The Behaviorist Character: Action without Consciousness in Melville’s “Bartleby”

Hannah Walser

Narrative, Volume 23, Number 3, October 2015, pp. 312-332 (Article)

Published by The Ohio State University Press

For additional information about this article
http://muse.jhu.edu/journals/nar/summary/v023/23.3.walser.html
The Behaviorist Character: Action without Consciousness in Melville’s “Bartleby”

I. Introduction: Bartleby’s Ordinariness

Over the past century, Herman Melville’s “Bartleby the Scrivener” has generated a wild proliferation of critical interpretations, drawing upon theoretical lineages ranging from continental philosophy to American historicism and presenting the story as anything from a Marxist fable to a psychoanalytic confession.1 Behind these diverse arguments, though, rests a single shared presupposition: Bartleby is a thoroughly alien intruder into the world of Melville’s narrating lawyer, pushing the latter to a moral and epistemological crisis by evading or indeed destroying the basic heuristics by which the narrator has hitherto made sense of the world. Within the continental tradition, such heuristics have typically been presented as linguistic (as in Deleuze’s characterization of Bartleby as a “pure outsider” whose speech refuses any shared syntax) or metaphysical (Agamben’s appropriation of the scrivener as an avatar of “pure potentiality”)—not so much specific and adaptive strategies of interpretation as conditions of sense-making itself, whose contours and limitations Bartleby exposes by existing stubbornly beyond them.

Yet even recent efforts within cultural studies and cognitive narratology to define the lawyer’s interpretive paradigms as psychologically localized and historically
bounded persist in representing Bartleby as the victim of a pathology that sets him apart from “normalcy,” whether the norm is defined by contemporaneous cultural discourses or by the storyworld itself. Ralph Savarese’s historically nuanced reading of the story, for instance, links Bartleby’s asocial, mechanistic behavior to the symptoms of post-traumatic melancholia and “dyspepsia,” both objects of increasing attention among mid-nineteenth-century physicians treating railroad accident survivors. Although the trauma suffered by Bartleby turns out in Savarese’s account to be not a train crash but the constant disruptive motion of capital, the result is much the same: Bartleby becomes “the quintessential . . . ‘accident figure,’” a sick soul whose refusal of treatment constitutes an “unmasterable utopian communiqué” (“Nervous Wrecks” 49). In reframing Bartleby’s presumptive illness as a mark of distinction and resistance rather than an impairment to be fixed, Savarese aims to undermine the gospel of human perfectibility and reductive “radical democratization” that he sees as characteristic of nineteenth-century medicine, trying to salvage Bartleby’s disruptive difference rather than to cure or conceal it. Like its theoretical predecessors, however, Savarese’s understanding of “Bartleby” starts from the presumption that its title character is different—one of Melville’s asocial “isolates”—and that this alienating difference stems from a wound, whether physical or psychic.

The interpretive limitations of this assumption become more evident in the context of a similarly pathologizing reading, this time from the perspective of neurological disability. Rosemarie Garland Thomson again shifts the terrain of the scrivener’s exceptionality from the absolute to the contingent, but remains convinced that Bartleby’s disorienting impact on the lawyer results from the former’s “differences from normative expectations,” which “constitute a problem that the narrator takes as his mission to solve” (783). By assuming that Melville’s narrator, and Melville himself, had available only one (culturally dominant) set of strategies for explaining behavior, Thomson makes of Bartleby a mere prompt to assimilative activity for the normative subjects that surround him. Because the character is assumed neither to fit within existing societal concepts of agency and humanness nor to inhabit an epistemically viable alternative to such concepts, Bartleby elicits from those around him only endless, vain attempts to reinscribe him within those societal norms—not unlike, perhaps, the hermeneutical effusion that Melville’s troublingly reticent story has incited among literary scholars.

Although Thomson, following Melville himself, refuses to specify Bartleby’s ailment, others have not been so hesitant: as Amit Pinchevski notes, schizophrenia, the favored diagnosis of the 1960s and 1970s, has given way in subsequent decades to assertions of Bartleby’s “autism.” Such accounts read into the stereotyped actions of the minor character the symptoms of a real-world disorder clinically characterized not only by perseverative behavior, but also by a deficit in the “Theory of Mind” skills, which allow us to explain our own and others’ behavior as motivated by propositional mental states like beliefs, desires, and intentions. Given the central position that mind-reading has so far occupied in cognitive theories of fictional narrative, and especially of the Anglophone novel, the appeal of such a reading of “Bartleby” seems obvious: if the chief purpose of fiction, as Lisa Zunshine and others have suggested, is to “prompt us to exercise our ToM” (Why We Read Fiction 27), an apparently “mind-
blind” character like Bartleby not only represents a pathological condition—autism, or something like it—but constitutes a kind of textual pathology, a deviation from fiction's evolutionary mandate. Whether this deviation is interpreted as a deliberate artistic choice or a result of authorial ineptitude will depend on the prestige of the text in question, but, in either case, mind-blindness is seen less as a counterpoint or alternative to readerly ToM than as a simple obstacle to it, inspiring either defeat or an even more vigorous burst of mind-reading—an attempt to read through the text's formal surface to what Alan Palmer calls the “underlying mental reality” (140) beneath.

If I attempt here to explore the manifest representations of behavior in “Bartleby,” rather than diving immediately into the mentalistic depths that supposedly must undergird them, it is not out of a lack of investment in the generation of narrative explanations for behavior; on the contrary, I agree with previous readers that “Bartleby” works precisely to probe the limits of our existing heuristics for social cognition. Rather, much as Stuart Murray critiques the identification of autism with mere “mind-blindness” on the grounds that this model represents the autistic individual as a “non-reader” who “functions in terms that are supplied to him or her” from outside rather than generating his or her own epistemology (27), I resist the impulse to foist mentalistic terminology and reasoning upon a text that seems both formally and thematically to reject them. Instead, this essay raises the possibility that Melville may also be interested in installing or experimenting with a non-mentalistic set of heuristics for explaining and predicting human behavior—one that, unlike ToM, would be optimized for making sense of the “obstinate . . . machine” that is Bartleby (Pinchevski 41). The urgency of such a shift in predictive strategy, of course, would not be obvious if Bartleby, crucial as he is to Melville's story, remained an anomalous figure in it, an individual whose actions and motives, if any, are immediately recognizable as aberrant. A central fact about “Bartleby,” however—and one that the readings above almost universally elide—is that the title character is hardly the only psychologically empty “machine” in the narrative.

