Chapter 4

Narrative Reliability

Margaret Atwood’s Surfacing

As the preceding chapter begins to suggest, one of the most fundamental interpretive questions about discourse concerns reliability. Indeed, the very distinction between narrator and implied author is in large part motivated by the desire to clarify the difference between reliable and unreliable narrators. Specifically, here as elsewhere, the implied author provides a norm against which one may judge the narrator. Fundamentally, the narrator is unreliable to the extent that he or she diverges from the implied author or “is at virtual odds with the implied author,” as Seymour Chatman put it (Story 149).

On the other hand, this may not tell very much. As Nünning points out, “Critics who argue that a narrator’s unreliability is to be gauged in comparison to the norms of the implied author just shift the burden of determination” (“Reconceptualizing” 91). This is particularly problematic because the very idea of an unreliable narrator, like that of the implied author itself, should serve an explanatory function. As Tamar Yacobi points out, for any given reader facing a particular text, the notion of an unreliable narrator is a “hypothesis” that is “formed in order to resolve textual problems” (110). But this, in turn, raises the need for a “systematic account of the clues to unreliable narration”—in other words, the “textual data” and their relation to “interpretive choices” (Nünning “Reconceptualizing” 105). To begin with, there is a need for a clearer sense of what constitutes unreliability.
What Is Reliability?

People routinely consider real individuals to be unreliable informants. To say that someone is unreliable in ordinary life is, first of all, to say that he or she misinforms one about the real world. In other words, an unreliable informant is someone whose discourse leads one to imagine some aspects of the world incorrectly. The idea of unreliability is fundamentally the same in fiction. A fictional narrator is unreliable to the extent that he or she misrepresents the storyworld, leading one to imagine the storyworld incorrectly. Recognizing unreliability is, then, inseparable from recognizing this incorrectness. As Shen puts it, “A narration is regarded as unreliable precisely because the reader has come to the conclusion that things are not or cannot be as the discourse represents them” (“Story-Discourse” 567).

There is, however, a difficulty here. In the case of real life, there are facts. In real life, if I am told that Jones married Smith, there is a fact about whether she did or did not marry Smith. But when the narrator of Surfacing says that she got married, what determines the fact that she did not?

By this point, readers will not be surprised to discover that narratologists do not entirely agree on the topic. Some stress authorial intent. The facts of the storyworld are, in this view, determined by the author. So too are the facts about the narrator, and his or her reliability or unreliability. In contrast with this, there is a “reader-centred . . . approach.” This posits “an interpretive strategy that naturalises textual anomalies by projecting an unreliable narrator figure” (Nünning “Reliability” 496). But, in order to make a distinction between reliability and unreliability, one needs a norm of some sort. It is difficult to see how a reader’s “naturalization” of “an interpretive strategy” constitutes such a norm. Here, again, the different views are at least partially reconciled in the implied author, understood as the receptive or reader-simulating intent of the real author.

But here another problem arises—how is it even possible to recognize unreliability in fiction? In the real world, one discovers unreliability by discovering facts, usually by some means outside the report of the unreliable informant (e.g., by meeting Jones and finding out that she did not marry Smith, despite what one was told by Gallstone). But how does this occur in fiction, since the only source of information is often that unreliable informant?

In order to consider this problem, it may be helpful to distinguish different sorts of unreliability. In fiction, as in real life, an unreliable dis-
course is likely to have one of three faults. First, it may directly mis-
represent the facts of the story. This is “misreporting,” in Phelan’s
terminology (see “Rhetoric/Ethics” 205). Second, an unreliable discourse
may leave out information that is crucial for valid inference (“underre-
porting,” in Phelan’s terms [“Rhetoric/Ethics” 205]; not being “forthcom-
ing,” as Bordwell puts it [60]; limiting communicativeness, as Sternberg
would say). Finally, such a discourse may misdirect attention, promoting
mistaken inferences.3

The first “mode of factual unreliability” is the simplest. Having been
given a piece of information (or misinformation), one is most likely to be
given a piece of directly contradictory information (or misinformation).
In some cases, this involves a narrator’s self-contradiction; in other
cases, there may be a shift in narrators; in other cases, the contradictory
information may be reported by a character in dialogue. In Surfacing,
the narrator first reports that she was married, and then subsequently
reports that she was not. This sort of direct contradiction signals to the
reader that the narrator is unreliable in one of those two reports. It
thereby requires the reader to infer which report is correct.

The complete withholding of information is usually fairly straight-
forward as well. In this case, the reader is most often faced with a gap in
the Iserian sense. He or she simply cannot figure something out and ulti-
mately realizes that information is lacking. This situation is a bit trickier
for the author in that he or she typically must provide enough clues that
the reader can infer the withheld information (though some works sim-
ply leave the information incomplete, making it part of the profile of
ambiguity).

The misdirection of attention is perhaps the most interesting of the
three. The direction of attention is one of the main concerns of liter-
ary authors and film directors, since readers simply cannot notice every-
thing, as Rabinowitz discusses (Before 47–52; on direction of attention
in film, see Bordwell On the History). One may distinguish two types of
attentional misdirection. In the cruder case, some true information is
made salient and that information suggests a false inference. This hap-
pens all the time in detective fiction, though it may be developed in such
a way that the reliability of the narrator is not impugned. For example,
suppose Jones has been shot. The reader then learns explicitly (and accu-
rately) that Smith had a fight with Jones the previous day and that he
had a rifle. This may lead the reader to infer, falsely as it turns out, that
Smith was the killer.
This form of “misdirected foregrounding” is commonly paired with the second sort of unreliable direction of attentional focus—the reduction of salience or “misdirected backgrounding” of information that is crucial to valid inference. Thus, along with the misleading information about Smith, a reader may find himself or herself distracted from information that would lead to valid conclusions. In the case of film, the information may be literally foregrounded or backgrounded. For example, the audience might be shown Smith cleaning a rifle, while in the case of Doe (the real killer) rifles appear in the background on a wall in his home.

Usually, the trick with foregrounding and backgrounding of information is manipulating the relation between spontaneous encoding and processing, on the one hand, and self-conscious inference, on the other. If the reader is explicitly told that Smith went out and bought a rifle after the fight with Jones, that will almost certainly lead to self-conscious reflection on the likelihood that Smith is the murderer. Given the nature of detective fiction, that may actually lead the reader or viewer to reject the inference as too straightforward. On the other hand, simply showing rifles in the background of a shot may not lead to encoding. In other words, the viewer may simply not notice the rifles. In that case, the conclusion of the story—in which Doe is exposed as the murderer—may not produce the required feeling of “retrospective necessity,” which is to say, the feeling that, once revealed, it makes perfect sense that Doe is the culprit.

Whether the narrator provides false information, fails to provide necessary, true information, or misdirects attention, his or her unreliability may result from deceit or incompetence, as Chatman notes (Story 149). It may also arise from cognitive or emotional bias (e.g., in-group preference). Incompetence—or fallibility—is further divisible into “range and depth of knowledge,” as Bordwell explains (Narration 60) or, one might add, articulateness. A reader’s emotional and cognitive response differs across these types of narrator. One is likely to feel antipathy toward deceitful or biased narrators. In contrast, one may feel a sort of identification with ordinary fallibility and compassion or pity for a cognitive or perceptual disability. In response to deceit, people typically look for motives, particular interests that would guide the narrator’s misrepresentations. Understanding these motives may help to correct for the misrepresentation. If the problem is a disability—as in the case of Benjy, a mentally challenged narrator in Faulkner’s The Sound and the Fury—the
reader may try to infer what facts would lead this particular character, with his or her particular disability, to this particular misunderstanding. In any case, one is likely to seek further information, from the same or other narrators.

