

Reconsidering the Unreliability and Treatment of Mentally Ill Narrators

Ellen Stenstrom

He was frightened of the craziness he saw around him because it was an extension of something inside himself. He wanted people to be crazier and more bizarre than they really were so that he could see the line which separated him, his inclinations and random thoughts, and his half-wishes, from the full-bloomed, exploded madness of the patients.

- Joanne Greenberg, *I Never Promised You a Rose Garden* (64)

Throughout her memoir, *Girl, Interrupted* (1993), Susanna Kaysen includes the images of several pieces of official medical documentation from just before and throughout her stay at McLean mental hospital. Each note is marked with the date and an official signature, some also with a time stamp. Among the first memos, she includes an Inter-Office Memorandum from her physician to McLean Hospital, stating that his evaluation “extended over three hours” before suggesting she sign herself into McLean (Kaysen 13). Later is an admission note from McLean, stating her entry at 11:30 am (73). As she calculated the wait time at the hospital, the time driving from doctor to hospital, the hour her doctor spent calling her parents and making arrangements, and her memory of leaving the house at 8:00 am for a 9:00 appointment, the time does not add up. No more than seventy pages into her account, Kaysen writes, “That doctor says he interviewed me for three hours. I say it was twenty minutes...between my walking in the door and his deciding to send me to McLean...We can’t both be right. Does it matter which of us is right? ...But now you believe him. Don’t be so quick. I have more evidence” (71-72). She proceeds to recount the events of that morning as best as she can recall, lining them up to fit her own remembered timeline.

As recently acquainted readers thrown almost immediately into a situation of conflicting accounts from differing types of authorities, what are we supposed to believe and in whom do we place our trust? The embedded documents are, after all, officially time stamped; however, the memoirist who is positioned as the reader’s primary source of intimacy and knowledge explicitly renounces the very documents she chose to include. In any narrative with conflicting accounts, readers are trained to examine the trustworthiness entrenched in each character’s story, and to make judgments about whom to believe. The situation in *Girl, Interrupted*, although autobiographical, requires readers to do the same. Do we trust the medical institutions involved

(despite contemporary criticism of historic mental health treatment practices), or do we trust the author who has just signed herself into a mental hospital but is also the most immediate source for her own experience? Though a vast body of research exists in the studies of madness, in feminist studies and gender dynamics, and in studies of narratological un/reliability separately, there is very limited scholarship in narrative theory addressing how we as readers interpret the reliability (or lack thereof) of narrators (especially female narrators) with mental illnesses of some sort, and how that choice impacts a reading of a text. This examination will work toward filling this sizeable gap in scholarship through its reconsideration of the un/reliable binary itself, primarily in regard to its use with mentally ill narrators and its gesturing toward larger uses as a whole.

It is also worth asking why the question of reliability itself is so pervasive in critical discussions of literary narratives. Does it really matter whether someone reads a given narrator, like Kaysen in the example above, with a different amount of reliability than is suggested in her provided official documentation? If one reads with the interpretation that Kaysen's hospital admittance forms are evidence of her mental instability and that she is therefore unreliable, one could learn that Borderline Personality Disorder (Kaysen's later revealed diagnosis) can cause altered perceptions of reality and leave a person with fragmented memory even years later. If, however, one reads Kaysen as a reliable narrator whose personal experience with mental illness lends the most direct insight into a segment of the history and evolution of psychotherapeutic treatment, one would conclude that Kaysen's health professionals made a rash and stereotyped judgment on the best plan of treatment, revealing some of the many flaws in treatments of mental illness. The difference between these two readings is so dramatic that it does not seem a discussion of the text can exist without consideration of Kaysen's un/reliability.

Out of the body of narratives with mentally ill narrators, there is a section of semi/autobiographical accounts for which un/reliability and its implications are particularly poignant. These texts include Charlotte Perkins Gilman's "The Yellow Wallpaper", Sylvia Plath's *The Bell Jar*, Susanna Kaysen's *Girl, Interrupted*, Joanne Greenberg's *I Never Promised You a Rose Garden*, and Lauren Slater's *Lying: A Metaphorical Memoir*. "The Yellow Wallpaper" was an early precursor to this subgenre when it was first published in 1892, and the rest of the texts followed throughout the period between 1960 and 2000. One also cannot ignore that each of these texts additionally features female authors who experienced mental illness of some sort at some point in their own lives. Within the long historical context of "madness" as

linked or associated with womanhood, the role of gender in discussions of reliability is one of multiple dimensions to a text that have presented barriers to a reader's full trust and become intertwined with questions of un/reliability; these deep-seeded associations must also be addressed as they never have before. This examination of the "unreliable narrator" since Booth's original use of the term tracks its evolving usage through Nünning, Fludernik, Carracciolo, Richardson, McHale, and others. While each theorist has slightly altered the usage of un/reliability, they have by and large maintained a clinical distance that still does not fairly allow for the consideration of things like gender and ability (as well as race, class, sexuality, etc.) as separate from an assessment of the narrator's technical reliability.

Through the texts with the most problematic constructions of un/reliability, those written by women between 1890 and 1990, it is clear that the "unreliable" label, which is so often assigned to narrators with questionable mental clarity or even sanity, is inextricably linked not only to mental health, but also especially to gender. In order to construct new and more appropriate terminology and methodology through which to discuss narrative "reliability", it is necessary to also examine the multiple facets of identity that have been traditionally lumped together and collectively used to judge reliability under the front of theoretical technicalities. For the particular set of texts addressed here, it is gender that becomes most significantly intertwined with mental stability, and then reliability.