“Bartleby” hints at the hypothesis that humans are creatures of material causation rather than psychic motivation, bundles of conditioned reflexes whose conduct, although inaccessible to reason or persuasion, can be manipulated and managed through the adjustment of environmental or physiological circumstances. This psychological model, however, represents in Melville's narrative not a radical innovation suggested by the peculiarities of the title character but, on the contrary, the norm of the storyworld. The complementary quirks of Turkey and Nippers, whose regular fluctuations in job performance can be traced in each case to the introduction of a single stimulus (alcohol in Turkey’s case, a meal in Nippers's), evoke from the start a model of human behavior that refrains from attributing motivating mental states to actions whose repetitiveness, regularity, and simplicity render a mechanistic explanation perfectly sufficient. Far from representing Bartleby's “mind-blindness” as an abnormality that requires either assimilation or rejection, then, Melville's narrative establishes a baseline agnosticism about and indifference to the mental states of others—a position I will call, by admittedly anachronistic analogy with a later psychological paradigm, behaviorism. The explicandum of any reading of “Bartleby,” then, is not the scrivener's supposedly anomalous mechanical
behavior, but the mechanistic human universe in which the text takes place, a world in which the absence of interiority is perceived not as an alienating disorder but as a basic and manageable condition of interpersonal interaction.

Only by acknowledging automatism as an ordinary and functional feature of the lawyer’s milieu can we begin to understand the origins and significance of the massive cognitive alteration that Bartleby does produce in the story’s narrator—in other words, the specific way in which Bartleby does disrupt the interpersonal world of the office. Bartleby’s conduct, like that of Turkey and Nippers, bears all the hallmarks of a nonconscious conditioned response or behavioral program: inflexibility, immunity to verbal influence, irreconcilability with any imaginable self-interest or intentional goal that the individual might be pursuing. Whereas the external conditions of the other two clerks’ behavior are immediately visible in the narrator’s environment, however, those of Bartleby’s actions predate his employment in the lawyer’s office and are therefore impossible—not in principle, but in practice—to determine. Nor would the problem necessarily be solved if the scrivener were less reluctant to talk about his former life, or about anything else for that matter: Bartleby would probably be no more able to articulate in propositional form the causes of his odd pattern of “preferences” than Turkey or Nippers could consciously and verbally account for the physiological influences that wreak such profound transformations on their personalities.

Surmountable though this epistemic barrier may be, its interference with the smooth functioning of the lawyer’s office is significant enough to inspire his attempts to find some other means of accounting for and controlling Bartleby’s behavior. If Bartleby causes the narrator to question and revise his usual methods of explaining behavior, then, it is not due to some intrinsic incompatibility between the scrivener’s actions and the interpretive heuristics from which the narrator begins; a much more superficial friction, when it impinges directly upon the lawyer’s self-preservation and well-being (by threatening his power over his subordinates), suffices to drastically downgrade the attractiveness of the usual explanatory methods and to temporarily raise the status of any substitute strategy that has some claim to accuracy.

Yet the point, for Melville, is not simply that the lawyer’s application of the intentional stance to Bartleby fails spectacularly, demonstrating with both irony and pathos that if the scrivener is to be understood, it can only be as a mechanism. “Bartleby”’s subject is less the particular accuracy or inaccuracy of any given means of explaining and predicting human behavior than the very possibility and process of the narrator’s shifts in strategy. If, when confronted with behavior that in some way exceeds or escapes the explanatory parameters of his habitual social heuristics, the baffled interpreter need not simply persist with increasing frustration in trying to reconcile the anomalous conduct with his interpretive schema, but can instead attempt to improvise a new strategy that seems better able to account for the behavioral data, then a reader may respond to a text like “Bartleby” not by irritatedly attempting to force it into his own everyday (mentalistic) frame of explanation but by adopting an alternative, behaviorist heuristic.5 Just as the narrator of “Bartleby” may experiment with both behaviorism and mind-reading in order to better understand his employee, this same process—albeit in the opposite direction—can take place in the readers of Melville’s story or of any behaviorist text: relatively small, superficially
“realistic” deviations from the kinds of action best explained by Theory of Mind can, if sufficiently pervasive and fundamental to a story’s representational tactics, inspire a temporary privileging of behaviorist explanations and a corresponding inhibition (though not, of course, complete deactivation) of mentalistic ones. This essay attempts to delineate in detail the conditions under which a text might induce such an epistemic shift in its reader.

II. Character Minus Consciousness: Behaviorist Representations

For most human actions, Simon Baron-Cohen notes, “there just aren’t many simple, readily available, plausible, non-mentalistic explanations” (Baron-Cohen 2). Any attempt to interpret a simple action—such as entering a room, walking around briefly, then leaving it again—in terms of mechanical and unmotivated “temporal regularities” (by suggesting, for instance, that the individual enters this room and then leaves it every day at the same time) is not only, for most observers, surprisingly difficult to generate, but also “very likely to be wrong” (ibid.). Instead, we usually have recourse to a relatively automatic and effortless set of cognitive heuristics for interpreting and predicting others’ actions in terms of internal, motivating mental states known variously as Theory of Mind, folk psychology, or the intentional stance. Any literary text that aimed to induce its readers to adopt “non-mentalistic explanations,” then, would need to represent behavior that actively resists interpretation in terms of hypothesized beliefs and intentions, thus intervening in the routine process of ascribing mental states to human conduct. In articulating “behaviorism” as a literary concept, then, this section will attempt to define the plot devices and formal choices by which writers can represent actions for which mental-state explanations are untenable or nonsensical, installing heuristics for social cognition that conflict with the normative Theory of Mind model.

One prominent subtype of actions that elude mentalistic explanations, as Baron-Cohen’s example suggests, is the repetitive, compulsory, and stereotyped. Such context-insensitive behaviors represent a challenge to two essential components of Theory of Mind’s interpretive and predictive heuristics: a mind imagined as a private container for legible and propositionally specifiable beliefs and desires, and a causal connection between inward intentions and exterior actions. By depicting behavior as neither motivated by the intention to produce a particular result (since the action is repeated regardless of the effect it may have) nor translatable into propositional mental states (since the action is repeated regardless of the conditions in which it originates), repetitive behaviors dissociate action from any plausible intention or belief about the state of the world. Yet if a completely regular behavior pattern forecloses the explanatory attribution of mental states, suggesting that an individual’s actions are physically caused rather than consciously motivated, wholly irregular behavior similarly stymies Theory of Mind by implying, if not the utter absence of mental states, at least their complete disconnection—and therefore uninferability—from the real-world circumstances in which the behavior is carried out. As Daniel Dennett notes,
if "the . . . attitude we occasionally adopt toward the insane" can be characterized as "a species of design stance"—our usual strategy for "making predictions about the behavior of mechanical objects" or inanimate natural ones—it is because the beliefs on which the actions of the mentally ill seem to be founded bear no predictable relation to the information and inferences that a rational observer would draw from the situation (10, 4–5).

