Autofiction and Authorial Unreliable Narration

In recent years, the concept of “the unreliable narrator” has been among the most debated within narrative theory. In the wake of a series of provocative articles from the late 1990s by Ansgar Nünning (Nünning 1997, 1998a, 1998b, 1999), questions have been asked again and again regarding on what basis we determine whether a narrator is unreliable and how broad the scope of the concept is. Is the presence of an unreliable narrator in a given text the result of an author’s intentional decision, or is narratorial unreliability a historically variable reader response to textual inconsistencies and/or changing cultural norms? Does the concept belong exclusively to fiction, or does it make sense to approach factual or “real” narrators with the same concepts we encounter in fictional narrators?

In this article¹ I will address these questions with reference to the genre of “autofiction” (a genre parallel to or a sub-genre of autobiography) with special attention to the Norwegian author Karl Ove Knausgård’s six-volume novel *Min kamp* (2009–2011). I want, on the one hand, to discuss how autofiction makes itself vulnerable to narratorial unreliability due to the complex truth status of the told and, on the other hand, to use this genre to question the concept of unreliable narrator in rhetorical criticism as dependent on an intentional act by an (implied) author. This is not to say, however, that unreliable narration cannot be a narrative technique deliberately used by an author in the creative act of novel writing, but that by using the concept exclusively in relation to this phenomenon, we lose sight of important alternative (but comparable) versions. Among these is the kind of unreliability we can encounter in autofiction.

¹ I would like to thank Professor James Phelan (Ohio) for his valuable comments on an earlier draft of this article.
1. The Real Fiction of the Self

When Serge Doubrovsky coined the term “autofiction” in relation to his 1977 novel *Fils*, he defined it, rather paradoxically, as “Fiction, of strictly real events and facts.”

Doubrovsky’s work was inspiringly provoked by the fact that Philippe Lejeune, in his influential 1975 study *Le pacte autobiographique*, failed to leave room for blending novelistic fictional writing and factual autobiography. Lejeune defined autobiography as a retrospective prose narrative produced by a real person concerning his own existence, focusing on his individual life, in particular on the development of his personality. (Lejeune 1989 [1975], 4)

Lejeune stated that if a reader was confronted with a text where author, narrator and protagonist were the same, he or she would not accept counterfactual events or incidents as a matter of fiction. “False” information in autobiographical writing, Lejeune claimed, would instead be related to “the order of lying” (“l’ordre du mesonge”).

Doubrovsky broke this dogma two years after Lejeune’s study when he published the novel *Fils* and proclaimed a new genre: autofiction. For Doubrovsky, it was characteristic of this new genre that it required homonymy between its author, narrator and character and that it played on the generic ambiguity of its contradictory pact: on the one hand, the

---

2 Doubrovsky coined the term on the back cover of the novel, where he wrote: “Autobiographie ? Non, c’est un privilège réservé aux importants de ce monde, au soir de leur vie, et dans un beau style. Fiction, d’événements et de faits strictement réels ; si l’on veut autofiction, d’avoir confié le langage d’une aventure à l’aventure d’un langage en liberté, hors sagesse et hors syntaxe du roman, traditionnel ou nouveau. Rencontres, fils de mots, allitérations, assonances, dissonances, écriture d’avant ou d’après littérature, concrète, comme on dit musique.” (Doubrovsky 1977)

“Autobiography? No, that is a privilege reserved for the important persons of this world, in the evening of their lives, and in a beautiful style. Fiction, of strictly real events and facts, autofiction if you will; to have given the language of an adventure to the adventure of a language in freedom, without wisdom and outside the syntax of the novel syntax, traditional or new. Interactions, threads of words, alliterations, assonances, dissonances, writing before or after literature, concrete, as they say music.” (translation mine)

Today, the term is included in French dictionaries, and in Canada it is widely used as a genre conception, printed on the cover of novels.
Autofiction and Authorial Unreliable Narration

work is claimed to be absolutely referential and factual; on the other, it is claimed to be a novel, that is: fiction.