By examining this narrower case study of women's mental illness narratives since roughly 1900, noting the insufficiencies and injustices of current terminology around mental health, I will analyze the dangerous implications of the un/reliable binary for female mental illness narratives and challenge its continued usage. I also propose the discontinuation of the un/reliable binary and gesture toward an alternative method for studying what can be conveyed between narrator and reader in fiction - one that separates questioning the narrator's mental health state from questions of their trustworthiness.

Though the technique itself was used in storytelling long before 1961, narrative theorist Wayne Booth was positioned at that time to assert his theory of the unreliable narrator. In his groundbreaking study of fiction and narrative theory, Booth introduces the concept of an "unreliable narrator" as he explains the complex relationships not only between reader and author, but also between reader and narrator and author and narrator. In *The Rhetoric of Fiction*, he writes, "For lack of better terms, I have called a narrator reliable when he speaks for or acts in

accordance with the norms of the work (which is to say, the implied author's norms), unreliable when he does not" (Booth 158-159). From this original definition, numerous derivative definitions have emerged (from Riggan, Phelan, Nunning, Yacobi, and others as explored later), but all build on or critique this basic premise. Most criticism is founded on Booth's use of relative terms such as "in accordance with" and "norms". These terms leave significant interpretation up to the reader's perceptions about the constituted "norms" of a given author and which narrative moves are "in accordance" with it. Booth's definition also implies that what is normal to the author will also be normal to every reader. If a given author experiences a reality where there are no ghosts, for example, a narrator who sees ghosts in her home can then be considered unreliable.

The Rhetoric of Fiction explores several other facets to an author's reliable narration of which readers should be aware. "He must passively relay to us every detail, however insignificant, that is a genuine part of the experience. He must avoid even the normal abridgments of dialogue...In short, selectivity must be eliminated – or is it simply that all recognizable signs of selection must be eliminated?" (52). It is now understood, though, that such "abridgments of dialogue" and "selectivity" are practically unavoidable aspects of any narrative and are thus unhelpful requirements for reliability. Additionally, Booth suggests that un/reliability is a consistent narrative trait throughout the account – or that the narrator's reliability status does not change. One of Booth's more timeless and helpful instructions reminds readers that, "We accept, by convention, the claim that what is reported as going on in [the narrator's] mind really goes on there, or in other words, that [the author] knows how [the narrator's] mind works" (163). In a story that has not produced evidence of an unreliable narrator, this is the premise on which readers must rely, giving the benefit of the doubt to a narrator's judgment that is truthful and as accurate as can be expected until the reader is given evidence to the contrary.

Around the time Booth was writing about fiction rhetoric, the documentary novel's popularity grew as well. Traditional realism was no longer sufficient, and authors began borrowing from historical events and discourse in hopes of gaining more authority (Søndergaard 57), which is particularly salient to note for the focal texts examined later, which all blur lines of genre. Since early modernism, the relations between author and narrator have required the careful differentiation of author and narrator, a distinction on which the very concept of narrative unreliability relies and a division that postmodern authors more recently have obscured

(Richardson 4-5). Especially in a documentary novel where many of the people and events resemble those of real historic relevance, it is even more imperative to remember that a fictional narrator who resembles the author may or may not accurately portray the author, as the narrator is a mere fictional construct.

A complete understanding of Booth's terminology also involves a thorough understanding of the context with which he conversed. About a decade before *The Rhetoric of Fiction*, the conclusion of World War Two brought many changes to the Western world, including the re-introduction of neurasthenia and its increasing sphere of influence. In an especially influential book credited with igniting the second wave of feminism, Betty Friedan explored the far-reaching causes and effects of what remained nameless by doctors, including numerous personal accounts of behavior and emotion that sound like what would later be called neurasthenia and today be considered depression. The reports reoccurred among a great many married women and included feelings of aloneness, denial, dissatisfaction, and shame, all of which they were afraid to admit (Friedan 15-16). The 1960s brought the eventual exposure of an undeniable problem among middle-class, stay-at-home women, but it was almost always dismissed as a superficial condition. Many men began suggesting that, perhaps women had become too educated for their own good, that when many would receive a college education and then marry, their new domestic role as a housewife made them unhappy. They proposed that less education would solve the problem, replacing a college education with classes in home keeping and bettering one's marriage and sex life (18). Under this ideology, women were usually not treated for their ailments, but merely counseled in adjusting to their role with no other options (21). These conversational narratives are representative, too, of the written narratives produced by women and repressed by men as both subpar writing and silly complaints. This is the context, then, in which Booth defined reliable and unreliable narration – one that would implicitly seep into the ideologies embedded within Booth's terminology.