In isolation, of course, a physical tic or sudden eccentricity would not be sufficient to disable the attribution of mental states and incite the dissolution or evacuation of the integral model of the psyche. As Dennett notes, "the presumption of rationality"—which we might describe in Theory of Mind shorthand as the principle that if an agent desires Y and believes that doing X will bring about Y, he will (barring interference or complications) intend to carry out X—"is so strongly entrenched in our inference habits that when our predictions prove false, we at first cast about for adjustments in the information-possession conditions (he must not have heard, he must not know English . . . ) or goal weightings, before questioning the rationality of the system as a whole" (9–10). To unsettle the presumed causal connection linking beliefs, desires, and intentional actions, then, an individual's behavior must persistently and irregularly depart from the norm predicted by Theory of Mind—a constant deviation, but one that cannot be reduced to any single, easily identifiable form of interference (whether as abstract and "mentalistic" as shyness or as concrete as a physical impairment). Only when the aberrant behavioral evidence begins to outnumber or otherwise outweigh the actions that conform to one's interpretive heuristics does a strategic shift take place.7

Behaviorist fictions may diegetically justify the stereotyped or utterly illogical behavior patterns of their characters through expedients such as somnambulism, mesmerism, or psychosis, as in the ventriloquism-induced insanity of Charles Brockden Brown's Wieland, the madness of many of Edgar Allan Poe's narrators, the title character's metempsychosis in Robert Montgomery Bird's Sheppard Lee, the hypnotized or otherwise tranced women in Nathaniel Hawthorne's The Blithedale Romance and Henry James's The Bostonians, or the demonic possession in Mark Twain's The Mysterious Stranger. Beyond these concessions to narrative causality (albeit of an often supernatural and implausible kind), however, the formal choices of the behaviorist text also tend to reinforce the involuntary and mechanistic nature of its characters' actions, preventing them from being taken as indicators of a particular set of beliefs and intentions even in the case of apparently normal subjects. What distinguishes the behaviorist from the merely Gothic or supernatural fiction is the ubiquity with which interiority is evacuated and the causal relevance of the psychic denied, resisting reduction to an exorcisable monster.8

This evacuation of consciousness, and the behaviorist reorientation it encourages, need not depend upon the elimination of all mentalistic terminology from the text. A behaviorist text may well contain characters capable of reflecting on their own conduct but persistently perplexed by its sources, experiencing their own minds as impenetrable and their own actions as unwilled, as when William Dean Howell's Silas Lapham or Theodore Dreiser's Hurstwood discover themselves taking irreversible action without their own awareness, or when Henry James's protagonists wonder what
they “will have meant” by a particular word or deed. Although the most extreme behaviorist texts use aberrant behavior patterns to encourage the complete negation of interiority, constructing characters whose actions can only be explained by the replacement of intentional explanations with mechanistic ones, the psychological model subtending Theory of Mind can equally well be rendered ineffective by the depiction of characters for whom the usual attribution of mental states is still functional, but has lost all explanatory power. Indeed, one might suspect that a reader will be more inclined to adopt the behaviorist paradigm on encountering a character who shares his mentalist biases, but finds itself in a behavioral situation that is obfuscated rather than clarified by those heuristics.

The reflective character in a behaviorist narrative registers the insufficiency of mind-reading when she experiences her own actions as “hyperopaque,” a term coined by Tamar Gendler to describe behaviors that result from “the activation of an associative chain” (650). This activation “can happen regardless of the attitude that one bears to the content activating the associations” (ibid.)—that is, regardless of whatever propositional mental states one happens to hold at the moment. Even as the hyperopaque character persists in consciously verbalizing her thoughts and attempting to explain her actions in accordance with those thoughts, she is forced to recognize the disjuncture between the beliefs and intentions that supposedly motivate her behavior and the materialistic causal chain that actually incites it. (One might look for illustration to the scene in Poe’s “The Murders in the Rue Morgue” in which the narrator, having supposed himself engaged in aimless daydreaming, finds his thoughts and actions given rigorous causal determination by Dupin, who interprets them as conditioned by physical interventions like the unevenness of a paving-stone or a chance collision on the street.) Such automatic activation of “behavior-inducing mental representations,” as Gendler notes, seems based on recent psychological research to play “a larger role in behavior than many had thought” (656), indicating once again that the aberrance of the behaviorist narrative results not from its representation of the impossible or even improbable (not, in other words, from its “unnaturality” in the sense of Jan Alber and Rudiger Heinze) but from its reconfiguration of the explanatory heuristics by which we interpret the mundane.9 Priming effects, evolutionarily programmed predispositions, and other forms of non-conscious and non-propositional brain activity, elided (or reinterpreted as conscious) by Theory of Mind’s psychological model, are pulled from the margins to the center by the behaviorist fiction, forcing the reader to confront behaviors that escape the motivational system on which mind-reading rests while, in many cases, offering a predictive strategy capable of accommodating these outliers.10

The recent absorption of “neurocosmopolitanism” by critics interested in cognitive approaches to literature offers one way of rendering such alternative heuristics visible to literary analysis without either assimilating them to the norms of Theory of Mind or diagnosing them as pathologies. Disability scholar Ralph James Savarese draws attention, in a recent conversation with Lisa Zunshine, to the mundane ways in which mind-reading is apt to fail: “how well do white Americans read black Americans, or rich Americans poor Americans, or Americans generally and the rest of the world?” (25).11 By casting Theory of Mind itself as a technology limited both by its
own internal logic and by its imperfect and biased social applications, Savarese not only redescribes cognitive “normalcy” as a framework that enables some interpretive moves while disabling others, but also hints that one may, by a kind of bricolage, patch up the holes in intentional reasoning with methods borrowed from non-“neurotypical” individuals or from peripheral social heuristics like those described by Gendler. Most relevantly for literary studies, this reimagining raises the possibility that literary kinds may be defined in part by the psychological models they specify and the strategies of behavioral interpretation they elicit: if there exist a multitude of cognitive maps by which to navigate the interpersonal universe, we might well expect literary works of different eras, cultures, and genres to show distinct preferences for particular maps or combinations thereof. While behaviorist strategies are doubtless dispersed throughout literary history according to the specific social needs and challenges faced by various communities of readers, the structural compatibility and functional coherence of its devices make it a likely candidate for existence, not merely as a collection of fleeting representational moments in unrelated texts, but as a strategy adopted by entire subcategories of texts in response to both intraliterary and extraliterary influences—in other words, as a factor in the formation of generic and national literary difference.

The fact that all of the examples used above to illustrate the techniques of behaviorist characterization are drawn from American fictions of the nineteenth century may point us in the direction of one such literary kind, perhaps initiated in opposition to the relentlessly mentalistic novels that, as Sunshine and others have noted, were being produced in Britain during that century—the literary lineage that Blakey Vermeule calls “the high mind-reading tradition” (129). It is of course well beyond the scope of this essay to argue for the behaviorist commitments of the nineteenth-century American fictional canon, let alone to offer the kind of historical and sociological argument that would explain why behaviorist devices would flourish in that literary environment, and perhaps in other national literatures of the period, while remaining relatively peripheral in the novels of England and France. (Such a project, however, might well take Savarese’s rhetorical question about the weakness of mind-reading across racial and class boundaries as a starting point.) The following pages, instead of offering this history, will attempt to inventory the behaviorist representational toolkit and to delineate the model of the mind that underpins it through a single representative proof-text preoccupied not only with the possibility of behavior in the absence of consciousness but also with the malleable, contextually variable, and ad-hoc nature of our heuristics for social cognition. Nonetheless, it is my hope that the reading below will serve as a mnemonic prompt to scholars both of American literature and of narrative theory more generally, recalling other representations of character that de-emphasize intentional states in favor of material causality or extreme psychic opacity—representations that might be usefully filed under the rubric of behaviorist characterization. The conclusions reached through this analysis, then, should not be mistaken as exclusive to “Bartleby,” or even to Melville; rather, the unusual systematicity with which Melville’s short story portrays shifts in strategies of behavioral interpretation makes it a useful node where a diverse set of techniques for evacuating interiority can be observed concurrently.
III. Automatism Humanized: Opacity and Interpretation in “Bartleby”

