). This theme is familiar territory for Derrida's readers, but arguably it matters more for the premises of Cognitive Literary Studies. Those premises rely upon an intense compression of narrator, scriptor, author, and character, in order to maintain the fidelity of the original intention animating language that is required to disambiguate or arrest the runaway equivocation that otherwise afflicts linguistic representation. So much the better with *Robinson Crusoe*, presumably, and for those (neo-)Wittgensteinians among us who wish to object to the elision of Defoe, the cognitivists would confirm that even though the book is fiction, the writing still conforms as expected and confirmation of this fact is enthusiastically confirmed by the widespread and widely shared conviction among the readers and scholars Derrida and Naas identify. If we were considering this from the perspective that Genette articulated, we can see a striking fit between Derrida's citation of Woolf and the insight Genette offered into narrative realism; summarizing Woolf, Derrida informs the seminar:

> …*Robinson Crusoe* is a 'masterpiece' not only because Daniel Defoe was able to maintain

---

become untied from the body proper….The bandages do indeed appear; they are there all of a sudden; they leap into the light: it is a phenomenon that seems to signify, that makes a sign as in a vision. The time of this phenomenon of the bandages, their moment in the story and in the process is very remarkable (and if we had the leisure to do so, if it were the subject of the seminar, we would meditate on this time of the bandages as the lodging made ready for literature, for an ascension without ascension, an elevation without elevation, an imminent but not yet accomplished resurrection, etc.)" (33-34). This would be a direction for further research.

> and impose his own perspective on us in a consistent way, but because, in doing so, he annoys us, 'thwarts us and flouts us at every turn' (RC, xiv). And to show this, she describes the way our expectation is disappointed: we expect an experience of solitude, of isolation far from humans, on a remote island with only sunrises and sunsets. But everything we are shown is anything but states of mind and solitude. There is no sunrise or sunset, no soul or solitude, only 'a large earthenware pot' (RC, xiv). (Derrida 2009, 17)

Indeed the plausibility of Robinson's adventure, of the narrative contained in his journals which we are reading is generated by the narrative pressure of his eventual survival on the island and return to the mainland. Further still the passages Derrida cites from the next installment of Robinson Crusoe's adventures, *The Farther Adventures of Robinson Crusoe*, maintains the strong implication of *vraisemblance*, that is, that Robinson *must* have lived long enough to survive the island and publish yet another book, all of which Derrida himself points out from the outset: "This fiction that *Robinson Crusoe* remains had as a referent or real, non-fictional springboard the memoires of Alexander Selkirk, a Scottish seaman. Not a model, but a sort of basic plot or pretext" (Derrida 2009, 14). If all this is the case, what then is the issue?

The issue arrives in the Fifth Session of the seminar when Derrida advances the following interpretation of *Robinson Crusoe*, mobilizing the resources of fiction to develop the 'destabilizing effects of writing' that upset the narrative alignment of implication while nonetheless maintaining compatibility with cognitive reading strategies. Derrida announces the following provocation to the seminar casually: "Suppose now, as hypothesis or fiction, that I say the following. If I say 'Robinson Crusoe was indeed "buried alive," he was indeed "swallow'd up alive,"' you would not believe me" (Derrida 2009, 127). Derrida's declaration is strange for its outright denial of the published narrative's plot. In addition to that, the announcement of Robinson's certain death carries a force of surprise because it disrupts the necessary chain of narrative events, the documented narrative of Robinson's survival, that ensures the alignment required to produce Robinson's *Journal*. Robinson must have always already survived, so the thinking goes, in order to have written about his own survival (so-called "first-person" narrators usually cannot live to

recount their own death), and so the stakes of Robinson's narrative, of his intense fear of death *on the Island*, are called into question. It is not the fear of death that directs Robinson's narrative, but as Derrida specifies earlier, Robinson's fear of death is given precise shape from the "earthquake that almost swallows him up". He "Robinson Crusoe's fundamental fear, *the* fundamental, foundational fear, the basic fear [*peur de fond*] from which all other fears are derived and around which everything is organized, is the fear of going to the bottom [*au fond*], precisely, of being 'swallow'd up alive,' (Derrida 2009, 77). Robinson is not afraid of death but rather as Derrida shows:

> He is afraid of dying a living death [*mourir vivant*] by being swallowed or devoured into the deep belly of the earth or the sea or some living creature, some living animal. That is the great phantasm, the fundamental phantasm or the phantasm of the fundamental: he can think only of being eaten and drunk by the other, he thinks of it as a threat but with such compulsion that one wonders if the threat is not also nurtured like a promise, and therefore a desire. (Derrida 2009, 77)

This "living death" is precisely the condition of possibility of Robinson's survival, and this fabricated by the fiction that Robinson does not live. When Derrida gives voice to his seminar's objections, the law of fiction, of the institution of fiction as a legally prescribed designation, is cited as proof: "that's not in the published narrative," the fictional students respond, "*Robinson Crusoe*, the identifiable and self-identical narrative, which is moreover deposited in copyright libraries, which everyone can consult at the British Library, the Bibliothèque nationale de France and the library of Congress….In fact, as we know, he came back from his island alive and well…" (Derrida 2009, 128).

Derrida will seem to agree, that is indeed what *Robinson Crusoe* depicts, but that story is all just *fiction*. If you "read carefully," Derrida claims, you will notice that Robinson "sees himself as though in a hallucination—read carefully—he sees himself in advance (that's why he talks about it at such length), he sees the moment coming, he sees himself already buried alive or swallowed up alive"' (Derrida 2009, 128).

The seminar protests again, that this episode in Crusoe's narrative "is not a future indicative but a conditional," the textual account of his fantasy, the terror we are to understand is motivating him, "can remain a phantasm and not become a reality" (Derrida 2009, 128).

At this point in Derrida's staging of this encounter he doubles down on the rights of fiction, after all, the fictional text is legally unbound from all resembles and all implications it may be seen to have to the world, and the position taken by the fictional seminar students denies it that liberty. Such are the destabilizing effects of writing set loose, "…when I say, when I pronounce 'Robinson Crusoe,' where you do not see the quotation marks between which I am suspending this proper name or the italics in which I am inclining it, when I say '*Robinson Crusoe*,' I am naming the narrative, I am referring to the narrative (the narrative is my reference and my referent)" (129). The smooth integration of narrative structure has begun to come unfixed: "The narrative entitled *Robinson Crusoe* and, within it, the character and the narrator, the author of the journal and the characterization that the author of the autobiographical journal puts on stage are all different, other among themselves, but all are named by the same name 'Robinson Crusoe,' and as such they are all living dead, regularly buried, and swallowed up alive" (130). The very condition of possibility of this Robinson's survival is fiction, "as one cannot be both dead and alive, dying a living death can only be a fantastic virtuality, a fiction, if you like" (Derrida 2009, 130).

