BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg
BG PULSE JOURNAL

THE NARRATOR CATEGORY IN NARRATOLOGY AND THE EVOLUTION OF THE CONCEPT OF UNRELIABILITY FROM A RHETORICAL TO A COGNITIVE APPROACH

Narmetova Zebo Azamatovna
Trainee Lecturer, Faculty of Foreign Philology,
Urgench State University named after Abu Rayhan Beruni
E-mail: zebo.normetova @gmail.com
+998934674921
doi:https://doi.org/10.5281/zenodo.23597910

Abstract: Following Wayne C. Booth introduced the term "unreliable narrator" in 1961, the concept has become one of narratology's most productive yet most contested analytical instruments. The rhetorical model that inaugurated the concept defines unreliability as a discrepancy between the narrator's discourse and the norms of the implied author, whereas reader-oriented and cognitive approaches reconceive unreliability as an interpretive construct that emerges when textual signals conflict with readers' frames, knowledge, and expectations. Drawing on a critical literature review, comparative conceptual analysis, and brief illustrative readings of canonical cases (Poe, Ford, Ishiguro), this article traces the concept's evolution from its rhetorical foundations through its communicative reformulation to its cognitive reconceptualization. The analysis identifies the paradigms' divergent assumptions concerning the normative anchor, the model of the reader, and the explanation of interpretive disagreement, and argues that cognitive narratology complements rather than replaces rhetorical theory. A two-level synthesis—unreliability as an authorially designed condition actualized through readers' inferential activity—is advanced as the most comprehensive account. The article contributes a systematic comparative mapping of the debate and specifies directions for empirical, transmedial, and cross-cultural research.

Keywords: narratology, unreliable narrator, narrative unreliability, rhetorical narratology, cognitive narratology, implied author, reader response.

Introduction

Few categories of narratological analysis have proved as productive—and as contested—as the narrator. Narratology, the systematic theory of narrative that examines the structures, constituents, and functions of stories across media, assigns the narrator a pivotal role in the architecture of fictional communication. Genette treated the narrator's identity, level, and distance under the heading of voice, though his own account left unreliability largely untheorized. Chatman modeled narrative transmission as a communicative chain running from the real author through the implied author and the narrator to the reader, thereby distinguishing the biographical writer, the textual principle that organizes a work's norms, and the fictional speaker who tells the story. This triple distinction underlies virtually all subsequent theorizing, even though Bal has warned that

treating the narrator as a homogeneous category obscures structural differences between narrative levels and agents.

It was in this context that Booth introduced two concepts that reshaped narrative theory: the implied author and the unreliable narrator. A narrator, for Booth, is “reliable when he speaks for or acts in accordance with the norms of the work … unreliable when he does not”. The concept quickly became indispensable for interpreting a substantial body of fiction in which the act of telling distorts what is told: the well-meaning but obtuse Dowell in Ford's “The Good Soldier”, the anguished murderer of Poe's “The Tell-Tale Heart”, and the dutiful butler Stevens in Ishiguro's “The Remains of the Day” are canonical instances.

Over the course of four decades the concept remained embedded in the rhetorical paradigm, according to which unreliability names a discrepancy between the narrator's account and the implied author's norms and carries an irreducibly ethical dimension. Phelan refined the model into a systematic typology, yet the paradigm has also faced sustained criticism: the implied author's norms are inferred from the very text whose reliability they are meant to adjudicate—a circularity that Zerweck regards as disabling—and Nünning pressed the objection in the question that titles his essay, “Unreliable, compared to what?” Yacobi had already shown that identical textual data can be integrated by several competing reading hypotheses, of which unreliability is only one.

These criticisms converged with the cognitive turn in narrative theory, consolidated in Herman's programmatic volume. Fludernik grounded narrativity in experientiality. Herman recast narrative as a cognitive artifact, Palmer and Zunshine reoriented analysis toward fictional minds and theory of mind, and Nünning proposed a cognitive theory of unreliability grounded in readers’ frames and knowledge. In this paradigm, unreliability is less a property of the narrator than a function of the reader's inferential processing. The state of the debate was registered in Phelan and Rabinowitz's “Companion to Narrative Theory”, where Nünning explicitly called for synthesis.

While this debate is well documented, its terms have frequently been framed polemically, and syntheses have remained programmatic. The present study addresses a specific research gap: the absence of a systematic comparative framework that maps rhetorical and cognitive accounts of unreliability along explicit parameters and specifies the conditions under which each model affords the more adequate explanation. Building on Nünning's call, the article extends it in two ways: first, by confronting both paradigms with the challenges that revisions of the narrator category itself pose to any framework built upon it, second, by articulating a two-level account in which rhetorical design and cognitive actualization are complementary rather than rival explanations.