Like “Benito Cereno” and Billy Budd—which form with “Bartleby” a kind of alliterative trilogy of narratives that explore the metamorphoses undergone by psychological models under duress—“Bartleby” focuses its representational energies less on dramatic incident than on the interpretive confusion of an observer torn between disparate methods of explaining the equivocal behavior of a strangely silent and resistant subject, methods that not only propose incompatible origins for that individual’s actions but also imply contradictory constructions of the psyche itself. Yet Billy Budd and “Benito Cereno” are fascinated by the transformed mind-reading strategies of the central character primarily insofar as that transformation mandates a corresponding moral adjustment: Billy Budd draws attention to the casualties that result from a model of action that privileges consequences over intentions; and “Benito Cereno” reveals Babo’s brain to be a “hive of subtlety” rather than the simple and subservient mechanism that Delano had supposed it to be (258).

“Bartleby,” by contrast, strikingly refuses to put an end to its narrator’s constant reconfigurations of his heuristics for social cognition. The stakes of “Bartleby” reside not so much in the ethical implications of any given psychological model as in the recognition of the contingent and undecidable quality of all such models, in the persistent recurrence of behavioral phenomena that escape the explanatory purview of one’s preferred method of interpreting and predicting others’ actions, and that render all shifts in interpretive strategy not only painful but also tragically unproductive. If two of the three stories turn on an observer’s (never unconflicted, but narratively decisive) abandonment of one method of behavioral interpretation for another—a crude, racially biased behaviorism for universal and paranoid mind-reading in “Benito Cereno”; the empathetic projection of the interior states that supposedly cause behavior for the delineation of its effects in Billy Budd—“Bartleby” earns its central position in this essay largely by representing the choice of one psychological model over another as a matter of continual hesitation and backsliding, the improvisatory amendment of a paradigm rather than the definitive rectification of an error.

“Bartleby”’s interest in what prompts the transition from one mode of explaining behavior to another is all the more noteworthy given that the narrator’s initial mentalism, though confident, is distinctly selective. Although the lawyer at first does not hesitate to attribute motivating mental states both to himself and to the likes of “the late John Jacob Astor” (4), his interactions with the other members of his office are largely determined by physiology rather than intentionality, presenting the reader from the start with an internally contradictory psychological universe: the initially mentalistic narrator is nonetheless encouraged by the extreme predictability and repetitiveness of his employees’ actions to cheat, so to speak, by taking the occasional shortcut, imagining them more as machines than as intentional agents. The little society constituted by the story’s narrator and his three employees, Turkey, Nippers, and Ginger Nut—vaguely allegorical “nicknames . . . expressive of their respective persons or characters” (5), suggestive of tools rather than people, that already un-
dermine the illusion of autonomous interiority that the proper names of the British novelistic tradition attempt to construct—is kept smoothly functional not through its members’ inference and manipulation of one another’s mental states, but rather through the calibration of their respective automatic programs. Melville represents the two clerks as bundles of repetitive and nonvolitional “eccentricities” (Turkey’s alcoholism, Nippers’s indigestion) that conveniently “relieved each other, like guards: when Nippers’s was on, Turkey’s was off; and vice versa.”

In this respect, Turkey and Nippers unite the two functional and formal aspects of minorness identified by Alex Woloch: each is at once a “worker” and an “eccentric,” a purely instrumental “gear within the narrative machine” and a fragmentary challenge to the textual status quo. Yet rather than representing behaviorism itself as a consequence of mental distress or a mere corrective to Theory of Mind’s localized failures, Melville creates a character-system (to borrow Woloch’s term) in which mentalism’s dominance—as exemplified in the narrator’s complacent musings on the diversity of human personality types, “at which good-natured gentlemen might smile, and sentimental souls might weep,” at the story’s opening (“Bartleby” 3)—is supported by a strong undercurrent of unacknowledged behaviorism. Although the “red and radiant countenance” of Turkey, the tooth-gnashing of Nippers, and the verbal tic (“with submission, sir”) with which Turkey accompanies his every line (5, 7) conform to the repetition and fragmentation that Woloch identifies as key techniques in the representation of minor characters, these reflexive behaviors appear not as deviations from an intentional norm but as the very stuff of social relations, a non-conscious and nonvolitional fundament that all interpersonal arrangements must respect.

Indeed, although “Bartleby” suggests that all psychological models—mentalist, behaviorist, and otherwise—are vulnerable to breakdown and replacement in the face of aberrant social stimuli, this pragmatism should not be mistaken for complete indifference on the question of which model is more fundamentally correct: even during the narrator’s experimental forays into mind-reading, the “hyperopaque” behavioral routines of the workplace persist as a subterranean and non-propositionally-representable influence on the very process of attributing mental states. In his first bewildered reaction to Bartleby’s constantly reiterated preference “not to,” the narrator appeals to Turkey and Nippers for interpretive guidance, with inconsistent results: whereas the former rages at Bartleby’s insubordination and repeatedly offers to “go and black his eyes,” the latter mildly excuses the new clerk’s conduct on the grounds that it “may only be a passing whim” (18). This difference of opinion, however, results not from any primary discrepancy in the propositional mental states attributed to Bartleby himself, but rather from the disparate behavioral programs on which each character is running: because the narrator makes his inquiry in the afternoon, he arouses “Turkey’s [alcohol-fueled] combative ness” while meeting with a “very gent[le]” answer from the satisfied stomach of Nippers (ibid.). Automatism, in other words, does not simply elicit a particular (behaviorist) model of cognition and action from observers, but shapes the act of interpretation itself—a reminder that reflects tellingly on the narrator’s attitude toward Bartleby’s actions in the paragraph preceding the exchange with Turkey and Nippers, where he alternately (at his most
complacent) attempts to assimilate Bartleby’s “involuntary” “eccentricities” to those of his other employees by approaching them as circumventable quirks and, in moments of irritation, feels tempted “to elicit some angry spark from him comparable to my own” (17).