"What does that mean?" Derrida asks, and his response outlines the specific conditions of the textual configuration that accentuate and amplify the destabilizing effects of writing, the fact of its remaining on to be read, or not read:

> You have already understood that a book, and, still more acutely, a book the text of which is a fiction in the first person, inserting into the living narrative quotations, inserts, inscriptions from a journal speaking in the first person, etc., that such a book is both alive and dead or, if you prefer, neither dead nor alive; and everything that not only Defoe, but, in *Robinson Crusoe*, Robinson Crusoe himself, both the Robinson Crusoe who speaks and the one keeping a journal, all that they — there are already a lot of them — might have

> desired is that the book, and in it the journal, outlive them: that they outlive Defoe, and the character called Robinson Crusoe. Now this survival, thanks to which the book bearing this title has come down to us, has been read and will be read, interpreted, taught, saved, translated, reprinted, illustrated, filmed, kept alive by millions of inheritors—this survival is indeed that of the living dead. As is indeed any trace, in the sense I give this word and concept, a book is living dead, buried alive and swallowed up alive. (Derrida 2009, 130)

This is the structure of writing, and Naas reminds us of the extent of Derrida's work on this problem in his own discussion of the seminar: "[a]s Derrida argued more or less everywhere, writing is in its essence and not just accidentally, from the beginning and not just eventually, separated from its author. It is in the very structure of writing or of the trace to be readable in the absence of the author, that is, in principle if not in fact, after his or her demise" (Derrida 2009, 62).

It is along this path that my project has proceeded, to outline the how and in what way the deconstructive analytic of 'writing in the general sense' destabilizes the representational model at the core of cognitive reading practices. The generativity of fiction that survives in writing, thereby expands in writing, to reach a complexity that cannot be processed according to the computational models proposed to govern thinking.

The linkage between writing and death is a long establish topos for Derrida's deconstruction, but I believe with the reading he gives of *Robinson Crusoe*, guided by Sedgwick's analysis of Proust, we can identify a more nuanced reading of what is amplified in *Robinson Crusoe* for Derrida. This account will bring together multiple thematic and analytic concerns, the energies of which have circulated in, through, and around the critical readings developed so far, the readings of Cognitive Literary theorists such as Zunshine and Vermeule, as well as the readings of Fodor and Genette. The purpose of this account is to formulate not a conclusive or definitive take-down of Cognitive Literary Studies, but to provide a substantive response to the question that motivates this project, namely, how can affect theory be used to inform critical reading practices, and how does the use of affect, as elaborated by Silvan Tomkins, either in

combination with cognitive concepts or separately from them, open new pathways for literary interpretation and critical reading? How would these reading strategies differently address the obstacles, the effects of writing, that destabilize cognitive readings and the computational model of mental processes?

This reading takes us back to the Second Session of Derrida's seminar, when Derrida returns to *Robinson Crusoe* after first introducing the text the week before and opens the session by reading a passage from the very end of the book that invokes the thematic focus of *The Beast and the Sovereign* seminar: "…our first incursion [was] in search of…the beast and the sovereign in *The Island of Despair*, and some passages from *Robinson Crusoe* on these two themes and especially the theater of an autobiography or an Autopresentation of the sovereign himself" (31). Derrida informs the seminar that he intended to read this passage during the first session, as the preliminary formulation of "the beast and the sovereign in *Robinson Crusoe*," how this text fits with those primary focuses (28). Here is the full text of his closing remark:

> Since I am coming to the end of this introductory and scarcely even preliminary session, I'd really like, faced with so many possible readings of Robinson Crusoe (and there are certainly more still than those I've just schematically mentioned), [I'd really like] carefully to delimit, and thus also limit, like an island in an island, the territory of our seminar and the center of gravity that we shall have to constitute as much as to privilege, that is also to restrict — namely, let's say, the beast and the sovereign in Robinson Crusoe. As for the beast, it is easy and it goes without saying, even though we have said little about it until now. The book is a long discussion between Robinson and so many beasts. (27)

On the question of sovereignty, Derrida continues, "As for the auto-affirmation of sovereignty by Robinson himself, I'll content myself with reading two other passages to which we shall have to return the better to reinscribe them in the time and consequence of the narrative" (28). In service of this theme,

Derrida's introduces Robinson's reflection on his success, his mastery of the Island, his "auto-affirmation of sovereignty":

> My Island was now peopled, and I thought my self very rich in Subjects; and it was a merry Reflection which I frequently made, How like a King I look'd. First of all, the whole country was my own meer Property; so that I had an undoubted Right of Dominion. *2dly*, My People were perfectly subjected: I was absolute Lord and Law-giver; they all owed their Lives to me, and were ready to lay down their Lives, *if there had been Occasion of it*, for me. It was remarkable too, we had but three Subjects, and they were of three different Religions. My Man *Friday* was a Protestant, his Father was a *Pagan* and a *Cannibal*, and the *Spaniard* was a Papist: However, I allow'd Liberty of Conscience throughout my Dominions: But this is by the Way. (RC, 222) (Derrida 2009, 31)

It is this primitive primal scene of sovereignty that leads to Derrida to affirm the nostalgia of *Robinson Crusoe*, as an organizing principle of the narrative text: " whether it be Robinson's nostalgia for the world he has lost…after returning to that other island, England, where he will continue to dream of returning to his solitary island—and to which he will indeed return, affected by a real tropism of return to the state of nature, of a nostalgia for a quasi state of natural childhood or native naivety, close to birth;" or as an affective relation between the text and the readers: "or the affect or phantasm of nostalgia that every reader feels as much for the state of nature and euphoric childhood which in spite of everything reigns over this island of despair, which reigns over this island and bathes it, surrounding it with all sorts of seas, good and bad[101]" (32-33). Here Derrida is explicitly articulating two significant points: first, that affect (here nostalgia) participates in exerting a narrative pressure on Robinson Crusoe and within *Robinson Crusoe* that

[101] Derrida's mention of "good and bad" is a clear reference to Melanie Klein as the translator's note confirms: "[Translator's note:] "L'entourant de toutes sortes de mers, bonnes et mauvaises," where the homophony of *mer* (sea) and *mère* (mother) is being exploited by Derrida.

causes him "to dream of returning to his solitary island" and indeed pulls him back to the island long after his initial escape; second, that this narrative pressure is a point of convergence of affect and Derrida's concept of the phantasm: "the affect *or phantasm* of nostalgia" (32).

Indeed, Derrida will devote much attention to the specific affects and affective experiences that motivate Robinson's actions on the island and the ways these affects or *phantasms* also structure the written text.