The aim of the study is to reconstruct the conceptual evolution of unreliable narration from a rhetorical to a cognitive understanding and to assess the theoretical relationship between the two paradigms.

The study pursues several interrelated objectives:

- to trace the historical development of the concept of the unreliable narrator within modern narratological theory;
- to compare the fundamental assumptions, analytical principles, and interpretive frameworks of rhetorical and cognitive approaches;

- to examine how readers identify, interpret, and evaluate narrative unreliability within these theoretical paradigms;
- to determine whether cognitive narratology extends, replaces, or complements traditional rhetorical models of narrative interpretation.

Methods

The study is a qualitative, theory-based investigation in literary narratology. Its results are theoretical findings—conceptual reconstructions and analytical conclusions—rather than empirically measured outcomes, and the methods were selected accordingly, because the research questions are conceptual and historical in nature, procedures of reconstruction, comparison, and clarification, standard in theoretical literary scholarship, are the appropriate instruments.

The study employed four complementary methodological approaches A critical literature review was conducted to examine foundational and contemporary scholarship on narrative unreliability Historical and theoretical analysis was used to trace the conceptual development of unreliability from Booths rhetorical model through reader oriented reformulations to cognitive narratology A comparative conceptual analysis was undertaken to examine the two paradigms across five analytical dimensions including the definition of unreliability its normative basis the conceptualization of the reader the role of textual signals and the interpretation of divergent reader responses Finally illustrative textual analysis was employed to assess the explanatory scope of both paradigms through brief analyses of selected canonical literary examples.

Theoretical sources were selected according to three criteria:

a) foundational status for the categories of narrator, implied author, and unreliability;

b) explicit theorization of unreliability;

c) representative standing within the cognitive turn.

Priority was given to peer-reviewed journals and university-press monographs in English, or canonical English translations, published between 1961 and 2014.

Three literary examples—Poe's “The Tell-Tale Heart,” Ford's “The Good Soldier”, and Ishiguro's “The Remains of the Day”—were chosen because they recur as standard test cases in the secondary literature, permitting direct comparison of how rival models handle identical material. Primary works are cited by year of first publication and quoted from standard editions, page references are accordingly omitted. Two delimitations should be noted: no empirical reader study was conducted, and references to “the reader” throughout denote the reader as modeled within the theories examined. Where the article speaks of readers’ cognitive processing, it reports the claims of those theories, not independently verified results.

Results

The findings of the theoretical investigation are organized under five themes.

The analysis demonstrates that the narrator remains a fundamental category of narratology whereas also being internally contested. Genette's typology of voice and Chatman's communicative chain established the narrator as analytically distinct from both author and implied author. Yet the category is not monolithic Bal stresses the structural heterogeneity of narrating agents across levels, Wall defends the narrator–author distinction against suggestions that authors wholly masquerade as

their narrators, and Nielsen documents first-person narratives in which no personal voice can plausibly be posited, testing the category's limits. Any theory of unreliability inherits this instability, because unreliability is defined as a property of the narrator.

Booth's definition makes unreliability a relation between two textual instances: the narrator's discourse and the implied author's norms. Chatman systematized the model, Rimmon-Kenan treated unreliability among the textual indicators of narrative voice, and Phelan differentiated it along three axes—misreporting (facts), misreading (interpretation), and misevaluating (ethics)—while distinguishing bonding forms, which invite ethical sympathy, from estranging forms, which alienate the reader. The paradigm's strength is its account of communicative design and ethical effect. It explains, for example, how Stevens's punctilious reporting of a lifetime of service generates an ironic distance that is simultaneously an ethical judgment. Its weakness is the opacity of its normative anchor because the implied author and the norms are themselves reconstructions from the text, the criterion risks circularity, and different readers may reconstruct different norms—hence Nünning's question, "Unreliable, compared to what?"

An intermediate, reader-oriented stage recast unreliability as a communicative and interpretive problem. Yacobi argued that problematic textual data are resolved through five "integrating mechanisms"—contextual, generic, conventional, perspectival, and centrifugal—of which unreliability is only one, and that the prevailing hypothesis depends on the reading's framework. Nünning catalogued the signals that prompt readers to posit unreliability, distinguishing intratextual signals (contradictions, gaps, discrepancies between the narrator's and other perspectives), structural signals (violations of generic conventions), and intertextual signals (echoes of established unreliable narrators). In his cognitive reformulation, unreliability is defined relative to the reader's conceptual knowledge: world knowledge, genre conventions, and historically variable value systems. On this account, disagreement among readers is not a failure of the model but one of its predictions.