The result is not only a contradictory pact but what Poul Behrendt (2006) has labeled a “double contract.” On the one hand, the text is subject to the rules governing our social and cultural interactions, that is, where people have real names and responsibilities, where (at least in our part of the world) there is freedom of speech, but also laws protecting us from defamation, etc. On the other hand, the text also relates to the communicative system of fiction, where reader and author have agreed that exceptions exist. Here, the distinction between lying and truth is disregarded: ideas and beliefs can be expressed, even if they are counterfactual, as contributions to the ongoing negotiation of the constitution of our culture and “reality.”

Here is not the time and place to follow up on the development of the concept suggested by Doubrovsky in detail, since it has been scrutinized and developed by several since then—most importantly perhaps by Gérard Genette in his *Fiction and Diction* (1993 [1991]). Genette suggested that all cases in which an author of fiction includes his own person (or a character with the same name as the author) in his fictional story should be considered autofiction. If so, works like Cervantes’ *Don Quixote* (1605–1615) and Dante’s *The Divine Comedy* (1555) would be included in the category. But as it has been noted, Genette thereby disregarded the second part of Doubrovsky’s characterization, namely that the work has to play on the generic ambiguity that comes out of the claim of absolute referentiality and absolute fictionality at one and the same time. Illustrated with another example: when in *City of Glass* (1985), the first part of Paul Auster’s New York trilogy, a character with the name “Paul Auster” appears and is furthermore presented as an author living in New York in an environment that reminds us of the real Paul Auster’s surroundings, it is less an autofictional than a metafictional strategy that is being implemented. The intention is not to tell the story of Paul Auster in fictional terms, but rather to enter a mode comparable to romantic irony.

What autofiction does is quite radical in the sense that instead of demarcating fiction from reality it blurs the border. This can be accomplished by promoting a picture of the authorial self which confirms, negates, transforms or plays with the public understanding of this self. Such is the case of Bret Easton Ellis’s *Lunar Park* (2005), where the author mixes known public events and incidents from his real life with counterfactual and even paranormal events and incidents. The blur
between fiction and reality can also result from using the novel form and the narrative techniques related to this form to investigate and depict the history of the author's own self. This was what J. M. Coetzee did in the three volumes of autofictional memoirs—Boyhood (1997), Youth (2002) and Summertime (2009)—when he chose to let the first two volumes be told heterodiegetically in third-person narration and the third to take the form of a fictitious biographer’s interviews with five factual people from Coetzee’s past. Coetzee himself is claimed to be dead in the novel, and his voice only appears through a number of third-person fragments of the kind the reader already knows from the first two volumes. But even though no one would question that the trilogy most certainly is written in fictional terms and therefore cannot be understood literally, the books are also understood to be about Coetzee’s own life. In this perspective, autofiction pushes what is a general paradoxical characteristic of practically all storytelling, namely that stories tell the (or some) truth, even though what they are telling might not have happened.

It is therefore also evident that we cannot always rely on the factuality of the story being told by the author. But can we approach this authorial unreliability in terms of unreliable narration?

2. Factual Unreliable Narration?

It has been claimed that only in fictional narrative can we have true cases of unreliable narration. The argument goes that narrative unreliability depends on, if not difference, then at least on distance between narrator and authorial agent. In her essay on discordant narration, Dorrit Cohn claims

that the diagnosis of ‘discordance’ can apply only to a fictional narrative, not to the kind of storytelling (oral or written) that presumes to refer to real facts: though we often apply the term ‘unreliable’ to voices we regard as wrong-headed in non-fictional works (historical, journalistic, biographical, or autobiographical), the narrator of such works is the author, the author is the narrator, so that we cannot attribute to them a significance that differs from the one they explicitly proclaim. (Cohn 2000, 307)

This assumption was challenged by James Phelan in the chapter from Living to Tell About It on Frank McCourt’s memoir novel Angela’s Ashes: A Memoir (1996).