> His island is an isolated world within the world, and we see him, and he constantly shows himself, solitary in this insularity, constantly in the process of deciding as to the best path, given that he has no map, neither a map of the world nor above all a map of the island. Refer for example to the moment when, having not yet found any trace of human life on the island, having not yet heard any voice other than that of his parrot Poll who echoes his own voice, Robinson discovers "the Print of a Man's naked Foot on the Shore" (RC, 142). It is as though he had been struck by lightning or thunder ("I stood like one Thunder-struck") and as though he had seen a ghost, the vision of a specter (*an Apparition*): the footprint on the sand of the shore becomes not only a spectral apparition, a "fantôme" says Borel's French translation, but a paralyzing hallucination, a sign come from heaven, a sign that is as menacing as it is promising, uncanny, as diabolical as it is divine: the other man. What terrifies Robinson is the possible trace of the spectral presence of another, another man on the island. (Derrida 2009, 46-47)

This passage sets loose an entire constellation of thematic elements that I have been gathering across the surveys of Sedgwick's writing, such as the narrative intervention of the weather ("I stood like one Thunder-struck") that marks Robinson's affective experience of terror, the introduction of a the question of relationality to the divine ("a sign come from heaven") that could be "as diabolical as it is divine," and the circular, circulatory, recirculating movement of Robinson upon the island; alongside these themes that

are so important to Sedgwick, sit figures of critical importance to Derrida, such as the phantasm ("a '*fantôme*"') and the trace. It also constitutes an encrypted scene of reading, and demonstrates the unpredictably complex affective consequences that Robinson experiences as a result *of reading* the sign of the footprint—that is, and also—the divergent narrative possibilities that are here opened up. Derrida is very specific about Robinson's reaction: "As always, he is keen to hope that all this is a good sign of divine providence, but he is afraid that, instead of God, behind the God, the devil or an evil Genius (Robinson Descartes again) might have come to do his work, like a malign substitute for God who, instead of saving him, might have come to destroy him by sending him another man to be his enemy, another foe" (47).

Exemplifying the characteristic flexibility that Tomkins theorized in their ability to be "strong" or "weak" theories, Robinson's theory of fear strengthens and slowly expands its purview, feeding back to itself ever more experiences as confirmation. The footprint that so shocked Robinson, and provoked the fear of another person on the island, another person who would thus be a challenge to Robinson's struggle for mastery, only a few pages later is revealed to be even more terrifying for the possibility that it is his own—that is, indistinguishable as either his own footprint or the footprint of another it is a sign that Robinson no longer can claim mastery over himself:

> He then wonders even more anxiously if this bare footprint is not that of his own foot. His own foot on a path he had already taken. Just as Poll the parrot returns to him only the echo of his voice, so the bare footprint is the more unheimlich, uncanny, for being quite possibly his own, on a path already trodden, that he has always described without knowing it, described in the sense that to describe a movement is also to execute it. (Derrida 2009, 47)

Here Derrida positions Robinson's parrot, Poll, who "returns to [Robinson] only the echo of his voice" next to the footprint, presenting each as incrementally more destabilizing figures of the operations of writing and each that bring with them incrementally more disturbing affective experiences. For Robinson's

sake, Poll's appearance in the narrative is remarkably *less* affectively drenched in terror than is the discovery of the footprint in the sand; that is, in *lasting* terror. Robinson's affective response to the moment of "mechanical and automatic hetero-appellation come from Poll the parrot" rises instead only to the level of "Surprize" (87): Derrida quotes Robinson's account of the incident:

> …for I was very weary, and fell asleep: But judge you, if you can, that read my Story, what a Surprize I must be in, when I was wak'd out of my Sleep by a Voice calling me by my Name several times, *Robin*, *Robin*, *Robin Crusoe*, poor *Robin Crusoe*, where are you *Robin Crusoe*? Where are you? Where have you been? (87)

In breaking into speech, Poll startles Robinson out of a deep sleep by speaking his name, and Robinson's theory of affect is weaker in the Tomkinsian sense, it does not (at this point) continue to expand, and Robinson reads this sign, recognizing his voice in the words he has taught Poll to repeat. Derrida continues to quote from Robinson: "I began to wake more perfectly, and was at first dreadfully frighted, and started up in the utmost Consternation: But no sooner were my Eyes open, but I saw my *Poll* sitting on the Top of the Hedge; and immediately knew that it was he that spoke to me; for just in such bemoaning Language I had used to talk to him, and teach him; and he had learn'd it so perfectly" (86). Robinson is clearly brought great comfort upon recognizing his own words and the memory of teaching Poll to repeat them back him, or as Derrida will say, to "speak mechanically…so as to send his words and his name back to him, repeating them blindly" (86).

Here readers of deconstruction will have easily recognized that Derrida has quite insistently been reading Poll as figure for a certain understanding of writing, one that we might reasonably say, is not 'scary', writing as the means by which a message is carried, even if—as in this case—that message only repeats itself. That Poll's repetition does not yet qualify as *writing* for Derrida (nor does it for Robinson, either) can been confirmed by his specification in the following sentence that "a book—or a piece of writing or a trace in general" is different from repetition. "One could say of every autobiography, every

autobiographical fiction," Derrida writes, "according to a trick that constructs and leaves in the world an artifact that speaks all alone [*tout seul*] and all alone calls the author by his name, renames him in his renown [*le renomme en sa renommée*] without the author himself needing to do anything else, not even be alive" (87).

The discovery of the footprint, however, is much different as it exerts a far greater affective response and generates an enormous narrative pressure on Robinson and within *Robinson*. The key difference here is that as written sign, inscribed in the sand, it has become detached from its author; unlike Poll, the footprint arrives already structured by Robinson's possible death, and his fear begins to feedback, accelerating and generating a large narrative pressure. Derrida specifies this auto-affective activity in precise detail: "[Robinson] scares himself [*il se fait peur*: literally "he makes himself fear"]. He becomes the fear that he is and that he makes himself. And all these pages, among the most extraordinary in the book, on which he is shown, in which he shows himself, meditating, in terror, on this bare footprint" (Derrida 2009, 46).

Earlier in this chapter I cited the direct injunction Derrida gives to the seminar participants not to forget that Robinson's fear is the most fundamental, "the basic fear from which all other fear are derived"; Here is the sentence once again: "You need to know or remember, as you must have noticed, that Robinson Crusoe's fundamental fear, *the* fundamental, foundational fear, the basic fear [*peur de fond*] from which all other fears are derived and around which everything is organized, is the fear of going to the bottom [*au fond*], precisely, of being "swallow'd up alive," (Derrida 2009, 77). We've just detailed how closely Derrida reading has stayed to affect in multiple senses; first, Robinson's own account of his affective responses (from "Surpriz" to "dreadfully frighted"); second, the impact of these affects have on the narrative; third, the very specific and deliberate attention Derrida's gave to quality of affect in the scenes of Robinson's encounters with technologies of writing, and thus of Robinson's affective response to a 'near death' experience in encountering the footprint in the sand. The destabilizing affective reaction Robinson has is caused not only by his encounter with writing, but in the fact that his encounters with writing, consisted in each case, with the fictional text questioning him (crucially, in Robinson's account, he is awoken by Poll's

questions "where are you, *Robin Crusoe?* Where are you? Where have you been?" but meditating and reflecting on the event, "satisfied it could be no Body but honest *Poll,*" the questions return as wholly his own; and back on Robinson's thumb Poll "as he used to do, and continu'd talking to me, *Poor* Robin Crusoe, and *how did I come here?* and *where had I been?*" (86, emphasis mine) and thereby in posing the question of its own conditions of possibility.