One valuable lens of study can happen through looking to the media as a source of stereotype propagation. During the Second World War, mostly women writers and editors staffed magazines. When men returned from the war though, the magazine staffs were replaced with men and the "spirited career girl" image replaced with a new image of women as house mothers (47). With men writing even women's magazines, women were continuously bombarded with the expectation that they can only truly be fulfilled by having children and maintaining the household for them. "In the feminine mystique, there is no other way for a woman to dream of

creation or the future,” Friedan writes. “There is no way she can even dream about herself, except as her children’s mother, her husband’s wife” (55). According to the male magazine staff, women were not interested in politics, global affairs, art, science, adventure, education, or even their own communities, though there is a particular irony to these men claiming themselves reliable sources for information on women’s interests. “ ‘Women can’t take an idea, an issue, pure,’ men who edited the mass women’s magazines agreed. ‘It has to be translated in terms they can understand as women,’ ” (Friedan 44) Friedan wrote. Yet if this had not been true before, it became a self-fulfilling prophecy made true by women who were not interested in such things simply because they had no access to information about them. Magazines were not the only problem, either. Advertisements, television, movies, novels, newspaper columns, books by experts, psychologists, and sociologists all followed suit (28-29). Friedan recounts the story of a New York woman who learned she was not able to breastfeed her newborn child, and who subsequently succumbed to a nervous breakdown (12). Incidents such as this make painfully clear the damaging effects of persistent media and societal portrayals of an idealized narrative when such a narrative is not attained. Perhaps under this pressure, when women were determined to marry and start a family and reach fulfillment, those who did not experience ultimate happiness were so let down by the expectation of what they were supposed to feel, that mental illness naturally followed.

As is already evident, the definitions and terminology surrounding mental illnesses evolve greatly with changes in cultural attitudes and values. In the case of hysteria/nervousness/neurasthenia/depression, the definition has always been constructed by male dictated ideals. Especially during the mid-twentieth century, the expectations often imposed on women made up a woman’s expected reality. Thus, when the ideals were not followed, a woman could be considered neurotic and out of touch with reality. Contemporary psychology professor and writer Andrew Solomon reflects on historic gender associations that, “Nonetheless, women are often seen as depressed when they fail to show the vitality that their husbands expect or demand, and which the women have learned to expect or demand of themselves” (Solomon 176). In this way, women whose actions did not match the model housewife were then declared ill. There may be yet another self-fulfilling prophecy here though, for as Solomon describes, “Women attempting to conform to ideals of femininity may act depressed out of conformism; or they may become depressed as a consequence of being unable to live within a stultifying definition of femininity” (176). Solomon helpfully articulates two possible implications for the

gender/madness tie, both of which offer additional insight for the variety of female authors writing female narrators with mental illnesses. The precarious position of a woman's mental health was (and is) so often interpreted through the lens of her gender that a female narrator's mental health would naturally be judged accordingly.

The entire history and evolution of the various ways mental illness has been socially constructed and physically manifested is extensive, and while the above excerpts come nowhere close to conveying a thorough overview, for the purposes of this literary examination they present a series of markers whose indelible stains are still visible in modern narratology and the "unreliable narrator." This history, then, cannot be separated from the conversation on mentally ill narrators; such discourse conveys a history where the mentally ill are consistently "othered" and physically separated from society, whether in ships, prisons, mad houses, and or hospitals before the deinstitutionalization movement of the fifties and sixties. Even assuming a narrator to be unreliable because of mental illness stems from the assumption that someone diagnosed with a mental illness cannot be trusted and is an "othered" type of narrator. It is not difficult to imagine the implicit influence of this context, then, on Booth's work. Since Booth birthed the "unreliable narrator" in 1961, definitions and diagnostic criteria of mental illnesses have changed, becoming increasingly dependent on other factors besides just physiology, and will continue to evolve with society. Particularly given contemporary feminist theory work, it is impossible to continue ignoring role of gender in analyzing and interpreting sanity and reliability. Therefore, the traditional definitions and diagnoses of mental illnesses cannot continue to be used as grounds for declaring a narrator unworthy of trust in telling a story.

With this framework, we must also examine the theorists whose progression of work with unreliable narrators built on Booth and one another, constructing the understanding we have of unreliability today. Narrative theorists after Booth primarily recapitulate variations of the same initial formula, and so imitative applications of his "unreliable narrator" definition still permeate literature discussions and dictionaries of literary terms. This lasting legacy has led to inaccurate usage of the term and perpetuation of the binary it established despite newer understandings of psychology that would suggest otherwise (Olson 96). In 1981, William Riggan next attempted a four-division organization of first-person unreliable narrators in *Picaros, Madmen, Naifs, and Clowns*, reducing the complexities of an "unreliable" narrator into one of four caricature-like categories (listed in the book's title). He claims that narrators' lack of reliability can be identified

on grounds of the “unacceptability of that moral philosophy in terms of normal moral standards or basic common sense and human decency” and that they assume the normative validity and universality of their “moral “philosophy”. Once again, though, this definition’s use of terms like “human decency” and “normal moral standards” remain culturally dependent and individually varying (Riggan 36). James Phelan later made more direct changes to Booth, posing a distinction between restricted narrators and unreliable narrators, for the latter of which he created a taxonomy of six types. Phelan suggests that a narrator must perform three primary tasks: report (facts, characters, events), interpret (knowledge or perception), and evaluate (ethics) (Phelan and Rabinowitz 33-34). A narrator performing only one of three tasks as used by an author wishing to communicate something that the narrator doesn’t know about, qualifies as restricted narration. When a narrator performs any of the three tasks inadequately, it is considered unreliable narration. Phelan does not alter, however, that the narrator’s perceived “inadequacy” is still measured against “how the implied author would perform it”, as Booth originally wrote (35).