The topic of emotional response leads to another key issue regarding unreliability—trust. For example, as noted above, a reader or viewer of detective fiction often does not trust the narrator who gives apparently definitive information in a highly salient form. Indeed, trust is arguably the initial basis of readers’ and viewers’ sense of reliability or unreliability. Talking about unreliability solely in terms of facts may make the process appear too inferential, particularly at the start. Perhaps one’s first sense of reliability or unreliability is not an abstract, cognitive relation to ideas, but rather a concrete, affective relation to a person.4

Indeed, it seems very likely that, in fiction and in life, emotional trust or mistrust of a person precedes a rational, self-conscious inference regarding the likely validity of particular statements. This is roughly the same process that is found in the well-known experiments testing ventromedial frontal damage. Test subjects are presented with four decks of cards. Unbeknownst to the subjects, the decks are rigged. Normal test subjects develop a spontaneous, emotional aversion to the losing decks and a spontaneous, emotional attraction to the winning decks “before they could consciously articulate the best strategy” (Oatley, Keltner, and Jenkins 136, discussing work by Bechara and colleagues). One may guess that things operate similarly in the case of trust. Indeed, one could construe normal subjects’ response to the different stacks of cards as one of trust and mistrust.

There is one possible difference between trusting a stack of cards and trusting a person. It appears that one’s default response to strangers is a mild and tentative distrust. People only “build up trust” (or enhance the response of distrust) over the course of personal interactions—or, in some cases, by just seeing the other person repeatedly even without interaction (see Oatley Emotions 73). However, there may be an exception to this. It seems likely that people feel a default trust for people placed in positions of institutional authority. The point is suggested by the “Dr. Fox lecture” experiment. In that study, test subjects were presented with an actor playing the role of an academic. Simply because of his stated position as an authority (the participants were not told he was an actor), they accepted what he said and responded positively to this (pseudo-)learning experience, forgoing their usual spontaneous distrust for strangers (see Naftulin, Ware, and Donnelly).
The point about authority is consequential because narratives place narrators in a position of structural authority. Thus, in all likelihood, readers begin with a presumption of narrator reliability and lose trust in a narrator only once they have reason to do so. This leads us to the issue of just how such distrust develops. However, before considering this, we need to address a logically prior issue. Even if an affective relation of trust is primary, it seems clear that distrust is connected with a sense of discrepancy between what the narrator reports and what the facts of the storyworld appear to be. The nature of such facts is fairly straightforward in the case of nonfiction. But the case of fiction is less clear. Thus, before addressing the development of distrust, we need to consider the nature of fictional facts—thus how there can even be anything to distrust a narrator about.

**Fiction’s Facts**

In considering the facts about a fictional world, it is helpful to begin with a broad division between general principles and particular instances. General principles are the patterns that govern physical events, individual actions, group dynamics, and so forth. For example, people generally act in such a way as to further their goals; people generally experience empathy when they witness the face of someone in pain; people generally have five senses. Such principles range from strict laws, such as the law of gravitation, to looser tendencies, such as the idea that empires tend to become overextended and disintegrate. Particular instances are cases of these principles—Smith is a case of a person with five senses, or perhaps four senses, if he is blind or deaf; Imperial Rome is a case of an empire; Jupiter’s orbit is a case of the law of gravitation.

People create or understand a fictional story most obviously by constructing or reconstructing particulars. To a great extent, these particulars are explicitly stated in the discourse. Or, rather, the discourse presents enough information so that one may draw on general principles to reconstruct “nondefault” or unique aspects of the story’s particularity. Indeed, this is true in nonfictional storytelling as well. In a fictional story, one might read “Jane tripped when running up the stairs. Now she had a prominent, bluish bump on her cheek.” Similarly, if this happens to a real person, she may explain her prominent, bluish bump by saying “I tripped on the stairs.” In both cases, recipients use general principles about the world to fill in the causal connections. They have, however,
been given enough information to know what is unique about the situation, what is not simply part of the world generally.

Most often, the general principles in a work of fiction are those of the real world. As Marie-Laure Ryan has discussed, readers follow a “Principle of Minimal Departure” in their assumptions about the relations between the real and fictional worlds. However, there are exceptions to this continuity of principles. One may distinguish three types or levels of general principle. The first is most easily identified as the level of the real world, but that may not be quite right. It is more accurate to say that it is the level of the real world as understood by the implied author. For example, some works may rely on a belief in the miraculous intervention of saints. A reader need not believe that saints really do cure illnesses from beyond the grave in order to recognize that this is a principle presupposed in some narratives (e.g., in the Hindi blockbuster *Amar Akbar Anthony*, where the heroes’ mother is cured of blindness after—and presumably because—her son prays at a Sufi shrine).

The second level comprises general principles derived from some category in which the work is located. The obvious case of this is a genre category. In science fiction, for example, one commonly assumes that there are civilizations roughly like one’s own on other planets. In this context, “roughly like one’s own” means that real-world psychological and sociological principles apply unless we are given other, differentiating information. Categories here also include groupings by culture or discourse mode. For example, when reading Sanskrit literature, one may reasonably expect the appearance of cakravāka birds to anticipate the separation of lovers. This is presumably not a belief that ancient Indians had about the real world. But it is part of the shared literary world of such works.

Finally, there is the level of general principles that are unique to the story at hand. Like particulars, these must be specified in adequate detail. That specification may be explicit or implicit. For example, in Salman Rushdie’s novel *Midnight’s Children*, people do not generally have supernatural powers. However, each of the children born at the hour of India’s independence has some special capacity. This is a straightforward general principle that guides the (implied) reader’s understanding of the story.

I speak of these as “levels” because they form a hierarchy. In general, information given in narrower contexts displaces information that applies more broadly. Thus the (conventional) principles of a genre displace any contradictory principles from the real world. Similarly, any
principles particular to a story displace principles from a genre or other category. Put differently, in cases of conflict, the more narrowly confined principle is the one that applies.

In addition to principles, particulars enter at all these levels as well. It may seem initially that information from the real world enters only at the level of principles, not at the level of particulars. However, on even brief reflection, it becomes clear that this is not true. Indeed, authors and readers assume extensive particular information in constructing and reconstructing a story. This particular information ranges from the existence of commercial products (e.g., a certain type of chewing gum) to world-changing historical events (e.g., Indian independence).

Needless to say, this does not mean that all potentially relevant particulars of the real world are carried over into the fictional world. Just as the author may change general principles, he or she may change particulars. Here, however, a complication arises. Particulars in the real world are, one might say, “referential.” There are real people, real actions, real experiences there. Some particulars include other particulars—thus the Second World War includes particular battles, which include particular troop movements, which include particular individual soldiers, and so on. For any given level of referential particularity, there are certain things that are readily knowable and others that are not. Moreover, there are certain things that are widely known and other things that are not.

For instance, virtually every Russian reader of *War and Peace* would have known the outcome of the battle of Borodino. In contrast, the precise disposition of troops is presumably known by experts, but not by the average reader. Finally, one can assume that no one does or can know the identities or actions of all individual soldiers—even of all individual soldiers who engaged in particularly significant actions.

Such differences regarding historical knowledge have a bearing on the facts of a fictional world and on the reader’s response to that fictional world. In a fictional account of the battle of Borodino, an author may create wholly fictional soldiers as he or she likes. At the other end of the spectrum, when there is common knowledge, the reader or viewer automatically assumes that the particulars hold. When watching a film about World War II, treating some event in May 1944, one automatically assumes that the end of the war is one year away, that the end is a triumph for the Allies, and so on. Readers or viewers connect relevant story events with those facts from the real world, even though the events are still an indeterminate future within the story itself. Of course, an author
may create a fiction in which the war had a different outcome, as in Philip K. Dick’s novel *The Man in the High Castle* where the fascists have won the Second World War. In this case, readers implicitly (or explicitly) recognize that this is a case where fictional information excludes real information. Intermediate cases, however, are less clear. If Tolstoy presents Napoleon as giving certain sorts of order, as having a particular knowledge or lack of knowledge about the battle, one may take this to be a partial fictionalization that further specifies known facts, a complete fictionalization of unknown facts, a counterfactual fiction (i.e., a fictionalization that alters known facts)—or perhaps unreliable narration.