This is precisely the question currently at issue, and a short recap is warranted. We have already established that writing is structured always already by the necessary possibility of the absence of the author, and we have returned to the passages above to closely follow Derrida's attention to affect as a primary narrative element, "around which everything is organized," and to note Derrida's insistence on the specificity of Robinson's fear of dying a living death, "of being swallow'd up alive" (77). Again, as we saw before, Robinson's fictional survival in writing terrifies him *and simultaneously constitutes and organizes* the narrative pressure that drives his account of a safe return to England.

The interpretive 'blur' happening here is captured in two separate moments in Derrida's seminar, which cluster importantly around a series of diacritical storm 'fronts' of narrative pressure, in the opposition between thinking one's death and imagining one's corpse, to which Derrida comments that "perhaps the supposed difference between thinking and imaging finds here its ultimate root"; later the opposition between the "future indicative" of "he *will be* buried or swallowed up alive" and "a conditional" the moment "he *would be* buried or swallowed up alive," to which Derrida responds that "the difference between the conditional and the indicative, the difference between the conditional, the future, and the present or past indicative are merely temporal modalities, modalizations at the surface of conscious phenomenality or representation that count for little in view of the fantasmatic content that, for its part, happens, really did happen, to Robinson" (117, 128). As we learned from Fodor, how the cognitive psychology of the computational model differs from Hume's account of the mind is that, per Fodor, the imagination "does a job for which [Hume] really need traces (together with the computational mechanisms required to operate on them" (Fodor 2003, 133).

While Fodor is comfortable dispensing with imagination, what we seen in returning to the passages on *Robinson Crusoe* from Derrida's seminar is that we are in sore need of reading strategies that account for affect within the fictional, literary narrative. It is Sedgwick who offers the most robust set of reading strategies specifically tooled for literary interpretation, strategies that combine Tomkins's theory of affect with the object-relations psychology elaborated by Melanie Klein, Winnicott and Balint, and her reading of Proust demonstrates a profound facility for addressing the problems of affect and of writing.

To illustrate the immense contribution that Sedgwick's analytic of affect makes, I want to first return to her citation of Proust's description of the Robert fountain with which she opens the essay, to elaborate the point of connection between her reading of Proust and Derrida's reading of *Robinson Crusoe*. As she notes, "the conspicuously emblematic description" of the fountain "seems to offer a crux for articulating a number of issues—architectonically as well as thematically important ones—in *A la recherche du temps perdu*" (Sedgwick 2011, 2). Here, once more, is Sedgwick:

> Early in the fourth volume of *A la recherche du temps perdu*, Proust's narrator goes to a party at the home of the Prince de Guermantes, where he sees a fountain designed by the eighteenth-century architect Hubert Robert. His description of the fountain is one of the novels (admittedly many) descriptive set pieces:
>
> It could be seen from a distance, slender, motionless, rigid, set apart in a clearing surrounded by fine trees, several of which were as old as itself, only the lighter fall of its pale and quivering plume stirring in the breeze. The eighteenth century had a refined the elegance of its lines, but, by fixing the style of the jet, seemed to have arrested its life; at this distance one had the impression of art rather than the sensation of water. Even the moist cloud that was perpetually gathering at its summit preserved the character of the period like those that assembled in the sky around the Palace of Versailles. But from a closer view one realized that, while it respected, like the stones of an ancient palace, the

> design traced for it beforehand, it was a constantly changing stream of water….[Its] continuity, apparently complete, was assured, at every point in the ascent of the jet where it must otherwise have been broken, by the entering into line, by the lateral incorporation, of a parallel jet which mounted higher than the first and was itself, at a great altitude which was however already a strain upon its endurance, relieved by a third. From close to, exhausted drops could be seen falling back from the column of water passing their sisters on the way up, and at times, torn and scattered, caught in a eddy of the night air, disturbed by this unremitting surge, floating awhile before being drowned in the basin. They teased with their hesitations, with their journey in the opposite direction, and blurred with their soft vapour the vertical tension of the shaft that bore aloft an oblong cloud composed of countless tiny drops but seemingly painted in an unchanging golden brown which rose, unbreakable, fixed, slender and swift, to mingle with the clouds in the sky. (Sedgwick 2011, 1-2)

The intricate description of the fountain serves as a "crux" because it provides with Sedgwick a model of literary writing congruent with Proust's novel, representing "a brash manifesto for the *roman-fleuve*, the fictional form in which destinies high and low are relayed and transmuted through a series of generations and where, for example, a single character may recycle under serial names or titles, while a single name or title can be forwarded by a series of distinct characters;" and also "from the point of view of iconography," the fountain "links to earlier fountains in European literature, art, and landscape to invoke a tradition that is specifically Neoplatonic" (2). These insights are *crucial*, as they establish the primary avenue by which we can connect the reading of Proust with the reading of Robinson, in the overlapping question of characters' *survival* and *reincarnation:* "…among the Neoplatonic associations of this fountain perhaps the most pointed, in the career of the water drops anthropomorphized by Proust in terms of their exhaustion and transmutation, involves a narrative of reincarnation" (Sedgwick 2011, 2).

Objections to this on the grounds of an incompatibility between the concept of "survival" and

"reincarnation" are resolved because the concept of reincarnation straddles the life/death frontier, and the specifically Neoplatonic version or vision of reincarnation entails "the possibility of non-oppositional relations of many important kinds: between pattern and contingency; the internal and the ephemeral; the universal soul and that of the individual" (2). The 'activity' of survival that Derrida outlines in the seminar is, quite directly, the movement of dissemination, the proliferation of the text, the 'sprawl' of ipseity; survival describes a complex, circulatory movement of the "living dead"[102] as the Neoplatonic reincarnation figured "by this fountain is the possibility of the souls' enacting serial lives" (Sedgwick 2011, 3). From the initial description of the fountain, Sedgwick returns to the literary question and suggests that the fountain "might be taken as representing a novelistic vision that combines flexibility with an extraordinary economy, in an endlessly mutable but ultimately closed system, where linear narrative is propelled through a perpetual recycling of elements, lives, positions, structures, and desires that honor the conservation of matter and energy, that operates according to law" (3). Such a vision of an "ultimately closed system" is precisely what "survival" and "reincarnation" breach, and Sedgwick is quick to point this out, again and *importantly*, drawn from Proust's text: "But compelling as this vision may be, it is no sooner finely articulated than it goes wastefully, farcically off course. The full-scale weather system comes athwart the fountain's condensed and elegant version: as Mme d'Arpajon crosses the garden in search of her errant lover, suddenly "a strong gust of warm air deflected the jet of water and inundated the fair lady" (Sedgwick 2011, 3). The interruption of the "closed system" of the novel by the open system of the weather, leads directly to a necessary rethinking, a reading otherwise, that provides a different approach to conceptualizing narrative affect:

> The important question in Proust of how open systems relate to closed ones, or perhaps

[102] "Now this survival, thanks to which the book bearing this title has come down to us, has been read and will be read, interpreted, taught, saved, translated, reprinted, illustrated, filmed, kept alive by millions of inheritors—this survival is indeed that of the living dead. As is indeed any trace, in the sense I give this word and concept, a book is living dead, buried alive and swallowed up alive" (Derrida 2009, 130).

> better put, how systems themselves move between functioning as open and closed, seems like an invitation to explore some of the literary and psychological connections to the scientific insights that are nowadays grouped under the rubrics of chaos and complexity. The weather has a privileged place in discussions of complexity. (Sedgwick 2011, 3)

Sedgwick's motivation to address "literary and psychological connections" between Proust's text and scientific elaborations of complexity is not separable from the developments in cognitive science and psychology that precipitated the theorization of the computational model of cognition. Much like the weather, modelling human cognitive processes advanced significantly because of "the increasingly sophisticated understanding of feedback processes that developed [in] leapfrog fashion with the computer's vast increase of human computational power" (3). The technological obstacle to such models, Sedgwick writes, has only recently begun to shift, as it has become "possible for science to conceptualize together the absolutely rule-bound cyclical economy of these processes, on the one hand, and on the other hand the irreducibly unpredictable contingency of the actual weather" (Sedgwick 2011, 3-4). In contrast to how mental complexity was addressed in cognitive science, with the theorization of computation as a cognitive mechanism capable of executing defined operations upon mental representations, Sedgwick's purpose here is focused on how literary and psychological complexity can be read through literary affect, in terms of "[Proust's] continuing access to a psychology of surprise and refreshment, as well as a nourishing relation to work" (4).

Sedgwick's specific attention to the literary affective experiences of Proust's narrator consists in a consideration of the "habitual relationality in the novel as a whole" and the subsequent examination reveals that "while the surprise and refreshment in Proust may respond to a logic of mysticism, so too do the deeply motivating experiences of desolation and of dread" (4). The "mysticism" that Sedgwick invokes here is precisely the literary implication of the question of psychological complexity, an infinitely relational field is opened up in the "environmental order of Proust's reality orientation, which coincides with his mysticism" (4).

Proust's "mysticism," that is, his literary mysticism is to be understood as the "Proustian atmosphere in which every act and landscape brims with a proliferation of genii, demigods, and Norns" and as Sedgwick confirms, her goal here is to approach this relational field "using a different strand of psychoanalytic theory [than Freudian and Lacanian approaches], organized around object relations and most closely identified with Melanie Klein, to explore the meaning and structure of these more complex energies" (5). The "complex energies" that we're interested in are the configurations of survival and reincarnation in Proust, such as "rebirth, transmigration, metempsychosis, metamorphosis, reincarnation," that provoke affective responses of "surprise and refreshment" (Sedgwick 2011, 6).

To clarify the terms on the table, what Sedgwick proposes is a reading strategy that attends to the relation of implication between the "complex energies" of characters' affective experiences and "change[s] in the weather" (7). This relationality is dictated directly by Proust's narrator: "Atmospheric changes, provoking other changes in the inner man, awaken forgotten selves" (Proust, quoted in Sedgwick 8). As to the question of how, precisely, to describe or observe these atmospheric fluctuations, the answer is presented again by the narrative text, by Proust's narrator's self description as an "animated barometer" (Sedgwick 2011, 8).

The novel utility of the barometer is not lost, because it allows the narrator to outline a relation to "a subtle, invisible, and indivisibly systemic index of the weather" (Sedgwick 2011, 9). That is, to elaborate a relation between open systems, as "the measure of a barometric pressure…means nothing at all outside a dynamic interpretive context," Sedgwick continues: "it requires a full sense of how changes in the weight of a given column of air, will affect both the vertical movement of heat and thus the air's temperature and ability to hold moisture, and also the horizontal travel of air masses that circulate 'fronts' of pressure difference, and thus major weather events, across the earth's surface" (Sedgwick 2011, 9). For Sedgwick, the barometer represents a different mechanism of affective analysis, that figures a character's responsiveness to the narrative affective pressure system.

Given the narrow scope of this section, I restrict the remaining points to the question of how

Sedgwick's mode of reading affect offers an alternative model of the interaction of affective and cognitive processes in literary narrative. Key to this insight is that Sedgwick incorporates of Klein's understanding that "the truth or phantasy that the conditions of survival are at stake can also inject the kind of terror attached to survival into situations that traditional psychoanalysis would read in more structurally mediated, less affectively drenched terms of anxiety" (Sedgwick 2011, 12).

Just as Derrida called for greater attention to the narrative scenes of intense affect and narrative pressure produced by from Robinson's obsession with "dying a living death," Sedgwick's examines in greater detail the "primitive defenses" that the characters exhibit consonant with the "paranoid/schizoid position" when faced with one of the central tragedies of the novel: the death of the narrator's grandmother (Derrida, 2009 133; Sedgwick 2011, 24). On this point, again, just like the questions that are so important in *Robinson Crusoe*, the overriding question is how death is entangled with the process of writing. Sedgwick notes that the "fully melancholic Oedipal, monotheistic form of mourning practiced by the narrator's mother, with its permanent sacrifice of 'common sense' and 'gaiety' to the exigent new superego, is indeed what the *Recherche* offers as *the* image of mourning" (Sedgwick 2011, 31). But this is not the case for the narrator, whose form of mourning, Sedgwick asserts, "is a stubbornly successful version of an animistic, Kleinian dynamic of multiple objects, projected and internalized at many levels, encompassing new possibilities of *surprise*" (Sedgwick 2011, 31).

As the narrator begins the project of writing, the question that persists for him as an artist is not only that of the quality of his own art, but if "art itself has a life and value that are independent of him and beyond the direct will of the artist. In fact, it is impossible—not metaphysically but, for Proust, psychologically so—for either of these questions to stand alone in the *Recherche*….His fear, again, is of his own talent, unliving, being made with a dead ecology of art" (Sedgwick 2011, 33). The threat this poses is one of destabilizing the affective environment that supports and maintains him, that provides him access to experiences of surprise:

Here Sedgwick's reading returns to the terms of survival that arrived to Derrida in his reading of

Robinson Crusoe, and she advances one step further, with the crucial observation that the affective composition of his art enables Proust's narrator to modulate his fear that his art is 'unliving' and 'dead.' Sedgwick identifies this precisely: "For Proust, the ultimate guarantee of the vitality of art is its ability to surprise—that is, to manifest an agency distinct from either its creator or its consumer" (33).