It is not only the case, then, that priming effects can activate social routines that would otherwise be interpreted as motivated by the agent’s Theory of Mind-informed projection of his interlocutor’s mental states, as when, in a study by John Bargh and colleagues, subjects who unconsciously viewed words associated with rudeness were significantly less resistant to interrupt the experimenter in conversation with a confederate (despite the absence of any significant difference in self-reported impressions of the experimenter’s politeness and considerateness). “Bartleby” further and perhaps more radically suggests that the nonconscious behavioral scripts of the clerks, and of the narrator himself, influence their assessment of the motives behind others’ conduct in ways that no propositional representation of their mental states would reflect. What Bargh et al. identify as the “automaticity of social behavior” (230) extends not merely to our own interpersonal conduct, but also to our interpretation of others’ actions and intentions—even when the triggers involved stem not from the other’s unconsciously perceived behaviors but from nonconsciously installed associative chains in the interpreter himself.

Even though such research suggests that conscious mental states and articulate intentions account for a relatively small subset of social behavior, our reasoning about social behavior is of course still dominated by these categories: when asked why a person committed a particular action, we are more likely to answer in terms of motives or beliefs than in terms of priming or implicit associations. As abundant research on Theory of Mind has demonstrated, mentalism is the default mode of understanding action in the neurotypical human brain—and, arguably, in nineteenth-century fiction, whose narratives of concealed desires and misunderstood intentions presume the existence and integrity of propositional mental states. When Melville constructs a storyworld rife with behaviorism, he is likely aware that this casual elision of mental states will itself prove somewhat disorienting for his readers, whose cognitive predispositions and literary experience both pull toward mentalism. Even before “Bartleby” stages the confrontation of the lawyer’s jerry-rigged mentalism with the title character’s opacity, then, the story has already generated friction between two systems of explaining action by representing a world in which the intentional reasoning favored by the average reader is less effective than mechanistic causal prediction.

In some texts, no doubt, such friction would inspire confusion or frustration in the face of what Porter Abbott, also in the context of “Bartleby,” calls an “unreadable mind”—a represented psyche that offers no foothold for interpretation but preserves the reader in a state of bewildered and fascinated “captivity.” Abbott constructs his own understanding of unreadability against the “unnatural narratology” of, for instance, Jan Alber, who focuses on the strategies that readers use to bring aberrant textual data into line with their normative expectations. Both of these critical approaches, however, assume that intentional reasoning is the only heuristic by which we can grapple with a character’s actions—even if the character’s own opacity forces us to
resituate those intentions on the level of other, more intelligible characters (treating the opaque character as a "catalyst") or of the text's author (the "symbolic" approach) (Abbott 451–52). My own reading of "Bartleby," by contrast, argues that, even before the scrivener's drama begins, the implicit behaviorism of the storyworld has already begun to pull the reader away from his own habits of intentional reasoning, habituating him to inferences based on physical causation rather than mental states. By the time Bartleby's peculiarities become apparent, the reader is surprised, not by Bartleby's lack of psychological depth, but by the opacity of the causal associations that make him tick—much as Melville's narrator is unnerved, not because Bartleby precipitates the breakdown of a richly conceived Theory of Mind, but because the scrivener forces him for the first time to feel epistemological uncertainty about others' minds, disrupting the unreflective, corner-cutting mentalism to which he is accustomed. Far from simply stymieing the reader's interpretations, then, "Bartleby"'s construction of a behaviorist storyworld revises the reader's own mentalistic default settings, tentatively approaching an explanatory heuristic that accommodates the non-intentional influences that dominate the office.

Yet if the narrator's environment points toward behaviorism from the beginning, why does the repetitive conduct and implied psychological emptiness of Bartleby—whose actions are after all scarcely more rigidly conditioned than those of Turkey or Nippers—provoke a crisis of interpretation for the lawyer, filling him with unaccustomed empathy ("pure melancholy and sincerest pity") and inspiring the novel conviction that Bartleby possesses a "suffer[ing]" "soul" that the narrator "could not reach" ("Bartleby" 24–25)? A simple comparison of Bartleby's verbal tic with Turkey's offers a preliminary explanation: whereas the latter's incessantly repeated "With submission, sir" functions as a continual reminder of his subordination, the former's "I prefer not to" signals a sudden hitch in the clockwork regularity of the narrator's office, serving approximately the function of a blinking warning light on a car's dashboard or a printer's screen—that is, indicating the disruption of one automated subsystem by another. That the simple introduction of a foreign element into an already calibrated character-system can unsettle the prevailing strategy of explaining others' actions in that system indicates Melville's awareness that any shift in interpretive paradigm is always reactive and relative: because no set of actions intrinsically mandates either mind-reading or behaviorism, an aberrant behavior pattern will prompt the observer to reach for any novel and untried interpretive strategy, regardless of its objective appropriateness to the situation.

Even as the contagious spread of Bartleby's preferred word through the office—where the narrator finds himself and his employees "involuntarily using the word 'prefer' upon all sorts of not exactly suitable occasions" (27)—alerts the narrator both to the hyperopacity of his everyday actions and to the peculiar vulnerabilities that this automatism entails, then, it also inspires compensatory efforts both to influence Bartleby's conduct by mentalistic means and to insulate the narrator's own thoughts and actions from behaviorist conditioning. The most novel—and the most embarrassingly failed—of the narrator's attempts to evict Bartleby is his "application of the doctrine of assumptions": rather than simply ordering Bartleby to leave, the narrator decides to "assum[e] the ground that depart he must; and upon that assumption build[d] all I had
to say” (32, 30). Insofar as this plan implies both a conscious but unspoken intention on the part of the narrator and a capacity on the part of Bartleby to infer that intention from the statements and actions that it is presumed to motivate, it accords precisely with the psychological model and heuristics associated with Theory of Mind.

Yet even before the narrator finds himself “thunderstruck” by the failure of his new tactic—Bartleby remains in the office, having missed altogether the “slight hint” that the narrator thought would be sufficient (32, 33)—he doubts its efficacy: “It was truly a beautiful thought to have assumed Bartleby’s departure; but, after all, \textit{that assumption was simply my own}, and none of Bartleby’s. The great point was, not whether I had assumed that he would quit me, but whether he would prefer so to do. \textit{He was more a man of preferences than assumptions}” (31, my emphasis). Suddenly unconvinced of the interpersonal legibility of mental states, the narrator locates the plan’s potential failure in Bartleby’s insensitivity to “assumptions” and confinement to “preferences”—that is to say, his functioning by nonconscious \textit{dispositions} rather than by intentional states. The inflexible, mechanical predictability of Bartleby’s behavior forces an interlocutor to approach him on the ground not of imaginatively projected, virtually manipulable mental states (“assumptions”) but of empirically established behavioral tendencies (“preferences”) that, whether externally conditioned or determined by some feature of Bartleby himself, remain equally invulnerable to propositional persuasion. Interestingly, then, the narrator’s next impulse, after the failure of this initial attempt, is to apply a nonverbal version of his “doctrine of assumptions,” indicating through \textit{behavioral} rather than linguistic cues—“pretending not to see Bartleby at all, walk[ing] straight against him as if he were air”—that he believes Bartleby to be as good as gone (32). “Upon second thoughts,” however, the narrator rightly deems the strategy to be “rather dubious” (ibid.): what is nonfunctional in Bartleby is not some ability to perceive verbal subtleties, but rather the very process of understanding \textit{any} action, verbal or otherwise, as emanating from an intentional state in the mind of the agent.