Around the turn of the twenty-first century, Booth’s original definition was taken in more diverting directions than it had before. Ansgar Nunning offered comparatively radical revisions to the unreliability formula. In his own definition, Nunning creates three new categories of unreliability (factual contradiction, lack of objectivity, and incompatibility of worldview/ideological unreliability) and suggests the following criteria:

...narrators are interpreted as unreliable if their behavior deviates from the reader’s expected standards of normalcy, objectivity or factual accuracy...the figure of the implied author is projected onto the text as a consequence of the reader’s discomfort with the narrator’s performance... It is because the reader’s construction of the events and the reader’s interpretation of them are contradicted by the narrator’s presentation and evaluation of the story and because the reader’s opinions are in conflict with the narrator’s views that the reader constructs the narrator as a scapegoat responsible for that incompatibility. The implied author is then pictured as intentionally dissenting from the narrator figure, who gives herself away by the unreliability of her own discourse” (quoted in Fludernik 75).

Where Booth’s original definition empowers the author to place unreliability between the implied author and the implied reader as an act of collusion behind the back of the narrator, Nunning instead suggests that readers assign unreliability to a narrator when her account does not align with what the reader believes to be the truth (75). With this newly adjusted agency,

Nunning also argues that narrator un/reliability is not an invariable characteristic, but merely a construction of the reader's textual encounter (97). Because each reader brings varying sets of beliefs to a textual interpretation, each will have differing concepts of the normative reality that had previously been universally attributed to every writer. Rather than misalignment between narrator's and implied author's truths, unreliability comes from inconsistency in the reader's and narrator's views. Despite this seemingly progressive approach to the narrator, Nunning proceeds to present a list of textual signals for unreliability, implying that there are stable commonalities with which readers can uniformly identify narrators. Is this not the approach against which he was just protesting?

In synthesis of this evolving terminology, Monika Fludernik uses Booth, Nunning, Yacobi, and others' works to specifically interrogate the unnamed narrator in Charlotte Perkins Gilman's short story, "The Yellow Wallpaper," (*TYW*) in her essay, "Defining (in)sanity: The Narrator of The Yellow Wallpaper and the Question of Unreliability." Using this combined narratology, Fludernik works to prove that the narrator is indeed unreliable (and was so from the beginning of the story) and therefore, that the traditionally feminist reading of the text is incorrect. If the narrator is mad from the start, she posits, then we cannot argue that the rest cure or the room itself or even the wallpaper caused her madness; thus, that it cannot represent the oppressive male patriarchy. In establishing her argument, Fludernik agrees with Nunning's claim that differing interpretations of a narrator's reliability often exist for a single text, except in overt instances of disparate telling of events or in detective-style scenarios where readers experience a moment of revelation upon discovering the secret truth and are suddenly "in the know," finally able to "congratulate themselves on having unearthed the figure in the carpet, the secret which the implied author was trying to impart" (Fludernik 78). The reader, now in collusion with the implied author, can easily detect that the narrator is not telling the story true to how it really happened. This moment of "discovery" does not happen with the type of narratives under question here though.

The work with which Fludernik engages throughout her analysis of Gilman models the type of critical skepticism with which I propose we need to approach narrators of questionable reliability. She presents evidence in support of both *TYW*'s narrator's sanity and insanity but does so by separating the question of sanity from the question of reliability. Her investigation is, however, limited to the scope of a single short story. While significant discourse and study of unreliable narration exists and the involved theorists pose variations on Booth's original

definition, none but Fludernik has specifically examined the ramifications and methods for questioning a mentally ill narrator's reliability. None have considered the contemporary understanding of mental illness and how that may impact the way we think about narrators with questionable mental clarity. In a literary community that increasingly reads with awareness for nuances of race, gender, and sexuality (to name a few facets of identity), a significant gap in understanding remains in regard to varying states of mental health. This void is not only a hindrance to meaningful engagements with texts, but also an injustice to anyone who has experienced mental illness. With Fludernik's work as a model and foundation for my own, I aim to continue her work of deciphering reliability at its intersection with mental health, now on a larger scale that encompasses the way mental illness narratives as a subgenre are read for and with an understanding of the narrator's "un/reliability".

As the way we understand mental illness in society as a whole has changed, it is time that we also update the way we talk about mental illness in literature, particularly as it relates to reliability. The terms "reliable" and "unreliable" are so frequently handled in literary discussions of varying degrees of formality, but with little awareness for the technical requirements so many theorists have established. They are used interchangeably with other terms of inequivalent meaning. Even when used in the proper context, the reliable/unreliable binary not only neglects to provide just, nuanced, or effective information about the narrator, but the binary also perpetuates harmful interpretations of narrators and even stigmas of mental illness outside of literature. The flaws with this binary's usage can be broken down into five key problems.

The tendency to automatically interpret mentally ill narrators as an "unreliable"

"What is the reality status of the story he tells," asks Marco Caracciolo. "[T]he fact that we cannot conclusively answer this question prompts the reader to pay attention to – and place an interpretive premium on – the narrator's mental illness, resulting in the foregrounding of the character-centered illusion and in categorizing readings" (Caracciolo 82). As readers who are inherently unable to entirely understand the "reality status" of a given fictional world, we are taught to seek other clues with which to pass judgments on characters and their actions. This practice, however, can be harmful and, in the case of semiautobiographical and memoir texts, dehumanizing.