Phelan demonstrates how McCourt has “built his narrative on the foundation of unreliability and the virtual absence of reflection in his
authorial voice.” McCourt’s “trick” is, as Phelan writes, to conceive “memoir,” not as an art of direct telling from author to audience, but as an art of indirection. Rather than speaking in his own voice at the time of the telling, McCourt uses the historical present and speaks in the voice of his former self […] at the time of the action. (Phelan 2005, 67)

What makes the narratorial unreliability possible in McCourt’s case is that he uses what Phelan recognizes as a “nonstandard technique” for his autobiographical purpose when he “re-invents” his childhood-persona’s perspective on the incidents. This technique is, on the other hand, a standard novelistic, fictionalizing technique: McCourt-the-Author has given the voice and the perspective to a less experienced narrator, Frankie (McCourt-the-boy-character), and even lets him tell from the moment of the action that is in the first person and in the present tense. Hereby the difference or distance claimed by Cohn as a necessity for the formation of an unreliable narrator is reinstated.

We might therefore also conclude that the extended use of this fictionalizing technique relates McCourt’s memoir-novel to the genre of autofiction more than it does to autobiography proper. Angela’s Ashes is “[f]iction, of strictly real events and facts,” and it suggests homonymy between author, narrator and character, although this homonymy cannot be considered absolute: it is nominal and personal, but not temporal. Furthermore, it plays on the generic ambiguity of what Doubrovsky called autofiction’s “contradictory pact”: on the one hand, the work claims to be absolutely referential and factual; on the other, it makes extended use of a standard novelistic technique and therefore relates to fiction, too, at least in Richard Walsh’s understanding of fictionality as a rhetorical rather than an ontological quality. Fictionality should not, in this perspective, “be equated simply with ‘fiction,’ as a category or genre of narrative: it is a communicative strategy, and as such it is apparent on some scale within many nonfictional narratives” (Walsh 2007, 7).

But what would the result be if one’s attention were directed toward a work that is not composed with the same obvious distance between the authorial I, the narratorial I and the character I? If we were to follow Phelan’s line of argument, where unreliable narration is defined as narration “in which the narrator’s reporting, reading (or interpreting), and/or regarding (or evaluating) are not in accord with the implied author’s,” (Phelan 2005, 219) then it is quite unlikely that the narrator would be unreliable in such a case, since the necessary distance between
the subject positions is eliminated from the very outset. The question, then, is whether Phelan’s exclusion would be correct—or whether it is the result of too narrow a conception of unreliable narration.

3. A Norwegian’s Struggle with his Past and Present

That it is the latter which is the case I will try to demonstrate by looking at an example of autofiction which makes us consider the reliability of the narrator without seeking recourse to the concept of the implied author. The example is the Norwegian author Karl Ove Knausgård’s six-volume autofictional novel *Min kamp*, meaning *My Struggle*.3

*My Struggle* is among the greatest literary sensations in Scandinavia in decades. Not only because it has been a huge sales success, but also because the publication of the six volumes has been accompanied by a heated debate about the use of autobiographical elements in fiction, and vice versa. In the six books, Knausgård gives a detailed description of his life from the day he was born until the moment he types the final sentence of the manuscript of volume 6, taking into account the reception of the first volumes of the project and the effect it has had on himself and his relationships. Knausgård tells the story in the first person and from the position of the writing situation, the first and the last volumes in particular containing long essayistic passages reflecting on life, death, art and literature. In long sections he changes the focalization and even the narrative tense from past to historical present, so that it isn’t the narrating I’s but the experiencing I’s (Karl Ove’s) perspective we are confronted with. In these chapters there are numerous examples of the kind of unreliability Phelan focused on in McCourt’s case, with the difference, however, that unreliability is established here due to the distance between a diegetic narrator and a fallible filter.

But these examples are less interesting in the present perspective. It is more interesting that there are aspects of the work which raise doubts about and even undermine the author-narrator’s reliability.