Defeated in his attempt to aggressively project interiority into the stubbornly mechanistic Bartleby, the narrator next entertains a subtler compromise: to accept Bartleby’s psychic emptiness as a means of confirming his own psychic depth and complexity. Having “looked a little into ‘Edwards on the Will,’ and ‘Priestly on Necessity,’” the narrator begins to explain Bartleby’s aberrant actions not in terms of the latter’s \textit{own} obviously nonexistent mental states, but by appealing to a postulated divine intention, “some mysterious purpose of an allwise Providence” that, however inscrutable, at least establishes a comforting link between the clockwork world of repetitive actions in which the narrator moves and \textit{someone’s} motivating mind (35). This reinstallation of intention on the spiritual plane, in addition to offering a last-ditch explanation for the bizarre behavior patterns that the narrator finds himself unable to manage either by simple causal inference or by the projection of interiority, also enables the narrator to complacently regard himself as the privileged locus of private, propositional consciousness within his otherwise mechanistically conditioned workplace. By assimilating Bartleby’s automaticity to simple inanimacy—“you are harmless and noiseless as any of these old chairs”—the narrator is able to use Bartleby’s emptiness to confirm his own unassailable interior depth: “I never feel so private as when I know you are here” (ibid.).
Yet although this coping strategy conforms to the relation between “round” and “flat” characters (that is, those with exhaustively represented interiority and those reduced to a few behavioral quirks) posited in most accounts of novelistic characterization—which ascribe narrative utility only to those flat characters who, in Blakey Vermeule’s terms, inspire “a fit of reflection in a round or major character” (83), thus reinforcing the text’s reigning mentalistic paradigm and encouraging readers to ratchet up their Theory of Mind skills—Melville’s character-system is notable for its refusal to join the narrator in relegating Bartleby to this purely instrumental role. Not only does the inscrutable scrivener become an increasingly central focus of narrative attention, inspiring pity at his death not only in the narrator but, arguably, in the reader; more immediately, the narrator himself is forced to abandon his scheme to magnify his own interiority via Bartleby (and so, it might be surmised, propel himself to the status of protagonist) when his colleagues begin to “whisper” censoriously about “the strange creature [he] kept at [his] office” (36). The narrator’s project proves unsuccessful, then, because it depends on the fiction that Bartleby’s lack of interiority is an alien intrusion rather than an integral feature of the workplace and, thus, of the story’s character-system itself: to these external observers, unaware of the interpretive contortions that Bartleby has awakened in the narrator, the “strange creature” appears not as some indifferent piece of furniture, dissociable from the narrator’s consciousness, but as an inextricable echo of the narrator’s own cognitive stance and status.

By placing the most “minor” of all characters at the center of “Bartleby”—and by emphasizing the continuity between the scrivener’s eccentricities and the comfortable automatism of his employer—Melville disconnects behaviorist representational strategies from the peripherality with which they are typically associated in the character-systems of the nineteenth-century realist novel. More fundamentally, however, the story reframes such representational techniques as indices not of “flatness” but of emptiness—a phenomenon at once more compelling and more disconcerting. The mental states of the typical “minor” character, after all, are not negated but merely oversimplified—trite, repetitive, and exhaustively expressed (rather than incompletely and ambiguously signaled) in behavior: it is not that Austen’s Mr. Collins or Dickens’s Mrs. Jellyby are assumed to have no interiority, but rather that their interiority is presumed to hold no surprises for the reader. In the context of a novel where Theory of Mind is the dominant method of interpreting characters’ words and actions, then, the author encourages us not to waste that mind-reading energy on peripheral figures through a battery of representational tricks (verbal and physical tics, exaggeration and fragmentation of physical traits, and the like) calculated to convince us that we already know all that there is to know about their mental states—not, as in the case of Bartleby, that there is nothing to know. If Turkey and Nippers, then, could be mistaken at the beginning of “Bartleby” for “flat” characters, one of the most disturbing side effects of the advent of Bartleby is to recategorize them as empty characters. The same formal devices that within the conventional realist novel merely indicate psychic shallowness, without disrupting the attribution of mental states itself, require a qualitatively distinct interpretive paradigm—one founded not on propositional interiority but on nonconscious automaticity—when brought to the center of a narrative.

For Woloch, of course, the “asymmetric structure” of the realist novel is generated by that genre’s insistence that “any character can be a protagonist, but only one
character is” (30–31)—making the minor character a distinctly melancholy figure whose consciousness is both theoretically acknowledged and structurally stifled. The peculiar pathos of “Bartleby,” by contrast, stems from Melville’s refusal to present Bartleby either as a purely functional narrative engine whose interiority is sacrificed for the sake of a protagonist’s psychological depth or as an intentional agent possessing causally efficacious mental states. Using behaviorism neither “aggressively” nor “ironically” (197)\(^\text{16}\)—neither as dehumanizing weapon nor as satirical barb—Melville defies assumptions that a character empty of mental states must always be the target of the reader’s scorn, hatred, or apathy, depicting instead a being that positively demands our pity and care even as his non-agentic behavior forecloses the psychological structures and strategies by which we would usually attempt to offer it. If “Bartleby” places the reader not in a community of confident mind-readers but amidst a multitude of beings whose actions are opaque both to others and to themselves, then, it also insists that he learn to extend his sense of moral obligation even to those who lack the legible beliefs and intentions that he finds in members of his own social sphere—precisely because that legibility has been revealed as a partial illusion, subverted by countless hyperopaque behavioral programs over which one has no control. The ethical force of this recognition, of course, is exactly proportionate to the feelings of helplessness and displacement that it inspires. Our mourning at the text’s close is, like the narrator’s, not only for Bartleby—understood now not as a conscious and feeling entity, but as a kind of displaced cog or obsolete habit in human form—but also for the illusory knowledge of and power over mental states that Bartleby’s behavior both tempted us to pursue and, ultimately, forced us to abandon.

IV. Conclusions: A Behaviorist Criticism?

In an essay entitled “Communicate With Me,” D. J. Savarese, who is autistic, advises potential acquaintances on the most effective way to interact with him: “First, ignore my involuntary gestures . . . They are just my body’s way of responding to stimuli. If you respond to them as meaningful, they fearfully rev my heart more, but if you wait patiently and wordlessly, you free me to finally respond voluntarily. Once I’ve freed my body to respond, I can skip over the autonomic responses and give faster motor replies as the conversation continues” (9). Savarese here makes a plea for the temporary suspension of mentalistic interpretation in social interaction: to treat his involuntary behaviors as though they were intentional communications, he suggests, is as disconcerting for him and unproductive for his interlocutor as it would be to interpret a neurotypical person’s blinks, swallows, or shifts in posture as necessarily meaningful. Only when an observer’s inclination to read action in terms of intentional states is temporarily muted—only, that is, when the usual machinery of “sympathy” is turned off—can Savarese engage in social interaction on his own terms. One can perhaps imagine Bartleby, if he had the power to articulate his needs, requesting similar treatment, but this echo, as I hope the preceding pages have demonstrated, should be taken not as a “diagnosis” of the character, but as evidence of the multiple ways that
human behavior, both typical and anomalous, can escape the categories of voluntary action presumed by Theory of Mind. Most of all, perhaps, we might take the passage as an injunction to perform a similar interpretive suspension when first faced with a text that seems to bypass or undermine mentalistic readings. Such an approach would advise us, in the presence of persistently aberrant or nonsensical answers to our standard critical questions about character, to ask new questions to which the text can, so to speak, respond in its own language.