At this point, the history of mental illness and its treatments becomes particularly relevant to understanding at least part of the dominating association between illness and unreliability. Conceptions of mental illness have historically separated the “mad” from the sane and hiding them away from the rest of society. At various points the insane were kept at sea on large ships, criminalized and sentenced to prison, committed to state asylums or mad houses; until America’s movement for deinstitutionalization, they were treated indefinitely in mental hospitals. The perpetuated separation and “othering” of the mentally ill may be moving toward more humane approaches on the institutional level, but it could be argued that our fascination with declaring mentally ill narrators to be unreliable is a modern textual manifestation of a historical tradition built on separation. Tamar Yacobi articulates this well in her own discussion of how readers resolve “fictive oddities and inconsistencies” by “...bring[ing] divergent as well as otherwise unrelated elements into pattern by attributing them, in whole or in part, to the peculiarities and circumstances of the observer through whom the world is taken to be refracted” (quoted in Caracciolo 93). Readers work to make sense of an inconsistent narrator, she says, by lumping together any number of contributing factors (age, gender, race, belief system, etc.) and attributing all of these potential causes of unreliability to the element that appears to be distorting the narrator’s account the most – in this case, mental illness.

Susanna Kaysen recalls feeling the impact of this stigma as she recollects watching TV with the other women during her stay at McLean and seeing the violence and riots in the world. She comments that the patients were calm during this time as they watched people no different from themselves, aside from them having “the one thing we would always lack: credibility” (Kaysen 92-93). Astutely, Kaysen’s reflection acknowledges that the discontent at the root of the events unfolding on TV was the same sort of discontent within many of the patients, yet the world broadcasts and listens to those who do not live under the roof of “madness” and will therefore never be listened to. Sylvia Plath writes of similar conversations in *The Bell Jar* through a conversation between Esther and her doctor. In Esther’s memory, her doctor suggested shock therapy after only two weeks and two appointments. The dialogue between the two exposes a scared young woman desperately attempting to convey her condition to a doctor who speaks down to her and responds as though nothing she says can be trusted (Plath 135). In both literary examples, characters demonstrate behaviors that represent those of readers’ engagement with a text, but also provide subconscious excuses for readers to do the same.

As theorists have created ways to dismantle sexist and racist institutions and discourses,

so we must also disassemble the ableist and dehumanizing discourse that associates mental illness and unreliability. Supported by a legacy of overgeneralizing mental illnesses into “madness” and “psychosis,” it is all too easy to similarly assume, upon the first hint of mental strain, that the narrator must be one of the “unreliable” type. It is accurate that a person suffering from a variety of mental illnesses may be genuinely unable to present a logical, cohesive, and true to (the widely accepted version of) reality narrative and could therefore be classified as “unreliable”. Simply having a mental illness or being in an altered state of mental clarity, though, does not necessarily preclude a person from being able to narrate an event.

As Rupert Read writes, “We have no criteria via which cognitively to evaluate [“mad” narrators], and so whatever we attempt to say of them by way of affirming characterization will be arbitrary, and in a way quite misleading...” (Read 2003, 122-123) (Grishakova 135). Each reader will, of course, bring to a text a unique set of beliefs and circumstances with which to read it; most readers, though, will not have experienced the same diagnosis or set of symptoms as a given narrator and will not have “criteria” to appropriately appraise a narrator’s character. In this way, it not only unfair but inaccurate and even “arbitrary” for the general reader, student, or scholar to judge what can and cannot be trusted from a mentally ill narrator unless there is explicit evidence against a narrator’s truth claim. Even the language used in literary discussions, such as the concept of “making sense” of an ill narrator carries the implication that the narrator fails to make sense on their own and needs others to “make sense” for them out of their personal experience (Ingram 242). These kinds of phrases are oppressive in their implications for what types of people are and are not able to speak for themselves.

The un/reliable binary terms do not allow for or acknowledge narratives that fall in-between

Particularly in recent discussions around other normative binaries such as male/female genders and homo/heterosexual orientations, it has become increasingly prevalent to think about people and literary figures on a spectrum rather than through traditional binaries. Unlike some of the other issues addressed here, contemporary narratologists have already spoken out against the inability to divide narrators cleanly between the reliable and unreliable categories. Brian Richardson suggests, “...we will be most effective...if we reject models that...insist on firm distinctions, binary oppositions, fixed hierarchies, or impermeable categories...numerous important narratives elude, problematize, or collapse these oppositions...Instead of dichotomous

typologies, we need to employ the concept of a spectrum...” (139). Like Richardson, others have favored a spectrum of reliability that allows a narrator to move along the gradient throughout the story, and disregarded Booth’s original model that required one of two categories of reliability between which a narrator could not oscillate (Olson 100). The un/reliable binary presents the additional problem of necessitating two possible readings of a text. In the very first textual example presented, a series of documents from *Girl, Interrupted*, two possible readings of Kaysen’s account are juxtaposed, each dependent on the reader’s declaration of Kaysen to be reliable or unreliable. There are, of course, countless concepts a reader could take away from Kaysen’s embedded documents and admittedly inconsistent memories. A focus on deciding who is reliable or unreliable, though, positions the reader to seek out one of two possible interpretations rather than exploring the full breadth and depth of interpretive possibility.

Monika Fludernik’s essay on *TYW* here offers a model of literary analysis reliant on declaring a narrator reliable or unreliable. While her synthesis and application of numerous definitions of unreliability to a specific text is laudable and useful, her analysis is problematic in its goal and perpetuation of an unforgiving binary. Fludernik posits:

This reading [Gilbert and Gubar’s] takes the insanity of the narrator for granted, and it likewise traces a development in the story from the depiction of an apparently sane narrator to the representation of an increasingly unbalanced mind that succumbs to outright lunacy at the end of the text.