In sum, the facts of a storyworld result from the interaction of quite extensive extratextual information with some minimal textual information. The minimal textual information serves primarily to signal where the storyworld deviates from common knowledge about the real world. That deviation is largely a matter of particulars, but it may include more general principles. There are complications and ambiguities with respect to real-world particulars and principles that are knowable, but not common knowledge. Technically, the role of such particulars and principles for the implied reader may be unclear. The crucial point in each case, however, is that the real and fictional worlds are always tightly interconnected, with the default presumption being one of continuity between the storyworld and the real world rather than discontinuity. This presumption of continuity is often of crucial importance in one’s distrust of a narrator.

**Distrusting the Narrator**

Again, the default attitude toward the narrator appears to be one of trust. Thus one needs some reason to doubt the narrator. Indeed, empirical research suggests such a strong tendency toward trust that, Bor-tolussi and Dixon explain, “the precise nature of the features in a text that lead the reader to interpret the narrator as unreliable” constitutes a “puzzle” (84). Technically, automatic processing has to be interrupted by some sort of contradiction—or, more precisely, a felt contradiction, something one experiences emotionally as a contradiction, something that has a motivational component. That requirement is not as restrictive as it may seem. As already noted, readers or viewers generally have an interest in understanding the storyworld, and if faced with mutually exclusive descriptions, they are likely to want to reconcile them.
course, one needs to have encoded those descriptions and brought them into connection with one another in order to recognize the contradiction. That is not insignificant. But readers and viewers do this readily with aspects of a work that are salient.

The point is clearer if one recalls that an emotional contradiction need not be a logical contradiction. Indeed, it seems likely that relatively few literary quandaries are logical contradictions. The contradictions that inspire effortful, working-memory-based processing—thus self-conscious interpretive work—are fundamentally task contradictions. Readers (including the author as reader) need to imagine some object, explain some event, understand some emotion through imagination. Ordinarily, they accomplish this task automatically. However, sometimes something goes wrong. In Heidegger’s terms, something is “broken” and the text, rather than being “ready to hand” for story construction, suddenly faces the reader as something “present at hand.” In Iser’s term, there is some gap. One tries to fill that gap through self-conscious interpretation.\(^5\)

The judgment of a narrator as unreliable begins with this sort of gap or task contradiction. Clearly, a gap does not imply that the narrator is necessarily unreliable. It may simply mean that one has to draw complex inferences, that one needs to learn something about history or culture that is carried over from the real world to the text, that one needs to familiarize oneself with genre conventions, and so on. (The relation to reality, culture, and genre are commonly recognized in discussions of unreliable narration; see Nünning “Reconceptualizing” 98–99 and Yacobi 110–11.) Moreover, the absence of any sense of contradiction does not mean that the narrator is reliable. For example, I may fail to see that a narrator is unreliable in reporting cultural or historical events from the real world because I myself lack relevant real-world knowledge. Rather, as the preceding example suggests, without a sense that something has inhibited automatic processing, one will never begin to distrust a narrator and thus to engage in effortful inference.

There are different sorts of task contradiction or gap that may arise in connection with constructing a story from discourse. The most obvious is some sort of apparent incompatibility. Again, this need not be a strict logical contradiction. It is more often a matter of unlikelihood or lack of explanation. For example, a reader may learn that a character is opposed to war and then learn that he has joined the army. This is not a logical contradiction. But it is unlikely and requires some particular explanation. As such, it will probably to give rise to effortful thought—to the asking of questions, as Noël Carroll stresses (see “Narrative”).
It is important to note here that gaps almost invariably involve real-world principles and/or particulars. Thus they are bound up with the continuity between the storyworld and reality. Some gaps may seem to arise wholly internally to a story. For example, it may seem that an antiwar character joining the army is an anomaly purely internal to the storyworld where it occurs. But, in fact, one cannot even recognize that armies are involved in war unless one incorporates real-world information into the storyworld.

It is also important not to understand gaps too narrowly. The examples just given come fairly close to contradictions. However, gaps may arise due to unexplained coincidences or, more generally, problems of probability. They may arise because, for certain events or characters, reference is vague or identity is unclear (e.g., if the shooting of a scene makes it unclear whether certain actions were performed by one or two characters). In each case, readers or viewers are motivated to understand something further about the story and they find that there is some difficulty in achieving that understanding.

Having identified a problem, one needs to first isolate its source. In technical, narratological terms, the first question concerns whether the source of the gap is the narrator, the implied author, or the real author. One may decide that the gap was produced by the real author, but has gone unnoticed by the implied author. If so, then one is likely to see the gap as a flaw in the narrative, perhaps an important one, perhaps not. Alternatively, one may interpret the gap as occurring in the implied author. In other words, one may view the implied author as, in effect, choosing to make something anomalous. In this case, there is no fact that the implied author is concealing. The anomaly is part of the profile of ambiguity. However, since the implied author is always free to resolve incompatibilities, one must infer some motivation for this gap. It is typically thematic, though it may also be emotional.6

Finally, and most importantly for present purposes, one may attribute the gap to the narrator, judging that he or she is unreliable. (In fact, readers or viewers almost certainly try to resolve any anomaly at the level of the narrator before turning to the implied author, and they almost certainly turn to the author only after failing at the level of the implied author.) Again, once one determines that the narrator is unreliable, one undertakes further interpretation. One determines whether the narrator is lacking in knowledge or in good will or is in some way biased. One tries to infer what the fictional facts are in opposition to what the narrator has (unreliably) reported. Perhaps most importantly,
one tries to determine the limits of the unreliability, because one generally continues to accept most of what the narrator recounts. The point holds whether the unreliability is motivational or epistemic, though of course one determines the limits of unreliability somewhat differently in the two cases.

**How Does One Infer the Fictional Facts?**

This leads to the question of how readers or viewers resolve gaps, how they determine just what the fictional facts are, once they have run up against a problem. Here too it is valuable to distinguish two levels of adjudicative processing. One is more automatic; the other is more effortful, but also more general.

The more general and effortful approach is the scientific one. Here, one relies on the usual principles of scientific inference to reach a conclusion. There is nothing magical here; there are no simple solutions. There is, rather, a synthesis of empirical evidence, the generation of hypotheses, evaluation by criteria of logic and simplicity, revision of hypotheses, and so forth. This is more likely to be the province of the literary critic than the ordinary reader.

Before undertaking such a highly effortful process, readers almost certainly begin by relying on some simplified, heuristic procedures. In the case of fictional narratives, these heuristics often involve what might be called *default preference hierarchies*. Drawing on Holland, Holyoak, Nisbett, and Thagard’s account of default hierarchies, one may understand these as pragmatic orderings, such that there is a default preference, but alternatives that displace the default when particular circumstances arise. More exactly, readers appear to have default *sources* of information and default *types* of information, with specified alternatives for certain “marked” conditions. The narrator is obviously the primary default for the source of information. The primary trigger for condition marking, thus for a shift from the default, is narrator unreliability, itself made possible by distrust.

The point is in keeping with the general operation of the human mind. The default for a bird is something like a robin. In one’s backyard in the American Midwest, the sort of bird one expects is a robin or something similar. However, in the context of the Florida seacoast, one is more likely to expect a gull (cf. Kahneman and Miller 140). One would expect similar processes—of shifting from a default—elsewhere. Indeed,
with respect to trust or judgments of reliability, people engage in pro-
cesses of this sort all the time. I might generally trust Jones’s character
judgments, but (shifting out of the default) not with respect to his ex-
wife (an instance of emotional bias). I may trust police officers with giv-
ing directions or dealing with theft, but (shifting out of the default) not
with respect to an antiwar demonstration.

Consider a simple literary example. Suppose there is a contradiction
between the statements of the narrator and the statements of a charac-
ter. It seems that, other things being equal, readers are likely to prefer
the claim of the narrator, who of course has access to both views (since
he or she is presumably the one who has reported the character’s state-
ment). However, if one judges the narrator to be unreliable due to a lack
of knowledge, one may prefer the character’s statement, if it falls into
the right epistemic area—as when the narrator is a child who does not
understand the topic under consideration (e.g., sex, finances, transcen-
dental Phenomenological reduction). If the narrator is motivationally
unreliable, then readers are likely to trust the version that is least in
keeping with the speaker’s biasing motivations or self-interests.