Sedgwick offers us the following survey of the various articulations Proust gives of the "autonomy of the work" and affirms that "[r]eaders of object-relations psychology will recognize in these passage of Proust the subtle crisscrossing of agency, interiority, and priority that characterize certain crucial situations in accounts of early, generative object relations" (33-34). Here are the descriptions she gathers from Proust:

> "It pre-exists us" is one of the ways that he describes the autonomy of the work (6:277), and only for that reason is it able to offer "celestial nourishment" to our true self (6:264). At the same time, that already existing, maternal plenitude is also gestated internally like a child: "I felt myself enhanced by this work which I bore within me as by something fragile and precious which had been entrusted to me" (6:513-14). Reincarnation is one way to delineate this balance of agency between the internal and external, where the true self becomes, perhaps, uncannily, "this being that had been reborn [with]in me" (6:264). "The writer feeds his book, he strengthens the parts of it which are weak, he protects it, but afterwards it is the book that grows, that designates its author's tomb and defends it…for a while against oblivion" (6:508).

"The book lives its beautiful death."
—Jacques Derrida, *The Beast and the Sovereign, v. II*

Conclusion

The critique of cognitive literary studies undertaken in this project took shape as an exploration of potentially shared ground between the critical reading practices informed by affect and Derrida's analyses of the destabilizing "effects of writing." Through a combination of the conceptual resources they established, I outlined a critique of cognitive reading strategies by attempting to 'read otherwise,' to approach Derrida's thinking of writing's relation to death in terms that would be productive for literary analysis and respond to the terms of cognitive reading practices. What does this constitutive, structural relation to death mean for literary writing?

Further than that, my aim was to achieve this task without recourse to the literary texts of High Modernist writers with which Derrida's highly technical and rigorous analytic style is so closely identified. The reason for this choice was, in part, due to the fact that cognitive literary studies largely avoided discussing these texts, but the choice also responded to the ambient 'contextualization' of Deconstruction, the presumption that its implications are legitimate only for a highly circumscribed set of literary writers, "mostly twentieth-century, and mostly modernist, or at least nontraditional" as Derrick Attridge characterized it in his interview of Derrida in "The Strange Institution Called Literature" (41). In Derrida's reading of *Robinson Crusoe*, the delightfully provocative suggestion of Robinson's 'actual' death on the island provided me with an opportunity to locate the effects of writing in a decidedly non-modern literary novel, a level playing field of narrative analysis, one that would grant to cognitive reading practices a certain 'home field' advantage.

Responding to Attridge, Derrida articulates what became the organizing question for my project by opening a lateral connection to Sedgwick's work on affect and the critique of cognitive literary reading

strategies she began. Here is Derrida's response to Attridge:

> These 'twentieth-century modernist, or at least nontraditional texts' all have in common that they are inscribed in a *critical* experience of literature. They bear within themselves, or we could also say in their literary act they put to work, a question, the same one, but each time singular and put to work otherwise: 'What is literature?' or 'Where does literature come from?' 'What should we do with literature?' These texts operate a sort of turning back, they *are* themselves a sort of turning back on the literary institution. (41).

Legible here is that what draws Derrida's attention to these texts is, one might say, a 'working otherwise' of the literary text, and, moreover, articulated in language laced with implicit resonances to his critique of speech act theory and the performative effects of language. But as he makes clear, the performative dimension of language does not precisely capture what occupies him with the question of literary writing; he continues:

> Not that [these texts] are only reflexive, specular or speculative, not that they suspend reference to something else, as is so often suggested by stupid and uniformed rumor. And the force of their event depends on the fact that a thinking about their own possibility (both in general and singular) is put to work in them in a *singular work*. (41-42)

The ability of literary works to question their own conditions of possibility emerges as the distinctive work that fictional narratives bear within themselves. Their power, "the force of their event," does not derive from a privileged access to the very material of cognition, but rather as so many 'ordinary words' swirling across the page, turning back upon themselves, feeding back and forward, much like the complex patterns of air circulation that accumulate into full-blown weather systems of tremendous strength.

Bill Streever, in his book *And Soon I Heard A Roaring Wind: A Natural History of Moving Air*, presents his readers with a quotation taken from a letter by Evangelista Torricelli, the inventor of the barometer,

depicting the atmospheric conditions to which his invention responds: "We live submerged at the bottom of an ocean of the element air" (31). Steever parses the image Torricelli gives as follows:

> The ocean is a liquid and the atmosphere a gas, but both are fluids, substances without fixed shapes, their molecules casually slipping past one another, unencumbered by strong attachments. Both have considerable depth. Both are, under the right conditions, blue. Both have tides and waves. Both flow with complex currents. (31)

The unifying conceptual frame of fluids, under which liquid water and gaseous air can be grouped according to their shared physics, both being "substances without fixed shapes," captures, I think, something very close to what fascinates Derrida and Sedgwick about literary writing. Composed of the 'same' words as 'ordinary language,' they nonetheless generate complex structures that question their own conditions of possibility. Why is it that these objects are so difficult to understand? Much, that is, like the weather:

> The difficulty comes in understanding why wind seldom moves in a straight line between pockets of high pressure and low pressure, why it never succeeds in reaching equilibrium, why the highs and lows that drive it form and disappear, and why it can provide gentle propulsion one day, threaten peace of mind the next day, and destroy life and property the day after. The difficulty comes in understanding why Lewis Fry Richardson's numerical forecasting does not work as well as one might hope. The difficulty comes in understanding the confusion that arises from the earth's incessant spinning below its atmosphere and from the friction that occurs where moving air meets unyielding ground and trees and buildings and mountains. The difficulty—and the fun—comes in understanding why something that appears at first glance to be so simple can be so wondrously complex. (32-33)

Difficulty and fun: the affective pressure system of literary writing turns around the fundamental question of how something "that appears at first glance to be so simple can be so wondrously complex" (33).

Works Cited

Austin, J.L. *How to do things with words*. Oxford University press, 1975.

Austen, Jane. *Sense and Sensibility*. Harmondsworth, 1969.

Baron-Cohen, Simon. *Mindblindness: An essay on autism and theory of mind*. MIT press, 1997.

Baron-Cohen, Simon, Alan M. Leslie, and Uta Frith. "Does the autistic child have a "theory of mind"?." *Cognition* 21.1 (1985): 37-46.

Barthes, Roland. *The Rustle of Language*. Macmillan, 1987.