There are several indications of such a progression, but that progression to a large extent resides in the incipient realization on the part of the reader that the narrator’s self-representation and even the factual basis of her account are increasingly less trustworthy (79).

She explains a traditional reading of *TYW*, which sees the unnamed narrator as initially reliable and sane, only becoming insane after she spends some time under the rest cure and locked in an attic. This most common reading is a feminist criticism of male control over women’s lives and treatments, S. Weir Mitchell’s rest cure, and inappropriately handled medical treatment of women. Fludernik’s reading is helpful in its detailed articulation of evidence that points toward *TYW*’s narrator’s unreliability, demonstrating a methodology with which one can arrive at a definitive decision regarding reliability status. In doing so though, she shuts down Gilbert and Gubar’s work that allows for a progression of sanity to insanity on more of a spectrum. Fludernik

returns to the traditional stance that the narrator must be either reliable or unreliable, and remains so consistently throughout the story.

Fludernik then discusses reliability as determined by what the narrator explicitly claims in writing compared to what the reader infers. She adds, “The narrator is unreliable to the extent that she pretends to be perfectly sane and, as we come to note with shock, turns out to be ravingly mad. In other words, in so far as there is a discrepancy between the discourse of the narrator and what we as readers reconstruct to really be the case, there is unreliability” (91). After referencing evidence such as the gnawed bedposts and “shmoosh” on the wall, Fludernik claims that the narrator may sound sane, but was actually quite “mad” from the beginning, which offers room to pose several countering questions to her reading here. Can we really say that the narrator is “pretending”? According to the Oxford English Dictionary, “pretend” means “To put forward as an assertion or statement; to allege, assert, contend, claim, declare; *esp.* to allege or declare falsely or with intent to deceive” (“pretend”). If the narrator truly is no longer in touch with reality, might she really believe what she tells us *without* intent to deceive? Perhaps more importantly, though, one must wonder if a narrator can only be sane or insane, reliable or unreliable. Though a reader may safely assume that the narrator is not communicating accurately when she says that there are women running around in the wallpaper, does that mean we must discount the entire narrative as being inaccurate? Fludernik does not allow for our narrator to exist in the grey space between terms, or for her “reliability” to fluctuate through the story.

All existing definitions of un/reliability depend on subjective characteristics that vary between readers and cannot be standardized in a single definition

One of the first qualms with Booth addressed here is his notion of what defines reliability itself. The reliability status is dependent, he claims, on the narrator’s proximity to a portrayed reality that readers interpret as normal. What an author deems normal will differ from what readers deem normal, and navigating so many versions of normalcy cannot be an effective way to also determine reliability. Charley Baker, Paul Crawford, B.J. Brown, Maurice Lipsedge, and Ronald Carter write the following of this narratological crisis:

Psychiatry relies on a relatively arbitrary but disciplinary necessary division between real and unreal. It is this binary that designates individuals as sane or mad... Because each individual’s reality is different – the same external occurrence will be experienced differently by each individual...fiction writers are

allowed to reinvent reality, and readers accept their versions as a way of experiencing alternatives (de Beaugrande, 1994, p.24) ... And by virtue of this very distrust, the problematics of “literariness” can never be fully disentangled from those of “madness.” We cannot maintain a genuine concern with literariness and yet ignore the commonplace defense whereby people who question established reality or propose an alternative have doubt cast upon their sanity... (Baker 29).

Though there is a clear need for separation between diagnosis and reliability, we also cannot ignore that changes in psychiatric terminology can and often do impact readings of reliability. Given mental illness’ long history of changing terminology and diagnostic criteria which designate them as social constructions relatively arbitrary in name, a character’s diagnosis, then, can be just as subjective as a reader’s judgment of the diagnosed character.

As discussed in both Booth and Nunning’s qualifications for un/reliability, the criteria they suggest seeks imprecise qualities of ambiguous meaning. Booth identifies reliability by actions or speech “in accordance with the norms of the work...the implied author’s norms” (Booth 158-159) and Nunning identifies unreliability as “...behavior deviat[ing] from the reader’s expected standards of normalcy...” (quoted in Fludernik 75). Marco Caracciolo presents concern against the ambiguity of these definitions by asking, “But what does it mean for a representation of a character’s mind to be convincing? In short, it means that this representation must match readers’ expectations about the workings of mind-expectations that are, in themselves, shaped by how the mind is conceptualized in a given culture at a given time” (Caracciolo 17). Caracciolo critically acknowledges here both that interpretations of un/reliability vary largely between individuals, and that an individual’s standard of norms is constructed from a culture’s dominant set of values at a particular point in time. He offers criticism, too, for Booth’s notion that reliability will remain consistent throughout a narrative; not only is a narrator’s mental state (and thus reliability) subject to fluctuation, but a community’s reading of a narrator’s reliability is quite fluid throughout time, too (1).

One other thing to note within this category of problematics is that reliability could change depending on what type of information a person is trying to learn about a narrative. To return to the opening Kaysen example, anyone but a physician who tries to tell the story of being a physician can be read as “unreliable” as they do not have accredited experience in such a role. Likewise, if one wants to know about the experience of being treated for a mental illness, would

a mentally ill narrator not be the most reliable source to speak on such a topic? In such circumstances, a flawed, anachronistic, or incomplete narrative may actually be more “reliable” in terms of learning about the actual experience and state of mind than a healthy-minded account of the same experience. In the case of Dr. Fried, psychiatrist of schizophrenic patient and protagonist Deborah in Greenberg’s *I Never Promised You a Rose Garden*, she appears to believe Deborah’s narrative accounts without doubt. After Deborah recounts a memory from her past intermingled with visits to the made-up world in her head, she proposes, “Maybe you doubt that I saw it all,” to which the doctor replies, “That is the one thing that I do not doubt” (Greenberg 107). Whether Dr. Fried truly believes every word Deborah tells her is unclear, but for her purposes as a physician, she counts Deborah’s account as trustworthy, an account that most readers would not.