Of course, these default-replacing alternatives may be overridden
also. They may generate their own gaps. Those gaps may, in turn, be
resolvable by further alternatives in one or another preference hierar-
chy, or they may require more effortful hermeneutic work.

Again, this shifting of a default preference hierarchy is intermedi-
ate between spontaneous processing and effortful interpretation. In this
way, there is a difference from the prototypes governing what sort of
bird one expects in what location. Those prototypes operate automati-
cally, with no prior sense of contradiction or gap. In the case of a narra-
tive, one must first come to challenge a source of information or sense a
conflict in types of information before one shifts from a default. The shift
involves some disruption of normal processing, even if it is not a matter
of complex, extended reflection. In this way, the situation with unreli-
ability in narrative information is parallel to first learning of some real
person’s unreliability, insofar as there is a relevant standard/prototypi-
cal alternative for trust in that case. I know that Jones is generally a fair
judge of character. However, I come to realize that he is unreliable with
respect to his ex-wife. This constraint is easily assimilated to prototypi-
cal cases of embittered ex-souses. I do not have to spend a great deal of
hermeneutic effort on the case, though it also does not occur spontane-
ously. In other cases, I may have to engage in more complex and difficult
inferences. This appears to be what happens with shifts from defaults in narratives as well.

Drawing on, but altering, hermeneutic terminology, one may refer to the three types of narrative processing in the following terms. First, the great majority of one’s reading is necessarily as simple and straightforward as one’s competence in the language allows. This may be called “spontaneous understanding.” It is the basis of the other hermeneutic processes, in part because it is necessary to the experience of gaps. For example, readers rely on spontaneous understanding to receive story information that allows for a sense of incompatibility. In this respect, one always has to rely on the narrator to some degree. Once one senses a gap and experiences distrust, however, one shifts from the default through what might be called “interpretive adjustment.” Relative to spontaneous understanding, this is an effortful process. However, it is effortful in a limited way, for it operates through a heuristic shift to a preset alternative that is itself largely unquestioned. In contrast, some gaps or some implications of gaps do not have, so to speak, “standard” solutions, particular and ordered alternatives to the default. They require a more diverse, vigorous, and creative exploration of possible solutions. This might be called “critical explication.”

As the terms indicate, the first process is undertaken by any reader who makes any sense out of a literary work. The second process is more variable and in many cases requires a reader who has some degree of skill and experience. This does not mean that the reader has to be an academic. Insofar as the default hierarchies rely on genre knowledge, a grammar school youth may be highly competent at interpretive adjustment with respect to, say, the mininarratives of rap songs, and a professor of literary theory may be almost entirely incompetent in that area. Finally, critical explication is the sort of sustained analysis that more typically characterizes criticism than the isolated reading process. Here, too, criticism is not confined to articles published in academic journals. It includes blogs, chat-room discussions, or even conversations among friends, though articles and books do tend to offer the largest scope for expansion and elaboration of critical explication.

Since this is a bit abstract, it is worth giving a brief example before continuing. As will be clear from the following discussion, Surfacing faces the reader with several anomalies. The central gaps concern the memories of the narrator—did she marry and have a child, or did she have an affair and an abortion? I have encountered students who fail to rec-
ognize the contradiction here and assume that both happened. They have obviously processed the novel at the level of spontaneous understanding, since they recognize the sequences of events. However, they have not noted the contradiction. Some know there is a contradiction, but are not quite sure of the solution. Most, however, sense the gap and the solution. I take it that this is because they recognize the narrator as motivationally and epistemically unreliable, with the particular sort of combined inhibition that results from (psychoanalytic) repression. If her unreliability were solely motivational or solely epistemic, it would be far more difficult to understand. (How could one simply not know that one had an abortion? Alternatively, why would one make up a story about abandoning one’s child?) It would require effortful critical explanation. In contrast, if the unreliability is a matter of repression then all one needs is some basic psychoanalytic heuristics. One heuristic may be that people are more likely to repress something sexual than nonsexual; another may add that people are more likely to repress something traumatic than nontraumatic. Both heuristics favor the affair-abortion as the repressed content. (There are other heuristic reasons for this choice as well, reasons that do not rely on any knowledge of psychoanalysis. Some of these will be taken up in the following section.)

Of course, even with the psychoanalytic heuristics, critical explication enters. For example, the reader still needs to understand the story told by the narrator as a substitute for the memory of the abortion. A critical explication may explore some of the ways in which the imagery of the novel is consistent with the psychoanalytic heuristics. More significantly, a critical explication may go on to consider a further difficulty. Understanding the story relies on a sort of psychoanalytic schema. But there is a problem with this, because the operation of repression here does not quite fit psychoanalytic principles. Nor is it entirely in keeping with the general understanding of traumatic memory loss, which typically involves an experience that was unexpected, terribly painful, and incomprehensible (for example, Christianson and Lindholm cite “rape and extreme and prolonged acts of violence” [772–73]). This may lead back to the implied author or the real author and the question of whether this discrepancy has an emotional or thematic function or is simply an error.

Before going on to explore Surfacing in greater detail, however, it is important to get a more concrete idea of what the usual preference hierarchies are. As already noted, such hierarchies may bear on sources of information or on types of information. In addition, gaps may arise due
to different sorts of relation between internal and extratextual information. This may be termed the *epistemic context* of the gap. Thus we may distinguish three varieties of hierarchy—bearing on source, type, and context—which might be briefly considered in turn.

Regarding information source, as already discussed, readers seem to begin with a default, tacit assumption that the narrator is reliable. But there is a complication here, since there are different sorts of narrator. Most importantly, one may distinguish between nonpersonified and personified narrators. This is probably the most basic division that bears on narrator reliability. Specifically, in cases of conflict, we assume that a nonpersonified narrator is to be preferred over any personified figure, either a narrator or a character.

This hierarchy fits with the common view that “If anonymous,” the narrative voice “is equated with the voice of truth.” In contrast, “In first-person narrative . . . the teller is individuated, and his or her vision considered subjective. Such a teller normally has access to their own mind only, and the completeness or reliability of any information or judgment they provide can be questioned” (Margolin “Person” 423; see also Margolin “Authentication” 33, citing Doležel). This appears to overstate the case, however. Nonpersonified narration is not absolutely unquestionable, even in comparison with personified narration (e.g., when one is embedded in the other). At the very least, as Nünning explains, “there is an ongoing debate about whether unreliability is a property of first-person narrators only, or whether it can also be attributed to third-person narrators” (“Reliability” 496). In any case, it seems clear that readers are far more likely to distrust a person when faced with some incompatibility. It is both emotionally unlikely and cognitively difficult to question a nonpersonified narrator. When a nonpersonified narrator delivers a particularly secure type of information (e.g., visual perception), then it is almost impossible—almost, because the ending of *New York* indicates that unreliability can arise even in this context.

Among personified figures, one tends to prefer some types to others. It seems likely that, other things being equal, one will first follow the communicative discourse hierarchy of embedding. Thus, after nonpersonified narrators, one will tend to trust the information provided by a personified narrator or a focalizer, then a character who is a topic or protagonist. For example, when, in his role as narrator, Nick directly or indirectly contradicts a statement from Gatsby, we are, it seems, more likely to trust Nick than Gatsby. A rather subtle case of this occurs when Gatsby says, “I suppose Daisy’ll call too.” The comment expresses
a belief that Daisy will call. Nick comments, “He looked at me anxiously, as if he hoped I’d corroborate this” (218), suggesting that the comment is more a desire than an expectation. I suspect that most readers automatically process the passage in such a way as to accept Nick’s implication. Note that this involves accepting Nick’s view at two levels. First, a reader accepts that Gatsby looks anxious. Second, he or she accepts that Gatsby looks anxious because he doubts whether Daisy will call (he could be anxious about many other things, given the nature of the preceding conversation, and given recent events).