Read in this way, “Bartleby” is the story of a man who “learns” how to mind-read only when confronted with an individual whose perplexing conduct seems to escape his preexisting interpretive paradigms—a man whose halting approaches toward Theory of Mind are ultimately thwarted when that scrivener’s seeming opacity is revealed to arise not from the mental depths that the narrator has been led to infer but from his transplantation into a context where his carefully honed and rigidly determined behavior no longer serves any function. The story is about Bartleby, centers on him, could not exist without him—but it is not of him, since he does not take part in the change that drives the narrative. More significantly, the story of Bartleby is not even what Woloch would categorize as a “possible story” (40): that is, Bartleby is not a complete consciousness that the constraints of narrative form have bound and distorted. In depicting the pressures that lead to the formation of a new strategy for interpreting behavior, Melville manages also to suggest that the structures of belief and intention that we read into each other’s minds reflect not the presence of actual consciousness, but rather the confusion of an interpreter faced with an unconscious mechanism of whose design and function he is ignorant.

Of course, the very fact that we refer to Melville as an agent in the construction of his story—or, more abstractly, personify the narrative itself as an agent—indicates the degree to which Theory of Mind is interwoven with our strategies of literary interpretation and explanation. Even the “symptomatic” readings that rely on unconsciously produced textual cues, like the Freudian analyses or detective narratives to which they invite comparison, isolate these features from a presumed background of intentional communication, locating literary meaning in precisely the lacunae created by the author’s imperfect execution of a conscious purpose. To expunge this mentalism from our critical practice seems neither feasible nor desirable, given both its formidable record of success and its intuitive logic: literary texts are complex human productions that lend themselves to intentional reasoning. When I suggest an interpretive strategy that is at least partly behaviorist, then, I intend not to invalidate or foreclose mentalistic readings, but rather to add to those readings the recognition that human minds sometimes intentionally adopt heuristics that ignore intention—whether in particle physics, utilitarian philosophy, or behaviorist psychology. Any attempt to explain how and why a non-mentalistic model of human behavior may at times be useful to authors and readers must draw upon all levels of brain function, from the conscious, Theory of Mind-driven social goals that such a model may advance, to the reflexive physiological responses it may use as evidence. My goal thus far has been to render visible and interpretable those aspects of social cognition that fall outside of Theory of Mind’s explanatory schema, with the hope of enabling a
literary history in which characters like Bartleby are recognized as not anomalous but perfectly at home in a lineage of fictions that represent behaviorist worlds.

After identifying the sociocognitive profile of such fictions, the next step will be to define their historical and geographical scope—both doubtless broad, but perhaps punctuated by periodic flare-ups in particular cultural contexts such as the nineteenth-century United States mentioned in Section II above. The full description of behaviorist characterization's historical distribution, of course, would require the integration of neuroscience and cognitive psychology with quantitative, economic, and sociological data—and, even then, will represent but one step in the continuous task of delineating the models of the mind implicit in the formal devices associated with various genres, schools, and national traditions. This essay has attempted to demonstrate that, rather than simply reflecting (or even ramping up) our innate or habitual models of the mind and practices of behavioral interpretation, literary texts are capable of identifying lacunae within those theories and building, through their representations of character, alternative models that better account for those neglected aspects of behavior and cognition. Which of those gaps is addressed in a given body of fiction will depend on the specific cognitive challenges of the historical period and cultural context in which it finds itself; how that gap is filled (or simply pointed out) can only be determined through the investigation of the formal and narrative structure of the texts; but why some gaps exist and not others, why some solutions succeed and not others, and why a particular change in the social structure would place a particular strain on our capacity for understanding other minds—these questions will best be answered by appealing to the cognitive sciences' body of knowledge. The constraints by which our cognitive architecture influences the shape of fiction are visible, no doubt, on all levels of the text; perhaps, however, their true significance will become appreciable only on the relatively large scale (the scale of oeuvres, of genres, of narrative fads and historical trajectories) at which their demands are forced to compete with the comparable claims of social demographics, economic structures, cultural sensibilities—and, of course, the environmental parameters of the literary system itself.

Endnotes

1. The historicism-deconstructionism binary has been best articulated and critiqued in Nancy Ruttenburg's "'The Silhouette of a Content': Bartleby and American Literary Specificity." The dynamic interplay of exceptionality and normalcy that Ruttenburg identifies as one of the most persistent backdrops to "Bartleby" criticism, although used in her essay to argue for the story's exemplarity as an instance of what Jacques Rancière calls "the suspensive existence of literature," is equally relevant to my argument against Bartleby's supposed characterological uniqueness (Ruttenburg 139). Dan McCall concisely summarizes the Marxist and psychoanalytic approaches, along with their weaknesses, in The Silence of Bartleby, although not without falling himself into the trap of portraying Bartleby as outside the pale of any interpretive strategy and therefore as a singular anomaly: "Bartleby is intractable. If he means anything other than the riddle of himself, he is not who he is" (77). Although this paper is similarly wary of attempts to decode Bartleby or replace the character with a concept of which it is supposedly the empty symbol, McCall's declaration of
interpretive impasse misses the extent to which Melville does give his readers the tools with which to understand Bartleby—even though those tools replace the hermeneutic project of excavating interior meaning with the essentially empirical one of mapping external correlations.

2. For historical arguments for the diagnosis, see, for instance, William P. Sullivan’s “Bartleby and Infantile Autism: A Naturalistic Explanation,” and Ashley Kern Koegel’s “Evidence Suggesting the Existence of Asperger’s Syndrome in the Mid-1800s.”

3. I am condensing and simplifying, here, an ongoing debate over the relation of cognitive literary studies to evolutionary psychology with which many readers of *Narrative* will doubtless be intimately familiar. Cognitive literary theorists differ in their commitment to the notion that an interest in stories about fictional characters is innate or evolved; whether humans’ taste for fictions is adaptive or epiphenomenal, though, the assertion that fictional narratives succeed primarily by “stimulating,” “exercising,” or “activating” readers’ Theory of Mind capacities is fairly widespread. Major book-length arguments for this position include Sunshine, *Why We Read Fiction: Theory of Mind and the Novel*; Blakey Vermeule, *Why Do We Care About Literary Characters?*; and Alan Palmer, *Fictional Minds*.