3 The first volumes have been published in several languages including German, English and French. Due to the obvious allusion to Adolf Hitler’s *Mein Kampf* (an issue to which attention is given in the sixth volume of the series), the books have appeared under other titles than in Scandinavia. The first volume is called *Sterben* (2011) in German, *La Mort d’un père* (2012) in French and *A Death in the Family* (2012) in English. In the latter case, the series has been called *My Struggle*. In the following pages I will refer to the series name.
In volume two, in the midst of a very detailed retelling of a long conversation during a dinner party at a restaurant, Karl Ove and his friend Geir agree upon the fact that Karl Ove has always had an extremely bad memory. Due to the frame, the statement becomes a variant of Epimenides’ paradox: “All Cretans are liars.” An author giving a hyperdetailed account of a long-past situation cannot in the same situation claim to have a bad memory. The impact of this inconsistency is quite remarkable. The conversation between Geir and Karl Ove is on a very highbrow intellectual level of almost Socratic-dialogic dimensions, focusing on, among other things, the personal costs connected with “the authorial call,” etc. But the remark concerning Karl Ove’s weak memory makes the reader speculate whether the whole dialogue isn’t more a matter of the authorial I’s creative attempt to put himself in a better light, due to the decisions regarding family life and relationships he has had to make to pursue his career as an author. We begin speculating whether Knausgård-the-author’s reporting is reliable.

Another example: over the course of the six volumes, Knausgård returns to some of the significant incidents in his life. Among these is his complicated relationship to his tyrannical and choleric father who died of alcohol abuse ten years prior to the time of writing. The last period of the father’s life, as it was experienced by Karl Ove, is focused on in both the first and the last volumes. But in the latter case, the description is much more downtoned than in the former. And where Karl Ove and his older brother Yngve were pretty much alone in cleaning up the mess after the father in the first volume, they receive attention and help from an uncle and aunt in volume 5. Both descriptions focalize the incidents through the experiencing I, but it seems as if the narrating I—the author—through the process of writing the novel, develops his understanding of the story as it is told, a contradiction that exposes a dynamic unreliability regarding values and judgments. In volume 6, Knausgård receives an angry letter from his uncle (his father’s brother) in which the author’s version of the incidents in relation to the father’s death is being questioned with reference to supposed hard evidence. And even though Knausgård becomes aware that there might be flaws and misjudgments in his former retelling of the incidents, he is not abandoning it. As he says to a prosecutor in an imagined trial: “This is how I remember it” (volume 6, 303; translation mine). But as we (and he himself) know: he has an extremely bad memory.

The death of the father is not the only example of this kind of unreliability. In volume 1 we are told that Knausgård spent four years
with a girl he didn’t love; but when the relationship is described for us in the fifth volume, it most certainly seems characterized by love. Again: the authorial perspective on and understanding of the incidents has taken over, and the result is that our natural urge to rely on the authorial voice of the narrative is challenged. Where our expectation with regard to the authorial narrative agent is that it is stable and sanctions the norms and values of the storyworld, we are instead engaging with inconsistencies, and the result for our reading is that we redirect our attention from the told toward the teller and thus see the misrepresentations and inconsistencies as an expression of character traits and unreliable narration. In that sense, Knausgård is, either intentionally or unintentionally, flouting the assumption of a stable author. In the beginning, we greet the representation as reliable. But as the telling and retelling progress, we start having second thoughts, just as Knausgård himself does in the sixth volume, when (as pointed out above) he refers to memory instead of fact.

The examples commented on so far are all intra- or internarrational insofar as the effect of unreliability is a result of contradictions within the narrator’s discourse or between his discourse and the differing perspective of others (e.g., his uncle’s) perspective on the same incidents.

But there are also examples of extratextual circumstances influencing the author-narrator’s reliability. Even though the books were published as “novels”, Knausgård claimed that everything he told was true. All material had been presented to the persons concerned, and only a few names were changed at their request.

Everything in this work is depicted with an attention to detail that marks the work as a fictional recreation of the past: no one can remember their past as clearly and elaborately as Knausgård does. And even though we as readers might accept the level of detail, bearing in mind that this is also a work of fiction, the authenticity is punctured from inside the storyworld by the aforementioned unreliability signals.