Bechtel, William. "Modules, brain parts, and evolutionary psychology." *Evolutionary Psychology*. Springer, Boston, MA, 2003. 211-227.

Boyd, Brian. "Fiction and theory of mind." *Philosophy and Literature* 30.2 (2006): 590-600.

Butte, George. *I know that you know that I know: Narrating subjects from Moll Flanders to Marnie*. Ohio State University Press, 2004.

Blackburn, Jared, et al. "A discussion about theory of mind: From an autistic perspective." *Proceedings of Autism Europe\'s 6th International Congress*. 2000.

Bloom, Paul. "I Who Cares About the Evolution of Stories?." *Critical Inquiry* 38.2 (2012): 388–393.

Carroll, Joseph. "III An Open Letter to Jonathan Kramnick." *Critical Inquiry* 38.2 (2012): 405–410.

Derrida, Jacques. *The Beast and the Sovereign, vol. II*. Trans. Geoffrey Bennington, University of Chicago Press, 2011.

Derrida, Jacques. *The Death Penalty, vol. II*. Trans. Elizabeth Rottenberg, University of Chicago Press, 2017.

Derrida, Jacques. *Dissemination*. University of Chicago Press, 1981.

Derrida, Jacques. *Learning to Live Finally: The Last Interview*. Trans. Pascale-Anne Brault and Michael Naas. Melville House Pub, 2011.

Derrida, Jacques. "Freud and the Scene of Writing." *The Post Card: From Socrates to Freud and Beyond*. Chicago: University of Chicago Press, 1987.

Derrida, Jacques. *Edmund Husserl's origin of geometry: An introduction*. U of Nebraska Press, 1989.

Derrida, Jacques. *Of Grammatology*. Trans. Gayatri Chakravorty Spivak. Baltimore: Johns Hopkins UP, 1997.

Derrida, Jacques. "Paper or Me, You Know…(New Speculations on a Luxury of the Poor)" *Paper Machine*. Trans. Rachel Bowlby. Stanford University Press, 2005.

Derrida, Jacques. "Signature Event Context." *Margins of Philosophy*. Trans. Alan Bass. University of Chicago Press, 1972.

Derrida, Jacques, and Alan Bass. *The Post Card: From Socrates to Freud and Beyond*. Chicago: University of Chicago Press, 1987.

Derrida, Jacques. ""This Strange Institution Called Literature": an interview with Jacques Derrida." *Acts of Literature*. Routledge, 2017. 33-75.

Fodor, Jerry. "How the mind works: what we still don't know." *Daedalus* 135.3 (2006): 86-94.

Fodor, Jerry. "Having concepts: A brief refutation of the twentieth century." *Mind & Language* 19.1 (2004): 29-47.

Fodor, Jerry. *Hume Variations*. Clarendon Press, 2003.

Fodor, Jerry. "The Big Idea: Can There Be a Science of the Mind" *The Times Literary Supplement* 1992.

Fodor, Jerry, and Massimo Piattelli-Palmarini. *What Darwin got wrong*. Profile books, 2011.

Genette, Gerard. "'Vraisemblance' and Motivation." *Narrative* 11.1 (1968): 5–21.

Hart, F. Elizabeth. "The epistemology of cognitive literary studies." *Philosophy and Literature* 25.2 (2001): 314-334.

Hogan, Patrick Colm. "Stories and morals: emotion, cognitive exempla, and the Arabic Aristotelians." *The work of fiction: cognition, culture, and complexity* (2004): 31-50.

Kramnick, Jonathan. "Against Literary Darwinism." *Critical Inquiry* 37.2 (2011): 315–347.

Kramnick, Jonathan Brody. "Empiricism, Cognitive Science, and the Novel." *The Eighteenth Century* 48.3 (2007): 263–285. .

Kramnick, Jonathan. "VII Literary Studies and Science: a Reply to My Critics." *Critical Inquiry* 38.2 (2012): 431–460.

Laqueur, Thomas W. "Masturbation, Credit and the Novel during the Long Eighteenth Century." *Qui Parle* 8.2 (1995): 1-19.

Lewontin, Richard. *Biology as ideology: The doctrine of DNA*. HarperPerennial, 1991.

Naas, Michael. *The End of the World and Other Teachable Moments: Jacques Derrida's Final Seminar*. Fordham, 2015.

Richardson, Alan, and Francis F. Steen. "Literature and the cognitive revolution: An introduction." *Poetics Today* 23.1 (2002): 1-8.

Richardson, Alan, and Ellen Spolsky. *The Work of Fiction Cognition, Culture, and Complexity*. 2004.

Richardson, Alan. "Studies in Literature and Cognition: A Field Map." *The Work of Fiction: Cognition, Culture, and Complexity* (2004): 1-30.

Tomkins, Silvan. *Affect imagery consciousness: Volume I: The positive affects*. Springer publishing company, 1962.

Savarese, Ralph James, and Lisa Zunshine. "The Critic as Neurocosmopolite; or, What Cognitive Approaches to Literature Can Learn From Disability Studies: Lisa Zunshine in Conversation with Ralph James Savarese." *Narrative* 22.1 (2014): 17–44.

Sedgwick, Eve Kosofsky. "Affect Theory and Theory of Mind." *The Weather in Proust*. Duke University Press, 2011.

Sedgwick, Eve Kosofsky. *Between men: Male homosocial desire and English literature.* New York: Columbia UP, 1985.

Sedgwick, Eve Kosofsky. *Epistemology of the Closet*. Berkley, 2008.

Sedgwick, Eve Kosofsky. "Jane Austen and the Masturbating Girl." *Critical Inquiry* 17.4 (1991): 818-837.

Sedgwick, Eve Kosofsky. "Melanie Klein and the Difference Affect Makes." *The Weather in Proust*. Duke University Press, 2011.

Sedgwick, Eve Kosofsky and Adam Frank. "Shame and the Cybernetic Fold: Reading Silvan Tomkins." *Shame and Its Sisters*. Duke University Press, 1995.

Sedgwick, Eve Kosofsky, Adam Frank, and Irving E Alexander. *Shame and Its Sisters*. Duke University Press, 1995.

Sedgwick, Eve Kosofsky, and Adam Frank. *Touching feeling: Affect, pedagogy, performativity*. Duke University Press, 2003.

Sedgwick, Eve Kosofsky, and Jonathan Goldberg. *The Weather in Proust*. Duke University Press, 2011.

Sedgwick, Eve Kosofsky. "The Weather in Proust." *The Weather in Proust*. Duke University Press, 2011.

Sedgwick, Eve Kosofsky, and Adam Frank. "Paranoid Reading, Reparative Reading" *Touching feeling: Affect, Pedagogy, Performativity*. Duke University Press, 2003.

Spolsky, Ellen. "Darwin and Derrida: Cognitive Literary Studies as a species of post-structuralism." *Poetics Today* 23.1 (2002): 43-62.