This may be a specifically discourse-related hierarchy, or it may be part of a general heuristic hierarchy of sources whereby one prefers information from those one knows well over information from those one knows less well. The latter is most obvious with two characters who operate as embedded narrators. For example, in book 10, chapter 35 of *War and Peace*, Kutuzov and Wolzogen, the second reporting from Barclay de Tolly, both summarize the Battle of Borodino. Their accounts are radically incompatible, with Kutuzov evaluating the battle positively and Wolzogen/de Tolly evaluating it negatively. The reader is strongly motivated to resolve this incompatibility. Much greater familiarity with Kutuzov seems to be a factor biasing a reader in favor of his account.

This preferential trust in familiar personified figures is precisely what one would expect from research on people’s response to familiar and unfamiliar people in real life. This same research suggests that people respond with fear or vigilance to people from out-groups (see Oatley *Emotions* 73; see also Ito and colleagues). Thus one would generally expect an in-group/out-group hierarchy to occur in fiction as well. In the preceding example, the association of Wolzogen and de Tolly with German language and ethnicity could readily contribute to at least a Russian reader’s preference in this case. Indeed, this identity-group division is highlighted in this chapter (for Wolzogen) and elsewhere (book 9, chapter 9, and book 10, chapter 25) for de Tolly.

Of course, there is also a potential motivational issue in this case. Even when readers do not have evidence that a particular narrator is unreliable, they have a general preference for information that is contrary to the self-interest of the speaker. In the case of Kutuzov and Wolzogen/de Tolly, this partially favors the latter. Kutuzov, in command of the Russian troops, is motivated to view the results of the battle positively. On the other hand, de Tolly is a rival of Kutuzov, which somewhat taints his version as well. Moreover, Kutuzov does not seem to be interested in any material rewards for success. Rather, his interested-
ness is a function of his sense of loyalties. In this way, his version is self-interested without being morally faulty. That does not mean it is fully trustworthy. However, it may mean that readers will be less critical of it on motivational grounds.

To some extent, these heuristic preferences share a single emotional basis. Readers have an emotional response of trusting some narrators and not others, some characters (who may contradict narrators) and not others. Spontaneous feelings of trust or distrust arise easily in relation to familiarity/unfamiliarity and in-group/out-group categorization. Decisions about a narrator’s self-interest may require somewhat more effortful, self-conscious processing, but they are often relatively simple and spontaneous. Moreover, they certainly bear on one’s feelings of trust or distrust.

It is possible to summarize the usual trust hierarchy as a preference hierarchy of the following sort. (A superior item in a preference hierarchy is chosen when all other factors are held constant. Thus the preference rule $a > b$ means that, if everything else is equal, choose $a$.)

$$
\text{nonpersonified narrator} > \text{personified narrator} > \text{focalizer} > \text{protagonist} > \text{other characters}
$$

If personified, then

$$
\text{familiar} > \text{unfamiliar} \\
\text{in-group} > \text{out-group} \\
\text{disinterested} > \text{self-interested}
$$

As this suggests, there do not seem to be general conditions in which a nonpersonified narrator is distrusted. This can only result from an accumulation of evidence, making it extremely rare.

The second area in which one finds a default preference hierarchy concerns types of evidence. This is in some ways more straightforward than the hierarchy of sources of information. It has a fairly clear relation to everyday experience, and is also closely related to the grammatical operation of evidentials. (Evidentials are grammatical markers indicating, for example, whether one witnessed or was told about the information one is conveying [see Aikhenvald].) Generally, it seems clear that people trust current perception more than they trust memory. Within current perception, people tend to trust vision over hearing (the other senses enter less often). Within memory, people trust current mem-
ory more than distant memory. People seem to generally trust memory more than empirical inference. Within empirical inference, people trust empirical inference by experts more than that by nonexperts. For example, the narrator of a crime investigation might be skeptical of the scientific contraptions used in the lab. But, other things being equal, readers are likely to give initial credence to the findings of the lab. Of course, if the narrator is a seasoned detective, the whole point of the story may be to show that the instincts of the seasoned detective are more accurate than those of technicians. But then readers have two types of expertise and this preference hierarchy does not come into play. Finally, below the level of empirical inference, there is something like hearsay—roughly, ideas for which there is no evidence other than someone’s assertion.

The hierarchy in types of evidence may be summarized in the following diagram:

perception [visual > auditory] > memory [recent > distant] > empirical inference [by experts > by nonexperts] > hearsay

Of course, like all default hierarchies, this can be overridden in particular circumstances.

The final default preference hierarchy concerns epistemic context. Given certain sorts of contradiction, people give preference to one type of consistency over another. Most fundamentally, in life and in literature, people prefer to preserve logical coherence over any other sort of coherence. This is simply a presupposition of all adjudication of gaps. If people did not seek to preserve the principle of noncontradiction, they would not even recognize gaps. For example, there would be no tension between Jones opposing fascism and joining the Nazi party. Anyone might empirically expect that someone who opposed fascism would not join a fascist party. But, without the principle of noncontradiction, there is no reason to balk even at the statement “He signed a party card and he attended an induction ceremony and he never signed a party card and he never attended an induction ceremony”; there is no reason to determine which alternative is correct or whether they somehow have different meanings.

In real life, below this founding assumption about logic, there are the laws—or at least generalizations—of various sciences or, more gen-
erally, modes of knowledge. These include not only the laws of physics, but general principles of psychology, and to some extent sociological or other patterns. For some authors and readers, they may also include supernatural principles, such as divine providence. These too tend to be hierarchized, with physics given preference over biology, which is in turn given preference over the psychological and social sciences, though this subhierarchy rarely has consequences in literature. The relation between natural and supernatural principles is more vexed. Even readers who reject supernatural causality in life must accept that it does often occur in literature. I suspect, however, that when natural and supernatural explanations are both available, authors and readers generally prefer the natural explanations. I suspect that this is the case even for authors and readers who accept supernatural causality in real life.

In literature, another epistemic context intervenes between logic and laws of nature particularly. These are category-specific conventions, such as principles of genre. In cases where the conventions of a genre come into conflict with general laws of physics (drawn from the real world) readers give preference to the former. Indeed, this is usually so automatic that no one is likely to notice. In that sense, it does not give rise to a gap. Particular genre conventions may also complicate the relation between natural and supernatural causality. They can certainly make the invocation of supernatural causality more (or less) likely than usual.

The final epistemic context is that of empirical particulars—not general laws, but specifics of history, culture, biography, and so forth. Of the various epistemic contexts, readers seem most willing to forgo these when reading fiction. On the other hand, these are also likely to prove the most troublesome because this is the level at which most fictional deviation from reality occurs. As already noted, at this level, it is sometimes difficult to tell whether a contradiction with facts is meant to indicate unreliability of the narrator, a fictionalization by the implied author, or an aesthetic or ideological problem deriving from the real author.

The default preference hierarchy for epistemic contexts may be summarized with the following diagram:

logic > principles of genre or other category > natural laws > supernatural principles > particulars of society, history, etc.
Margaret Atwood’s *Surfacing*, published in 1972, is a prime example of a novel with an unreliable narrator. But, at the same time, it is a novel where it seems fairly clear what the actual story really is. In other words, the narration gives enough information to infer the correct story. Indeed, the narrator herself gives this information as she corrects the story and, in effect, becomes reliable. In this respect, the novel is nicely designed to illustrate the various ways in which readers may judge and limit unreliability, and the ways readers may adjudicate incompatibilities or resolve problems arising through gaps. Put simply, the narrator changes her mind about things. This allows one to consider the question—why do readers (or critics) trust one version and not the other?

It seems likely that most first-time readers of *Surfacing* are unaware that the narrator is unreliable almost until the moment where she herself more or less announces that she has gotten the story wrong. This is in part because the contradictions are largely “unnarrated” (in Prince’s sense) until the point where the narrator begins to “denarrate.” (Denarration is “the narrator’s denial or negation of an event or state of affairs that had earlier been affirmed” [Prince “Disnarrated” 118; see also chapter 3 of Prince *Narrative*].) Moreover, it seems likely that, even then, most readers engage in relatively quick interpretive adjustment, rather than critical explication, more or less automatically assuming that the new statements by the author are reliable—even as they make further interpretive adjustments to allow for her increasing peculiarity of behavior and apparent hallucinations. The following analysis, however, will follow Roland Barthes’s “last freedom” of the critic, “that of reading the text as if it had already been read” (15). Thus it will be concerned with how someone who has previously read the novel could critically explicate both the narrator’s limited unreliability and the implied reader’s adjustment to this.