Cognitive narratology supplied the processes behind such inferencing. Fludernik grounded narrativity in experientiality and described reading as the "naturalization" of textual oddities by mapping them onto real-world frames; unreliable narration arises where the narrator's version cannot be so naturalized and readers must dynamically reframe the storyworld. Herman treats narratives as cognitive artifacts whose participants hold differently informed perspectives; his account of discrepant awareness explains unreliability as a knowledge asymmetry within the storyworld, while his later work situates interpretation in mental-model construction and story-driven sense-making. Palmer, building on Cohn's typology of consciousness representation, reconceived fictional minds as continuously constructed by the reader from behavior and discourse, making unreliability a problem of mind-attribution, Zunshine links the appeal of unreliable narration to theory-of-mind reasoning about nested mental states. Together these accounts relocate unreliability from the narrator's failure to the reader's inferential achievement.

Table 1 presents the principal characteristics of the two paradigms.

Table 1. Principal features of rhetorical and cognitive approaches to narrative unreliability.

PARAMETER	RHETORICAL APPROACH	COGNITIVE APPROACH
Definition of unreliability	Discrepancy between narrator's discourse and the implied author's norms	Divergence between the narrator's version and readers' frame-based inferences and knowledge
Normative anchor	Implied author's ethical, factual, and perceptual norms	Readers' real-world knowledge, genre conventions, value systems
Locus	Designed communicative structure (author–implied author–narrator–reader)	Interpretive processing of textual signals
Textual signals	Norm violations, ironic distance, contradiction	Intratextual, structural, and intertextual signals triggering frame revision
Reader	Implied reader; ethical co-participant	Actual reader; frame-builder; theory-of-mind attributor
Disagreement among readings	Treated as misreading of a recoverable design	Treated as a normal outcome of differing frames
Principal strength	Explains ethical effect and authorial design	Explains actualization, variability, historical change
Principal limitation	Norm circularity; opaque implied author	Risk of subjectivism; weakened ethical accountability

Building on the comparison presented in Table 1 three conclusions emerge from the comparison. These are the present study's analytical conclusions, distinct from established positions. First, the paradigms differ less in what they observe than in what they take as their explanandum: rhetoric theorizes the design of unreliable communication, cognition theorizes its actualization. Second, the models divide the empirical phenomena unevenly, as Olson shows, fallible, self-deceiving narrators such as Stevens or Dowell fit rhetorical analysis of bonding unreliability, whereas deliberately deceptive narrators strain the notion of discrepancy with norms the text barely reconstructs and reward frame-based analysis. Third, the two levels are logically complementary: a text can signal unreliability only if some designed discrepancy exists, yet signals become unreliability only when a reader’s frames register them as discrepant. Poe's narrator illustrates both levels at once: the protest “But why will you say that I am mad?” is rhetorically designed self-incrimination and, cognitively, a frame conflict between asserted sanity and the reader's diagnostic inferences. Cohn cautions, finally, that a purely cognitive definition loses the concept's discriminative force unless a standard of discordance between narrator and author is retained—supporting the two-level account over either monism.

Discussion

With regard to the first research objective the concept’s evolution is best described as a dialectical refinement rather than a linear replacement. Each stage preserves the founding insight

that telling can distort what is told while relocating the criterion of distortion from authorial norms to communicative integration to readerly frames. The trajectory is not unidirectional, rhetorical theory continued to develop after the cognitive turn, and the cognitive theory itself required a rhetorical question—"Unreliable, compared to what?"—to define its problem.

Regarding the second research objective, the differences are systematic—normative anchor, locus, model of the reader, treatment of disagreement, and explanatory scope (Table 1)—but they are differences of emphasis within a shared communicative framework, not incommensurable paradigms: both presuppose a text capable of indicating discrepancy and a reader capable of detecting it. The genuinely hard difference is normative. Cognitive models dissolve norms into frames, at the cost—identified by Cohn—of losing the standard that separates genuinely unreliable narration from mere naivety or limited knowledge; without such a standard, every ignorant narrator becomes "unreliable" and the concept is diluted.

Regarding the third research objective, the two traditions describe complementary phases of a single interpretive trajectory. Textual signals—whether rhetorical anomalies or Nünning's three signal classes—trigger frame revision, theory-of-mind reasoning models the narrator's mind and possible self-deception and the process culminates in the ethical evaluation that rhetoric alone theorizes. The cognitive tradition explains why this process varies across readers and historical moments, since frames and value systems change, the rhetorical tradition explains why its outcome is normatively charged rather than arbitrary.