Spolsky, Ellen. "Women's Work Is Chastity: Lucretia, Cymbeline, and Cognitive Impenetrability."." *The Work of Fiction: Cognition, Culture, and Complexity* (2004): 51-83.

Starr, G Gabrielle. "V Evolved Reading and the Science(S) of Literary Study: a Response to Jonathan Kramnick." *Critical Inquiry* 38.2 (2012): 418–425.

Streever, Bill. *And Soon I Heard a Roaring Wind: A Natural History of Moving Air*. Hachette UK, 2016.

Vermeule, Blakey. "VI Wit and Poetry and Pope, or the Handicap Principle." *Critical Inquiry* 38.2 (2012): 426–430.

Vermeule, Blakey. *Why do we care about literary characters?*. JHU Press, 2010.

Wilson, Elizabeth A. *Affect and artificial intelligence*. University of Washington Press, 2011.

Woloch, Alex. *The One vs. the Many: Minor Characters and the Space of the Protagonist in the Novel*. Princeton University Press, 2009.

Zunshine, Lisa. "Fiction and Theory of Mind: an Exchange." *Philosophy and Literature* 31.1 (2007): 189–196.

Zunshine. Lisa. “Introduction to Cognitive Literary Studies.” *The Oxford Handbook of Cognitive Literary Studies*. Oxford University Press, 2016.

Zunshine, Lisa. "Mind Plus: Sociocognitive Pleasures of Jane Austen's Novels." *Studies in the Literary Imagination* 42.2 (2009): 103.

Zunshine, Lisa. "Why Jane Austen was different, and Why we may need cognitive science to see it." *Style* 41.3 (2007) p. 275-298.

Zunshine, Lisa. *Why we read fiction: Theory of Mind and the Novel*. Ohio State University Press, 2006.

Works Consulted

Astington, Janet. "What Is Theoretical about the Child's Theory of Mind?: A Vygotskian View of Its Development." *Theories of Theories of Mind*. Ed. Peter Carruthers and Peter K. Smith. Cambridge: Cambridge UP, 1996. 184-99.

Banfield, Ann. "Describing the unobserved: Events grouped around an empty centre." *The linguistics of writing: Arguments between language and literature* 1987: 265-85.

Barrett, H Clark, and Robert Kurzban. "Modularity in Cognition: Framing the Debate" *Psychological Review* 113.3 (2006): 628–647.

Bering, Jesse M. "The Existential Theory of Mind." *Review of General Psychology* 6.1 (2002): 3–24.

Berns, Gregory S et al. "Short- and Long-Term Effects of a Novel on Connectivity in the Brain." *Brain Connectivity* 3.6 (2013): 590–600.

Cohn, Dorrit and Géard Genette. "A Narratological Exchange" (2013): 1–5.

Crane, Tim. "Intentionality as the Mark of the Mental." *Royal Institute of Philosophy Supplement* 43 (2010): 229–251.

Fodor, Jerry. "How the mind works: what we still don't know." *Daedalus* 135.3 (2006): 86-94.

Fodor, Jerry. "Why pigs don't have wings." *London Review of Books* 29.20 (2007): 19-22.

Frank, Adam, and Elizabeth A Wilson. "I Like-Minded." *Critical Inquiry* 38.4 (2012): 870–877

Harris, Paul - 'Thinking by Children and Scientists- False Analogies and Neglected Similarities'." "Harris, Paul L. - 'Thinking by Children and Scientists- False Analogies and Neglected Similarities'." (2013): 1–11. Print.

Herman, David. "Genette Meets Vygotsky: Narrative Embedding and Distributed Intelligence." *Language and Literature* 15.4 (2006): 357–380.

Herman, David. "Storytelling and the Sciences of Mind: Cognitive Narratology, Discursive Psychology, and Narratives in Face-to-Face Interaction." *Narrative* 15.3 (2007): 306–334.

Hernadi, Paul. "On Cognition, Interpretation, and the Survival of Literature: A Response to Hans Adler and Sabine Gross." *Poetics today* 24.2 (2003): 185-190.

Hirschfeld, Lawrence A., and Susan A. Gelman. "Toward a topography of mind: An introduction to domain specificity." *Mapping the mind: Domain specificity in cognition and culture* (1994): 3-35.

Hirschfeld, Lawrence A. "Who needs a theory of mind." *Biological and cultural bases of human inference* 131 (2006).

Hogan, Patrick Colm. "Stories and morals: emotion, cognitive exempla, and the Arabic Aristotelians." *The work of fiction: cognition, culture, and complexity* (2004): 31-50.

Jackson, Tony E. "Issues and problems in the blending of cognitive science, evolutionary psychology, and literary study." *Poetics Today* 23.1 (2002): 161-179.

Kunin, Aaron. "Characters Lounge." *Modern Language Quarterly* 70.3 (2009): 291–317.

Leys, Ruth. "How Did Fear Become a Scientific Object and What Kind of Object Is It?." *Representations* 110.1 (2010): 66–104.

Leys, Ruth. "III Facts and Moods: Reply to My Critics." *Critical Inquiry* 38.4 (2012): 882–891.

Leys, Ruth. "The Turn to Affect: a Critique." *Critical Inquiry* 37.3 (2011): 434–472.

Loewer, Barry. "Why Is There Anything Except Physics?." *Synthese* 170.2 (2009): 217–233.

Long, Hoyt, and Richard Jean So. "Literary Pattern Recognition: Modernism Between Close Reading and Machine Learning." *Critical Inquiry* 42.2 (2016): 235–267.

Miall, David S. "Anticipation and feeling in literary response: A neuropsychological perspective." *Poetics* 23.4 (1995): 275-298.

Rebecca Wanzo. "Against Proper Affective Objects." *American Quarterly* 61.4 (2009): 967–978.

Starr, G Gabrielle. "Poetic Subjects and Grecian Urns: Close Reading and the Tools of Cognitive Science." *Modern Philology* 105.1 (2007): 48–61.

Suzie Park. "Compulsory Narration, Sentimental Interface: Going Through the Motions of Emotion." *The Eighteenth Century* 50.2-3 (2009): 165–183.

Szabó, Zoltán Gendler. "Compositionality as supervenience." *Linguistics and Philosophy* 23.5 (2000): 475-505.

Walser, Hannah. "The Behaviorist Character: Action Without Consciousness in Melville's 'Bartleby'." *Narrative* 23.3 (2015): 312–332.

Whalen, D H, Lisa Zunshine, and Michael Holquist. "Increases in Perspective Embedding Increase Reading Time Even with Typical Text Presentation: Implications for the Reading of Literature." *Frontiers in Psychology* 6.266 (2015): 255–9.

Wilson, Elizabeth A. "Another Neurological Scene." *History of the Present* 1.2 (2011): 149–22.