There are two basic areas of the narrator’s unreliability. The following discussion will explore only one of these. The first area of unreliability, outlined briefly above, concerns her memories of her relationship with a lover and their child. Initially she recounts that she and this man were married; they had a child; and the narrator left both the husband and the son. Subsequently, the reader learns that she did not get married. Rather, she had an affair with this man, who was married to someone else. She did become pregnant, but had an abortion.
As she recovers these memories, it becomes clear that she had repressed them and substituted a fantasy, if an ultimately rather unpleasant one. The idea that there is a repressed memory works well with various aspects of the book. For example, in psychoanalytic theory, repressed contents will repeatedly manifest themselves in symptoms, symptomatic acts, ways of speaking, and so on. The repressed ideas or impulses will partially orient one’s trains of thought, behavior, and word choice. Thus one’s speech and action will often contain hints of those ideas and impulses of which one remains unaware. This is found throughout the narrator’s false narration of her marriage.

On the other hand, as already noted, there are important ways in which the narrator’s case does not fit psychoanalytic theory. Most importantly, the repressed memories are not infantile. Of course, there are cases of repressed traumas in later life. But these are usually much more limited. Thus it is not impossible that she would suppress some aspects of the abortion—for example, aspects that were particularly physically painful. This might be parallel to a soldier losing memory of certain parts of a battle. But the narrator has repressed the memory of the entire abortion, and the entire pregnancy, and the lover being married. It would require quite severe, pervasive, and probably debilitating psychological problems for someone to have repressed so much of her recent life (cf., for example, the rape victim discussed by Christianson and Lindholm [773]). But, for most of the novel, the narrator exhibits no signs of such severe, pervasive, and debilitating problems. This produces a gap requiring critical explication and suggesting some thematic points of the novel.

Before going on to this, however, it is important to briefly mention the second area of unreliability. This concerns supernatural experiences. At the end of the novel, the narrator sees the spirits of her dead father and mother. This could be understood as the result of a severe psychological disorder, manifest also in the distorted memories. However, it has a simpler explanation. She has taken hallucinogenic mushrooms. She clearly identifies the mushroom in conversation (see 177). The information is presumably included by the implied author to provide an understanding of the narrator’s hallucinations. On the other hand, this too is not entirely simple. The mushroom in question was perhaps involved in Ojibway spiritual practices (see Navet). This is particularly germane as the narrator is clearly imitating Native American/First Nations’s practices at this point. This too involves a sort of interpretive gap, also with thematic consequences. For example, it bears on whether one reads the
end of the novel as hopeful or hopeless (a point of conflict in the criticism\textsuperscript{1}). These are obviously important issues for one’s understanding of and response to the novel. However, in order to keep the discussion within manageable limits, the following analysis will focus on the narrator’s unreliable memories.

Though impossible to tell before reading the novel, the title already suggests the submerged nature of the narrator’s memories. It also points, quite literally, to the ways in which they will reappear eventually—when she is below the water. Even before one reads the novel, one knows that “surfacing” suggests something or someone temporarily concealed, then reappearing.

Already, on the first page or two, there are several points at which one could question the narrator’s reliability, if one did not have an initial, default trust in her. The narrator and three friends—including her current boyfriend, Joe—are driving to northern Canada, where she grew up. Her father has disappeared and may be dead. They pass “the city.” First, she explains that they have not gone through; then she describes its “one main street with a movie theater.” Relying on memory, she explains that the movie theater was “the itz, the oyal, red R burned out” (3). The theater was presumably either the Ritz or the Royal, not both. She presents this as if it were a trustworthy memory. But the duplication of the theater’s name indicates that it is not entirely trustworthy. She goes on to recount a memory from “before I was born” (3), a memory that also seems dubious. Her brother “got under the table and slid his hands up and down the waitress’s legs while she was bringing the food” (3). This simply does not fit with the verb “bringing” (with its suggestion of walking from the kitchen to the table). But even if “bringing” means “serving,” it hardly seems likely that the waitress would simply accept this behavior. Then there is a third memory of her and her brother having no shoes in winter (4). The image of them running barefoot through the snow once or twice is plausible enough. But how likely is it that her parents would leave them entirely without shoes in a north Canada winter?

The points are all small and apparently insignificant. However, they suggest that the narrator may not be trustworthy even at the outset. Though a reader may not self-consciously notice any of these small uncertainties, they may provide some preparation for the larger gaps in memory that follow. In other words, few readers will reflectively consider any of these points as genuine contradictions or errors. However, there is something strange about them. Readers may begin to sense at
least a “tall tale” element to the narrator’s memories, which may unre-
fectively prepare them for what follows. As a result, her eventual rev-
elation about her memories of marriage and divorce may be surprising,
but not incomprehensibly anomalous.

The end of the chapter introduces a peculiar self-consciousness about
the narration. Given that there is no narrational frame and no overt nar-
ratee, it is strange to find the narrator reflecting on her own practices as
a narrator. Specifically, at one point, she refers to something “they” used
to do (11). She then corrects herself, saying, “That won’t work, I can’t
call them ‘they’ as if they were somebody else’s family” (12). Again, at
least some readers are likely to respond to this with an implicit question
about just whom she is addressing and when. It is as if there is narra-
tion to an implicit narratee that is interrupted by a shift to the inter-
nal consciousness of the narrator. However, without the involvement of
some at least moderately motivated task, a reader is unlikely to follow
through on such a momentary question, particularly in the context of
default trust. The plot thickens when she goes on to suggest that she
has an impulse to misrepresent something. The next sentence is “I have
to keep myself from telling that story” (12). At the same time, this hint
of untrustworthy motives is mitigated by the narrator’s stated commit-
ment to not being unreliable. Thus, here as elsewhere, the hint might
serve to prepare the reader for the narrator’s unreliability without yet
directly provoking distrust.

In the second chapter, the reader supposedly learns about the narra-
tor’s recent past—her marriage and divorce. She recounts that she sent
her parents a postcard about the wedding. While certainly not impos-
sible, this is clearly an unusual situation (by real-world standards) and
requires clarification. A neighbor of her father’s, Paul, seems to know
about the marriage, and the narrator explains this by reference to the
postcard. This apparently confirms the story (i.e., makes it less anom-
alous, less likely to appear doubtful). However, she also notes that she
is wearing a wedding ring. In retrospect, one may understand that Paul
simply inferred the marriage from the ring.

Just after this, the narrator has a sudden memory of the “husband.”
Its location and time are unclear. She figures that “it must have been
before the child” (22). This uncertainty highlights the fallibility of mem-
ory (a point stressed in recent memory research; see Schacter). Though
not specific to the narrator, such fallibility clearly bears on the degree
to which we can trust her recollections. Merely noting this fallibility is
unlikely to provoke questions about her reliability. But it probably does
help to prepare the reader for what occurs later. The general point is recapitulated—and the preparation enhanced—subsequently, when the narrator thinks about Joe. She makes the strange statement “though I can’t reconstruct our first meeting, now I can” (28). The sentence simultaneously reasserts the uncertainty of memory and the subjective conviction that any given reconstruction is correct. It fits with the narrator’s apparent confidence in her memories and what we will discover to be their frequent falsity.

Readers are soon given more information about the marriage. She explains that her parents did not understand her marriage and that she herself “didn’t understand it.” She goes on to state that her son “wasn’t really mine” (29). Most of the earlier incompatibilities are probably ignored by first-time readers. But these last points may require some minimal interpretive effort. One can make some sense out of the statement that she did not understand her own marriage by assuming that the marriage was undertaken on the basis of a misconception. But that is not entirely adequate, since the misconception would presumably be about the man or about the reality of married life, not why she married this man at this time. The statement about the child is tougher. Readers may assume that the child has been made into the image of its father. But that would seem to suggest a child who is somewhat older than indicated in the rest of the text. (Later, the narrator says that she got married about nine years earlier. Could the child already be recognizably identical with his father at the age of, say, six or seven, when the narrator would have been contemplating the divorce?)