Due to the work’s claim to factuality, moreover, it was also punctured from the outside by the persons and family members depicted. Several felt exposed and misrepresented and expressed their displeasure and disappointment through the media. Fourteen family members even

---

4 The concepts intranarrational, internarrational and extratextual unreliability belong, together with intertextual unreliability, to a taxonomy of different ways of signaling/detecting unreliable narration. For further detail, see Hansen (2005, 2007, 2009).
announced that they wanted to bring Knausgår to trial for his “Judas literature.”

Now, Lejeune claimed that when an author writes in his own name, he signs an autobiographical pact in which everything stated is to be considered true. If something turns out to be false, it is not a matter of fictionalization but, as quoted earlier, relates to the “order of lying.” This rather strict understanding of autobiography’s truth value has been revised in later conceptions of the genre. As Phelan notes, autobiographical theory has repeatedly shown that “subjective truth is far more important to memoir than literal truth [...] because it is crucial to the autobiographer’s ability to give shape and meaning to experience” (2005, 73). But as he also remarks, subjective “truth must also be accountable to some extent to facts, people, and events that have an existence independent of the autobiographer’s perception” (73).

With reference to Lejeune’s initial distinction, it can be observed that a radical subjective recounting of the story jeopardizes the author’s reliability and makes us, the readers, react in the same way as when we get suspicious about a narrator’s account in fictional narrative: we read with precaution and look beyond the authorial representation of the facts; we try to figure out the true facts of the case and use our observations to construct a critical picture of the authorial self about which we make ethical judgments. We might even press charges against this self due to the fact that autobiography belongs to the system of factual communication that is restricted by responsibilities and protection against defamation.

It is also in this respect that we find the main source of the difference between the everyday use of the concept of unreliability and its use within the context of fictional narrative: where the two forms of unreliability both invoke the reader’s ethical judgment, only the latter opens up for an esthetic judgment.

Autofiction blurs this distinction. Insofar as Knausgår’s work is also claimed to be fiction, the system of factual cultural interaction and communication is partly suspended, and instead the rules governing fiction takes over: whatever is depicted is fictionalized, and thus a simulated reality is told from a given perspective. Considered an author of a work of fiction, Knausgår cannot be blamed for the misrepresentation.

5 Uncle Gunnar (which isn’t his real name) was among the most critical voices to the project, notably in a commentary published in the newspaper Fædrelandsvennen (cf. Kristensen 2011). Knausgår’s estranged ex-wife, Tonje Aursland, also retorted in October 2010 in a radio documentary broadcast on NRK.
of persons and events or for his exaggerated remembrance of details. Under the auspices of novelistic fiction, this kind of re-creational depiction of the past is expected and allowed.

This, of course, also has significance for the reader’s approach to the authorial self. Due to the many misrepresentations in Knausgård’s narrative, the ethical judgments passed on both the actions of the author and the characters and the critical and questionable aspects of the author’s telling are accompanied by esthetic judgments. The narrative is not read as an author’s telling about his past, but as a refuguration of that past.

By making himself the author, the narrator and the protagonist of his work and at the same time claiming the storyworld to be in accordance with reality, Knausgård’s result is more or less doomed to be an act of unreliable narration. Reliability is dependent on a perspective that can tell the truth. This is why, on a general level, it makes sense to speak of fictional truth, insofar as narrative fiction is the creation of a storyworld from a given perspective. But reality as such does not support one truth alone. Truth in reality is perpetually being negotiated. When Knausgård claims that he tells the truth, but does it in the framework of fiction, he suspends the negotiable nature of factual truth. By doing so, he is both hit and saved by what we could call the “kernel paradox” of autofiction: he claims that what he tells is true, thus opening up the possibility of negotiation, while at the same time claiming that what he says is fiction, authorizing the truth value of the told. It is in this tension between author and narrator that Knausgård’s unreliability comes into being. More generally, autofiction as a genre must at the very least be suspected of extratextual unreliability.