4. By “behaviorism” I mean to suggest not simply the rejection of mentalistic explanations for behavior, but also the insistence on nonetheless finding a coherent theoretical model of some sort that can handle the nuances of human action. Although literary texts can draw upon a number of representational devices to discourage the reader’s inclination to attribute mental states to characters—devices that the next section will examine in more detail—they cannot (and, indeed, are unlikely to have any reason to) switch off the desire to achieve some kind of predictive mastery of others’ behavior that informs such explanations; the behaviorist text simply changes the form that those explanations are allowed to take.

5. A heuristic modeled—albeit confusedly, conflictedly, and sometimes inadvertently—by the story’s narrator. My intention is hardly to suggest that “Bartleby”’s lawyer, who shares much of the self-satisfied obtuseness of “Benito Cereno”’s Amasa Delano, represents a kind of ideal social reasoner to whose skills a reader is meant to aspire; rather, the point of the story seems to be that even the complacent lawyer can be forced by circumstances into some degree of cognitive flexibility, even when those circumstances are far more mundane than the spectacular violence of “Benito Cereno.”

6. Although the precise psychological contours and neural correlatives of our tendency to explain others’ behavior by means of mental states are of course matters of continual debate among cognitive scientists, the existence of the propensity itself is confirmed both by the regular and predictable development of “mind-reading”—in Baron-Cohen’s terms, “the capacity to imagine or represent states of mind that we or others might hold” (2)—in healthy individuals and by its impairment in autism spectrum disorders. While the terminological divergence above results more from differing emphases than from mutually exclusive models, Daniel Dennett’s definition of “intentional systems” as those “whose behavior can be—at least sometimes—explained and predicted by relying on ascriptions to the system of beliefs and desires (and hopes, fears, intentions, hunches . . . )” (3), and the intentional stance as the “strategy” of ascribing mental states to those agents in order to understand and anticipate their behavior (7), possesses a few features that make it particularly congenial to this paper’s aims. These features include the classification of intentional reasoning as one explanatory system among many, albeit one that has proven exceptionally useful in understanding human behavior; the recognition of the provisional status and flexible objects of our mental-state attributions, which can be generalized to complex nonhuman entities (for instance, a chess-playing computer [ibid.]) or withdrawn from human ones (for instance, the mentally ill [10]) when the result is an increase in predictive power; and, finally, the non-prescriptive and ethically noncommittal attitude that defines mind-reading only as a “pragmatic” decision (7)—not a guarantor of moral depth, nor even a normative pattern of cognition.
7. Of course, insofar as neurotypical humans are genetically predisposed to understand others' actions in terms of motivating mental states—and insofar as this attribution often takes place too rapidly and automatically to be accessible to consciousness—no literary text can simply “disable” its readers’ mentalistic inferences; I am arguing that certain texts both make it much harder to trust such instinctive inferences (since they are within the narrative context demonstrably inaccurate) and push their readers toward alternative explanatory paradigms. A literary work cannot switch off its reader's Theory of Mind, but it can force the output of that module (if indeed it is one) to compete with other modes of interpreting and predicting behavior that have, within the world of the text, more explanatory power.

8. This, of course, hardly implies that the devices used to cast doubt on the integrity of personal consciousness and the causal efficacy of beliefs and intentions in Gothic narratives, sensation novels, and other “genre fictions” are completely distinct from those employed by behaviorist characterization; the difference in the representational strategies at work is one of degree more than kind, but the quantitative plurality of psychically evacuated characters and actions in the behaviorist text gives rise to a qualitative shift in the interpretive strategy by which readers attempt to explain and predict characters' actions.

9. Alber and Heinze define “unnatural” narratives as texts in which the fit between the represented world and our cognitive heuristics for making sense of the real one is less perfect; their use of the concept, however, focuses on the assimilation of anomalous fictional data to “familiar interpretive patterns,” touting the ability of readers to “use parameters that are based on real-world experience and their exposure to literature to grasp textual oddities” (10), whereas I mean to emphasize the extent to which these anomalous narratives are both more closely linked to everyday cognition and less capable of being elided by readers than this analysis would suggest.

10. For a recent popularizing (but well-written and persuasive) survey of the vast quantity of neural activity—in both ourselves and others—that takes place beneath our awareness and outside of propositional consciousness, see David Eagleman, *Incognito*.

11. To suggest that even neurotypical individuals may sometimes fail at mind-reading in the same way that autistics do, or may draw upon similarly non-intentional strategies of behavioral prediction, should not be taken to imply that these structural similarities derive from the same causes. Insofar as mind-reading falters when asked to cross racial, cultural, or class lines, these breakdowns are to an important degree socially conditioned and may be altered both by the large-scale restructuring of social institutions and by behavioral intervention on the individual level. Autism is certainly susceptible to behavioral therapy, but its origins are genetic and neurological, and any ameliorative interventions must adopt methods appropriate to these causes.

12. Since prose fiction is arguably the mode in which social cognition has been most important as plot point and formal constraint, at least for the last three to four centuries, such literary specialization would offer support for Lisa Zunshine's earlier suggestion that the rise of the novel may have enabled “increasingly diverse ways of engaging [readers'] mind-reading adaptations” (“1700–1775: Theory of Mind, Social Hierarchy, and the Emergence of Narrative Subjectivity” 175).

13. The plausible character names of the British novel, as Ian Watt noted, are strategically empty: offering no clue as to a character’s behavior and demeanor, these merely denotative signs force the reader to engage with the character by speculatively projecting mental states into it—that is, by filling the hollow sign of “Tom Jones” or “Emma Woodhouse” with personality traits inferred from the figure's represented actions. This engagement reflects the character’s status as, in Watt’s terms, “a particular individual in the contemporary social environment” (19). By providing a summary in miniature of each character's typical behavior patterns, on the other hand, “Turkey,” “Nippers,” “Ginger Nut,” and their ilk do not solicit the imaginative attribution of but rather replace and render moot the character’s interiority, transforming a name from a container for causally effective beliefs and intentions to a kind of statistical record of the individual’s most frequently repeated actions.
14. See, for instance, the research of Bertram Malle, who finds that, “[w]ith 84% of behavior explanations offering at least one reason, the reason mode is the clear default for explanations of intentional behavior” (106). (Malle defines reasons as “agents’ mental states in light of which and on the grounds of which they formed an intention to act” [86].) Malle also finds that reasons take priority over other explanations in the practice of observers: an observer trying to understand a particular action “tries to construct reasons first and, if she succeeds, ends there. If reasons cannot be constructed, she searches for a causal history explanation”—that is, an explanation founded on factors outside of the agent’s conscious awareness (119). One virtue of Malle’s understanding of behavior explanations is its inclusion of a complex ecosystem of explanation types; reasons or motives represent only one such category, albeit the dominant one.

15. One might note that this explanation, too, acknowledges an absence of intention in the diegetic world only to correct it by an appeal to a higher plane—in this case, that of the author, whose motivation provides an explanatory compensation for the flat character’s deficient interiority.

16. These are the two major possibilities addressed by Vermeule in Why Do We Care About Literary Characters?

Works Cited