The chapter ends with a memory of her brother drowning. She explains that “It was before I was born but I can remember it as clearly as if I saw it” (32). People do in effect fabricate memories from things they are told. Her assertion, in that sense, is not precisely untrustworthy, since she indicates that the memory is false. But this does establish that her memories are at least sometimes untrustworthy—and it is far from clear that she will always recognize this fact. Indeed, at this point, a careful reader may be starting to have a fairly strong, if probably still un-self-conscious, sense that the narrator’s reported memories are far from reliable. Or, rather, some readers may be starting to feel their trust in the narrator disturbed, even if they could not say precisely why.

Not by coincidence, right after the clause about her memory of her brother drowning, the narrator introduces the idea of “an unborn baby” with “its eyes open” looking out from “the mother’s stomach, like a frog
in a jar” (32). This is a crucial image because it prepares readers for the “surfacing” of her actual memories of an abortion. Particularly in the psychoanalytic context implied by the text, these become, in retrospect, small, unconscious expressions of the repressed memory. Of course, this sort of rhetorical patterning can occur even outside a psychoanalytic context. Specifically, of all the possible expressions open to a narrator, the implied author may choose just those words and images that serve to prepare the reader for something that is concealed, thus making it more plausible later on.

The beginning of the next chapter offers an explanation of the child not being hers. First, she points out that she “didn’t name it before it was born” (34). That is plausible in itself. But, at the same time, it may draw attention to the fact that one never learns any name for the child. Clearly they must have named the child at some point, so why is the reader never told the name? Moreover, if it was a boy, why does she continue to say “it”? This seems very out of keeping with ordinary biological responses to one’s offspring.

The rest of the passage presents what was at the time something of a cliché about male domination. The husband uses the wife as a baby machine. (Needless to say, something can be a real and important problem and nonetheless be turned into a cliché.) She particularly complains that “he imposed it on me” (34). It is true that he did impose something on her. But what he imposed on her was the abortion, thus not having the child. The point is significant thematically. Perhaps particularly in 1972, the women’s movement was deeply concerned with men trying to restrict women’s access to abortion. But if women’s choice is really the issue, then their freedom to have children should be similarly important.

The next chapter gives further imagistic manifestations of the repressed memory. Thinking about the end of her marriage, the narrator reflects, “A divorce is like an amputation, you survive but there’s less of you” (44). That makes sense, certainly. All the lost projects and aspirations, with their foundations in emotional attachment, are aptly analogized to a part of oneself. But it should be clear that the phrase applies at least as clearly to an abortion.

Later in the chapter, her emotional response to her husband “surpris[e]s” her. She is angry and resentful, even though “he didn’t do anything to me” (50). Clearly, her emotional response is in conflict with her explicit memory here. There does not seem to be a standard hierarchy to resolve this dilemma, so one is left with an ambiguity about which to trust—the emotion or the memory.
She goes on to note that “there aren’t any pictures” of the baby (again, referred to as “it”) and that she hasn’t mentioned him/it to Joe or her other friends (David and Anna). She then reflects on how “the baby . . . was taken away from me, exported, deported” (51). The idea is strange. Her claim up to this point is that she left the child. So it hardly makes sense to say that the child was taken away. The words “exported” and “deported” both suggest a separation across national lines as well. This “nationalization” of her separation may, in turn, serve to associate the “husband” with America. If so, this would relate to her extreme antipathy toward Americans, an antipathy that is almost pathological, even considering the historical context of the Vietnam War. The Americans and Americanism are consistently associated with violence, cruelty, senseless killing—which, in the end, is precisely how the narrator thinks of her abortion.

The following lines recur to the imagery of abortion. She characterizes her separation from her child as “my own flesh canceled” and “my own life, sliced off from me” (51). Here again one sees the narrator choosing images that may be construed as fitting her abandonment of her child. However, they are more fitting manifestations of an unconscious memory. The abortion does slice off her own life and cancel her own flesh in a literal way that abandoning her child does not.

Many of the same issues of memory recur in the following pages. At one point, for example, she thinks how safe she felt in the city. She then immediately tells herself “That’s a lie,” recalling that she was often “terrified” (82). In this particular case, she reflects on memories and how she has problems discriminating between her own memories and “the memories of other people” that she’s been told (82). The passage draws explicit attention to the fallibility of memory and thereby may foster thematic reflection. However, for present purposes, it most importantly continues to prepare readers for the revelation about her own false memories. In other words, it creates a complex of memories and emotional responses in the reader that affect his or her sense of trust in the narrator (just as the events in the experiment with the cards create memories and emotional responses that affect the test subjects’ sense of trust in the different stacks).

The narrator’s memories of the “birth” are along the same lines. The entire process is associated with being “dead.” The doctors are “butchers” who “take the baby out with a fork like a pickle out of a pickle jar” (92). By this point, it is actually becoming difficult not to recognize that this is an abortion rather than a delivery—even though first-time read-
ers probably still do not recognize this. She goes on to explain that “He wasn’t there” but “he brought his car to collect me afterward” (92–93). Here, the narrator herself is very close to recognizing that this was not a delivery. The procedure she recalls is what she understands of her abortion. Her departure, apparently immediately afterward, seems to fit with abortion more than a delivery. Finally, and most significantly, she remembers that he collected “me” after the procedure, not “us.” The reader still is probably not self-conscious about most of this at this point. For example, he or she is free to imagine “afterward” as referring to a day or so afterward. He or she can imaginatively envision the “husband” picking up the narrator with the child in her arms. In short, the reader need only make slight interpretive adjustments. But Atwood is constructing a set of associations that will make the subsequent change in memory believable and that will encourage readers to judge the abortion memory true and the birth/divorce memories false.

A crucial point comes soon after this. The narrator is supposedly remembering her “wedding.” They fill out forms, which prominently include blood type (101). The place smells of “antiseptic” (102). In retrospect, one may realize that this is not the justice of the peace. It is the abortion clinic. The following dialogue makes no sense at a wedding. The man asks if she feels “better” now that “It’s over.” He consoles her, saying, “it’s tough . . . but it’s better this way.” Her legs shake and she has an ache, presumably in the region of her legs. Crucially outside there is a “fountain” with “dolphins and a cherub with part of the face missing” (102). At least some readers will remember that this is precisely the fountain she sees on her drive, “dolphins and a cherub with part of the face missing” (9). Thus one is presented directly with a memory and a perception. It seems extremely unlikely that both are true. By the usual default hierarchy, the perception is more trustworthy.

Of course, the abortion is more often misremembered as a birth, not a marriage. The links connecting this medical procedure (birth or abortion), the title, and memory are made clear subsequently. She explains that the procedure involved anesthetic that “was like diving, sinking.” As a result, “I could remember nothing” (131).

Finally, searching in the water, she sees something, perhaps her father’s corpse—readers never know. But it triggers a series of memories. She now in effect explains the strange images that had permeated her (pseudo-)memories up to this point. She did not remember her brother drowning, contrary to her claim (see 32). Moreover, the association of the “unborn baby” with a “frog in a jar” (32) was fabricated. Rather,
she had envisioned the aborted fetus in a jar. Indeed, for a moment, she thinks that is a memory. But she corrects herself. Suggesting the image of the baby being removed with a fork, she recalls that the fetus was “scraped . . . into a bucket.” It flowed “through the sewers . . . to the sea” (168), from which it was now surfacing as a memory. She goes on to explain why “he” was not there. He had children of his own, and there was a birthday party for one of them (168). She goes on to recall that the fountain “with the dolphins and the cherub” was not there. She recurs to the image of feeling “amputated” (169, cf. 44) and to the smell of “anti-septic” (169, cf. 102).