4. Auto-Mocking and Self-Deception

From the perspective of unreliable narration as narration “in which the narrator’s reporting, reading (or interpreting), and/or regarding (or evaluating) are not in accord with the implied author’s” (Phelan 2005, 219), autofictional authorial unreliability of the kind we find in Knausgård’s novels will clearly fall outside the definition. We could therefore choose to invent a new term to cope with these deviant cases. But insofar as there are quite a few common aspects between the two

Phelan (2011) has suggested that we should distinguish between unreliable and deficient narration, the latter covering an example such as the one I have discussed here.
sorts of unreliability, I would like to avoid throwing yet another narrative term on the heap of narratological concepts and models. Let me instead close this article by considering whether room can be found within the standard conception that will allow us to accommodate narrative works such as Knausgård’s.

The concepts that serve as the toolbox we label “narratology” come from a great variety of disciplines and discourses and are brought to foreign areas thanks to the fact that someone has found an item that lacks description but shares significant aspects with another item that has already been described. The reason for this travelling applicability is the fact that narrative is a transtextual, transgeneric and transmedial phenomenon. Applying general narrative concepts developed in one context to comparable items in another context gives us the added advantage of providing our analysis with considering whether the conception has been biased by the fact that it is formulated within a framework of one genre, mediality or communicational mode.

Scholars from literary studies in particular have provided narratology with a nearly endless series of useful concepts for the transmedial and transdisciplinary study of narrative. But they have also quite often made themselves guilty of making general claims as to what narrative is and does, whereas what they are actually describing are specific features for prose fiction, not narrative in general.

The concepts of unreliable narrator and unreliable narration have travelled between genres and modalities ever since Wayne Booth coined the term more than fifty years ago in his study of the rhetoric of novelistic narrative fiction. Initially, at least, he did not consider whether the concept was of general relevance or limited to literary fiction. Even so, it seems that Booth did have a vague idea that there was more to the story than he told.

Booth has often been blamed for the inconsistency of his initial formulation:

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 1991 [1961/83]: 158–159)

All discussions of the concept since then have their basis in differences of accentuation in this definition. Rhetoricians working on literary narrative have focused on the parenthetically mentioned authorial agent,
intentionality and normativity (understood as ethics) whereas more semiotically oriented approaches have highlighted the lack of accordance between the narrator’s representation of the storyworld and the actual constitution of it. One might claim that the former has developed the Boothian concept with respect to the fact that he was working with literary, fictional narrative, whereas the latter have tried to include it in more transdisciplinary approaches.

If we can agree that the concept has proven its value as a traveling concept, I think we owe it ourselves to reconsider Booth’s definition and liberate the “narrative/narratological” part of the concept from the literary part.

By focusing on unreliability in Knausgård’s autofictional case, it is clear that the implied author component is not a necessary part of the unreliable narrator machinery. Rather, it is a concept that proves helpful in the distinct cases of unreliable narration in which it is used as technique for what Phelan calls “indirection.” This, we could claim, is a feature related to fiction, a matter of fictionality, and thus important to study in the study of narrative fiction.

With regard to the reliability of the narrator, seeking explanatory recourse in the implied author is a matter of framing. What relates unreliable narrators across the fact/fiction distinction, across genres and modalities, is that they do not speak or act in accordance with the norms, values or facts of the storyworld. As discussed above, it is characteristic for fictional storyworlds that their constitution can be authorized by an authorial agent. In factual narratives, by contrast, author and narrator are often (as Cohn stated) the same. This, however, does not rule out the possibility of unreliability. The storyworld is simply not governed by an implied author in these cases, but rather by sensus communis to the extent that readers have a stake in it.

Autofiction is a special case, since it blurs the borders between author, narrator and character as well as between fact and fiction. In itself, it is an extremely unreliable genre, and it can thus take either the form of intentional auto-mocking, as in McCourt’s example, or that of seemingly unintentional self-deception, as in Knausgård’s. So perhaps, after all, there is still room and a need for new conceptions and distinctions.

Works Cited