Traditional un/reliable distinctions are limited by genres with which they are used

Booth’s and others’ original definitions of un/reliable narrators speak of narrators in fictional texts where the narrator is not the author. To them, the relationship between author and reader is of utmost importance; it is the author and reader who collude behind the back of the narrator in order to establish unreliability by tipping off the reader to an inaccuracy in the narration (Tammi 225). Within the fiction genre, another set of narrative distinctions is made between homodiegetic and heterodiegetic narrators. A homodiegetic narrator, or one who partakes in the narrative’s events as a character, is the most likely sort to be declared “unreliable.” Heterodiegetic narrators, ones who are disincorporate and removed from the story’s action, have very little to no surrounding discussion regarding their qualifications for un/reliable consideration (Olson 99). That question remains for a separate debate, should the un/reliable binary persist, but the debate’s necessity is only further evidence of the insufficiencies of the binary as a productive tool for narrative analysis.

In, say, a memoir or semi/autobiographical text, the narrator and the author are the same person; the narrator then cannot conspire with the reader about the narrator (Tammi 229). As definitions for unreliability have changed over the years, so have discussions of whether a nonfiction writer can also be an unreliable narrator. Memoirs and semiautobiographical texts, then, seem to hold the most significant possibility for narrative blurring of genres. Memoirs do offer an implied narrator that functions quite like a fictionalized character and could allow the memoirist to create a narrator separate from herself. But, returning to the question of intention

for reading a text, a reader most likely picks up a memoir to learn about the author's personal experiences, so how could an author write unreliably about her own experienced reality?

To segue into the final problematic element of unreliability, also commonly accepted in discussions of reliability in nonfiction is the principle that all narratives are at least partially inaccurate representations of reality, that the likelihood of some degree of unreliability between the text and the author's real life is quite considerable (Riggan 26). Even in nonfiction, there are variations in multiple versions of a story and things that are inevitably omitted (Tammi 230). If we approach a nonfiction or memoir text with the expectation that every bit of it is true, an author/narrator who admits to imperfect memory will be considered "unreliable" while other less honest writers who are equally faulty might not.

All narrators are, to some degree, inherently "unreliable"

Every observation, experience, and sensory intake of a human being is made subjective by the one experiencing it, never perfectly representing its respective entity in the real world. Therefore, it is impossible for a narrator, as a product of the human imagination, to portray a narrative in exact facsimile of the widely accepted state of reality in which we live. As Riggan writes,

"...Precisely because of these narrators' simulated humanness and because of the realism inherent in the situation of a character's speaking to us directly, the natural limitations of human knowledge and judgment and memory come into play... Thereby any possibility of absolute reliability with regard to all facts and facets of the events and characters within such a narrative field is cancelled. First person narration is, then, always at least potentially unreliable, in that the narrator, with these human limitations of perception and memory and assessment, may easily have missed, forgotten, or misconstrued certain incidents, words, or motives" (Riggan 19-20).

If this is the case, does reliability even really matter? Will a narrative just inherently be unrealistic no matter what we call its narrator? Though no narrator will be the perfect narrator, some will identifiably come much closer than others, reinstating the concept of a spectrum of reliability. When faced with only two options (reliable and unreliable), perhaps we should then call all narrators unreliable, but when considering a spectral approach, many more nuances and degrees reliability become available. Virginia Woolf also describes this idea in her essay, *On*

Being Ill, in which she argues that there is no adequate language to write about illness; the ill have to create their own words to describe their plight (Woolf 34). Something as personal and individualized as illness, Woolf argues, cannot be conveyed with conventional language – and certainly not with a simple set of binary terms. If language is insufficient in and of itself, any narrative account is bound to distort its reality; this should only provide more reason to cautiously reconsider the way we sort those whose stories we study.

Even if there are inherent linguistic disconnects and moments of unreality in every piece of writing, these variations of so-called “unreliability” are still worth studying. As the first example from Kaysen so clearly proves, discussing both the narrator’s mental health and his ability to be trusted can be valuable to literary discussion and scholarship, just not in the form that it currently takes. Old terms are often upgraded to new ones, but simply replacing un/reliable with something else does not fix the problematic aspects that accompany it. Regardless of what it is called, declaring a narrator to either be or not be truthful and worthy of our trust is a value-based judgment call, differing for each person, and lumping together multiple other facets of a character that, as a result, can go unnoticed or become unfairly blamed as a cause for the untrustworthiness.

Half a century after its coinage, Booth’s un/reliable narrator is still largely used in classrooms, discussions, and research with only minor modifications of its technicalities. Its use as a simple marker of a characteristic that a narrator either is or is not, especially when used for narrators with some kind of mental illness, does not sufficiently provide information about the complex nuances of a narrator. Given the binary’s problems and the contemporary direction of critical literary and narrative theory, the “unreliable narrator” is no longer a reliable marker of narrative quality.