Regarding the fourth research objective, the evidence supports principled complementarity. Nünning himself synthesized the approaches. Phelan's rhetorical model implicitly incorporates reader inference, and cognitive models require a designed signal set to avoid unconstrained subjectivism. The two-level account proposed here makes this complementarity explicit: unreliability is an authorially designed communicative condition whose actualization depends on readers' inferential activity. Several issues nonetheless remain unresolved. First, the circularity problem is mitigated but not dissolved: if norms are frames, the question of whose frames fix the discrepancy recurs at a higher level. Second, the implied author's status under cognitive scrutiny remains undecided. Third, claims about readers' cognitive processing await empirical validation of the kind envisaged by psychonarratology. Fourth, unnatural narratives that resist frame-based naturalization press both models, and enactivist reconceptions of experientiality may revise the cognitive framework itself.

Conclusion

The present study has examined the evolution of narrative unreliability from Booth's rhetorical model through reader-oriented reformulation to cognitive reconceptualization. In direct answer to the research questions the concept evolved dialectically, each stage relocating the criterion of unreliability from authorial norms to communicative integration to readerly frames. The principal differences concern the normative anchor, the locus of unreliability, the model of the reader, and the explanation of interpretive disagreement. Readers, on the combined account, identify unreliability through textual signals and frame conflicts, interpret it through theory-of-mind attribution, and evaluate it ethically, and cognitive approaches complement rather than challenge or

replace rhetorical models, because unreliability is a two-level phenomenon—designed in rhetorical communication and actualized in cognitive processing.

The article's contribution lies in its systematic comparative mapping, its five-parameter framework, and its explicit two-level synthesis, which specify the conditions under which each paradigm affords the superior explanation. Its limitations follow from its design: the analysis is theoretical rather than empirical, the modeled rather than observed reader stands at its center, and the illustrative corpus is confined to canonical Anglophone works frequently discussed in the secondary literature. Future research should accordingly pursue empirical reader studies of unreliability recognition, extend the two-level account to transmedial and digital narratives, examine cross-cultural and historical variation in the frames that ground unreliability judgments, and test the account against unnatural narratives in which the narrator category itself is unstable.

References

1. Alber, J., Iversen, S., Nielsen, H. S., & Richardson, B. (2010). Unnatural narratives, unnatural narratology: Beyond mimetic models. “Narrative, 18”, 113–136.
2. Bal, M. (1997). “Narratology: Introduction to the theory of narrative” (2nd ed.). University of Toronto Press.
3. Bortolussi, M., & Dixon, P. (2003). “Psychonarratology: Foundations for the empirical study of literary response”. Cambridge University Press.
4. Booth, W. C. (1983). “The rhetoric of fiction” (2nd ed.). University of Chicago Press. (Original work published 1961)
5. Caracciolo, M. (2014). “The experientiality of narrative: An enactivist approach”. De Gruyter.
6. Chatman, S. (1978). “Story and discourse: Narrative structure in fiction and film”. Cornell University Press.
7. Genette, G. (1980). “Narrative discourse: An essay in method” (J. E. Lewin, Trans.). Cornell University Press.
8. Herman, D. (2002). “Story logic: Problems and possibilities of narrative”. University of Nebraska Press.
9. Nielsen, H. S. (2004). The impersonal voice in first-person narrative fiction. “Narrative, 12”, 133–150.
10. Nünning, A. (1997). “But why will you say that I am mad?” On the theory, history, and signals of unreliable narration in British fiction. “AAA: Arbeiten aus Anglistik und Amerikanistik, 22”, 83–105.
11. Olson, G. (2003). Reconsidering unreliability: Fallible and untrustworthy narrators. “Narrative, 11”, 93–109.
12. Palmer, A. (2004). “Fictional minds”. University of Nebraska Press.
13. Phelan, J. (2007). Estranging unreliability, bonding unreliability, and the ethics of “Lolita”. “Narrative, 15”, 146–164.
14. Poe, E. A. (1843). The tell-tale heart. “The Pioneer, 1”.

15. Rimmon-Kenan, S. (2002). “Narrative fiction: Contemporary poetics” (2nd ed.). Routledge. (Original work published 1983)
16. Wall, K. (1994). “The Remains of the Narrative” and the narrator retained. “The Journal of Narrative Technique, 24”, 31–38.
17. Yacobi, T. (1981). Fictional reliability as a communicative problem. “Poetics Today, 2”, 113–126.

18. Zerweck, B. (2001). A cognitive critique of the concept of narrative unreliability: The fallibility of the narrator in British crime fiction. “Style, 35”, 31–50.