Some time since Kaysen’s memoir, America began a gradual shift in focus toward issues surrounding mental health as institutions across healthcare, education, and politics begin to recognize the exigence of this debate. It is through this layer of narratological understanding and growing cultural urgency that we must reconsider the implications and ramifications of the theoretic understanding through which we engage mentally ill narrators, and real mentally ill people. After fifty years of critique and alterations to Wayne Booth’s original term, the “unreliable narrator” and its use with “mad” or mentally ill characters is still not sufficient to justly serve the complexity of the presuppositions that accompany it.

Unreliability and its relationship with mental illness has only been addressed with specific texts rather than within the larger context of narratology and literary theory. It is a gap in the conversation that must be undertaken, not only for literature, but for its larger implications on the non-written, or spoken narrative, too. Conversations between a doctor and patient are made of patient narratives relayed to a health professional who attempts to “read” and understand the patient’s story. Verbal narratives happen between the ill person and his family and friends, and then there is the much larger national and cultural discourse around mental health, something far greater than I can begin speaking to here. People living with mental illness are so often surmised to be in such serious condition that intermediaries are needed to interpret, translate, and make the important decisions for patients who are presumably unable to care for and advocate for themselves – people presumed to be unreliable. The values with which we critique literature reflect the values of a culture. In an age of desperate need for better understanding of mental health, let that too be reflected in the consideration and attention with which we approach literature.

Works Cited

- Baker, Charley, et al. *Madness In Post-1945 British and American Fiction*. Palgrave Macmillan, 2010.
- Booth, Wayne C. *The Rhetoric of Fiction*. 2nd ed., The University of Chicago Press, 1983.
- Caracciolo, Marco. *Strange Narrators in Contemporary Fiction: Explorations in Readers' Engagement with Characters*. University of Nebraska Press, 2016.
- Friedan, Betty. *The Feminine Mystique*. New York, NY, Dell Publishing, 1964.
- Fludernik, Monika, et al. “Defining (In)Sanity: The Narrator of ‘The Yellow Wallpaper’ and the Question of Unreliability.” *Grenzüberschreitungen: Narratologie Im Kontext*, Gunter Narr Verlag Tübingen, 1999, pp. 75–93, books.google.com/books?id=egN6LiSTjyMC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.
- Foucault, Michel. *Madness and Civilization: A History of Insanity in the Age of Reason*. New York, NY, Vintage Books, 1988
- Gilman, Charlotte Perkins. “The Yellow Wallpaper.” “The Yellow Wallpaper” & *Other Stories*. Ed. Paul Negri. Mineola: Dover Thrift Editions, 1997. 1-15. Print.
- Greenberg, Joanne. *I Never Promised You a Rose Garden*. St. Martin's Press, 2009.
- Grishakova, Marina. “The Voices of Madness: Performativity and Narrative Identity.” *Disputable Core Concepts of Narrative Theory*, edited by Göran Rossholm and Christer Johansson, Peter Lang AG, 2012, pp. 131–145.
- Ingram, Richard. “Reports from the Psych Wars.” *Unfitting Stories: Narrative Approaches to Disease, Disability, and Trauma*, edited by Valerie Raoul et al., Wilfrid Laurier University Press, 2007, pp. 237–245.
- Kaysen, Susanna . *Girl, Interrupted*. New York, NY, Vintage Books, 1994.
- Lutz, Tom. *American Nervousness, 1903*. Ithaca, NY, Cornell University Press, 1991.
- Olson, Greta. “Reconsidering Unreliability: Fallible and Untrustworthy Narrators.” *Narrative*, vol. 11, no. 1, 2003, pp. 93–109. *JSTOR*, JSTOR, www.jstor.org/stable/20107302.
- Phelan, James, and Peter J Rabinowitz. “Authors, Narrators, Narration.” *Narrative Theory: Core Concepts and Critical Debates*, The Ohio State University Press, 2012, pp. 29–38.
- Plath, Sylvia. *The Bell Jar: a novel*. New York, NY, Harperperennial, Modernclassics, 2013.

"pretend, v." *OED Online*, Oxford University Press, June 2018,
www.oed.com/view/Entry/150938. Accessed 22 August 2018.

Richardson, Brian. *Unnatural Voices: Extreme Narration in Modern and Contemporary Fiction*. The Ohio State University Press, 2006.

Riggan, William. *Pícaros, Madmen, Naiifs, and Clowns: The Unreliable First-Person Narrator*. University of Oklahoma Press, 1981.

Slater, Lauren. *Lying: A Metaphorical Memoir*. Random House, 2000.

Solomon, Andrew. *The Noonday Demon*. New York, NY, Scribner, 2001.

Søndergaard, Leif, and Christer Johansson. "Fictional and Factual Discourses in Narratives - and the Grey Zone Between." *Disputable Core Concepts of Narrative Theory*, edited by Göran Rossholm, Peter Lang AG, 2012, pp. 57–79.

Tammi, Pekka. "OTHIN RONG WIT HE UBTITLES. Remarks on Unreliability in Fiction and Non-Fiction." *Disputable Core Concepts of Narrative Theory*, edited by Göran Rossholm and Christer Johansson, Peter Lang AG, 2012, pp. 221–238.

Woolf, Virginia. "On Being Ill." *The New Criterion*, IV, no. 1, Jan. 1926, pp. 32–45.,
thenewcriterion1926.files.wordpress.com/2014/12/criterion-eliot-woolf-jan-1926.pdf.