Again, faced with this contradiction, how does one determine which story is true? At this point, the reader is no longer in the realm of small interpretive adjustments. However, some heuristics do come into play. For example, as already noted, the incompatibility between memory and current perception in the case of the fountain weighs heavily in favor of the perception (i.e., given the unlikelihood that both are correct, there is greater reason to trust the perception than the memory). More importantly, a reader’s sense of conviction in the story of the abortion results to a significant extent from the narrator’s creation of a network of associations that makes the new story plausible. At the very least, the prior narrative creates a wide range of gaps—tensions, questions, uncertainties, if not outright contradictions—and these, gradually diminishing the reader’s trust, prepare readers to find the earlier narration unreliable. Moreover, whatever one may think of psychoanalytic theory, the novel is readily categorized as relying on psychoanalytic ideas. Like principles of genre, the principles connected with this categorization take on an important adjudicative function. Thus repression and the recovery of memories become highly plausible explanatory principles here.

Finally, there is the larger, effortful, critical processing that goes beyond these heuristics. It operates here in the usual fashion, with the usual hierarchies. For example, in the adjudication of theories, people generally prefer the one that provides the more encompassing explanation. Here, readers have been given two accounts of the story—one in which the narrator was married; one in which she had an abortion. As presented, with its implication of a psychoanalytic context, the latter serves to explain the former, but not vice versa. In other words, if one assumes that the narrator’s story actually involved an abortion, one is able to explain why she would have the feelings, memories, and con-
ditions (e.g., no baby pictures and no wedding pictures [127]) reported earlier. Put differently, one can infer the nature and limits of her unreliability. One can link the epistemic limitations to her motivations, and circumscribe her motivated distortions to the sexual, traumatic, and guilt-ridden memories. Indeed, the fact that she feels guilt about the abortion (“Whatever it is, part of myself or a separate creature, I killed it” [168]) serves to explain why a sense of guilt seeps into her “screen memories,” that is, the apparent memories that both conceal and indirectly reveal a repressed memory (for a concise treatment of the concept of screen memory, see Laplanche and Pontalis 410–11). In contrast, if one assumes that the story of the marriage is true, one is left with a wide range of gaps, unexplained points of irresolution in the story.

Of course, the explanation provided by the abortion story is incomplete. It too leaves some gaps. Again, the novel tacitly invokes psychoanalytic principles of repressing memories connected with sexuality, trauma, and guilt. But, in psychoanalytic theory, it is, first of all, childhood memories that are repressed. This is clearly not a childhood memory, nor assimilated to a childhood memory. In these ways, the novel deviates from psychoanalytic accounts. Nor does it appear to be consistent with what is known about traumatic memory loss in adulthood. This could simply be a mistake on Atwood’s part. It seems more likely, however, that this deviation from psychoanalysis has thematic implications.

Specifically, there seem to be three important thematic complexes in Atwood’s novel. One is feminist. This is clear in the repeated discussions of the place of women, the sexual relations of men and women, the humiliating treatment to which women are subjected by men, and so on. A second concern is, roughly, nationalist. It is manifest in the repeated discussions of U.S. violence and excess, and the relation of Canada to the United States. The third concern is ecological. It is manifest in the attention to nature and the treatment of animals, among other things. Each of these thematic concerns is partially explicit in the novel. The narrator reflects on feminist, national, and ecological issues; characters discuss them.

The gaps created by discrepancies with psychoanalysis could contribute to any thematic concern. However, the relation of the narrator with her lover and the topic of abortion bear most obviously on feminist issues. The man appears domineering and manipulative. He uses the narrator sexually and belittles her talent as an artist. As noted earlier, the narrator represents him as something of a cliché as a husband—
particularly, as someone forcing her to have a child. This fits well with standard feminist views at the time. Due in part to political activism related to reproductive rights, many feminists rightly saw (many) men as trying to control women’s reproduction by outlawing abortion and forcing women to have children. In this standard view, however, one could reasonably argue that another issue had been repressed. A wide range of men would want women to have abortions. It is not really plausible to think that a normally functioning woman would entirely repress an adult memory of an affair, pregnancy, and abortion. But it is quite plausible to see mainstream feminism as metaphorically repressing the story of women who do not want to have an abortion, but are coerced into doing so. Indeed, it is relatively easy to explain why this would happen. Fearful of restrictions on access to abortion, many feminists may have been inclined to avoid stories that lent themselves to anti-abortion use.

This does not mean that Atwood is “anti-abortion” (or “pro-life”) in any usual sense. The novel does not seem to take a consistent stand on the status of the fetus. Sometimes, the narrator views the fetus as a baby who perceives the world in much the way an infant does (32). At other times, she insists that a fetus “wasn’t a child,” though “it could have been one” (168). Perhaps the most plausible interpretation is to see Atwood, at least in this novel, as pro-choice in the fullest sense of the term. In practice, “pro-choice” is used to mean “in favor of access to abortion.” But it should mean “in favor of women being able to choose whether to continue a pregnancy to term or not.” Thus they should not be forced to have a child when they do not want to—but they also should not be forced to abort a child when they do not want to. When faced with such a statement, I suspect that few if any feminists would disagree. But the suggestion of Atwood’s novel is that part of the story of choice has been repressed, and replaced with the single story of women being used as baby machines.

The point may be suggested when the narrator begins to question her own recollections and cautions herself. “I have to be more careful about my memories,” she says. “I have to be sure they’re my own and not the memories of other people telling me what I felt” (82). Such distortions are most obvious in patriarchal ideology. They turn up in the novel in the narrator’s notebooks as a young girl, her imagination of what she should be—“ladies . . . holding up cans of cleanser, knitting, smiling, modeling toeless high heels and nylons with dark seams” (105). But false memories can arise from political correctives to patriarchy as well.
Conclusion

A narrator is unreliable, first of all, to the extent that he or she is untrustworthy regarding the facts of the story. Those facts are defined by the receptive intent of the author. But that receptive intent is not simply a function of what the narrator says. In fact, the readers’ and implied author’s construction of the story involves the extensive integration of real-world presuppositions with deviations that are marked by the text—that is, the text serves primarily to indicate the points at which the story differs from the real world. The narrator may misrepresent facts, exclude important facts, or divert the reader’s attentional focus from some important facts that are presented. In connection with this, the narrator may be unreliable for epistemic reasons (error) or due to some motive, deceitful or emotionally biased (as in the case of out-group representation).

Readers begin with a default assumption that the narrator is trustworthy—or, more properly, a default sense of trust in the narrator. They tend to retain this sense even as incompatibilities or “gaps” accumulate. These incompatibilities need not be logical contradictions. They are, rather, primarily a matter of what readers experience emotionally as incompatible. Specifically, ordinary textual processing proceeds in a relatively routine way until one encounters a gap. At this point, one needs to resolve the incompatibility. Most often, readers do this with very minimal effort, following heuristic preference hierarchies. One may refer to this as “interpretive adjustment.” (Such adjustment temporarily displaces the process of “spontaneous understanding.”) Preference hierarchies may be isolated for sources of information (e.g., nonpersonified versus personified), types of information (e.g., direct perception versus memory), and epistemic context (e.g., genre conventions versus real-world laws of nature). However, as gaps accumulate and become more complex, readers may have to engage in more extensive and sustained reasoning of the general sort used in scientific inference. This is the sort of process that a professional critic commonly engages in. But it is also practiced by ordinary readers in conversations about books or films. It is commonly what leads to the clarification of a work’s thematic concerns or emotional strategies. This may be referred to as “critical explication.”

Atwood’s novel shows some of the ways in which narrator unreliability is established through the development of gaps. Specifically, it presents a complex series of limited incompatibilities. These prepare the reader to find the narrator unreliable, to distrust the narrator. However,
for the most part these gaps seem to remain below the level of self-conscious awareness. Thus they probably do not provoke a self-conscious response to unreliability before the narrator makes the issue explicit. When the full extent of the unreliability becomes clear, the novel nicely illustrates how readers come to adjudicate between alternative possibilities. In this case, the reader is likely to recognize that the second story line (about an abortion) allows the means for explaining a great deal of both story lines, while the first story line (about a marriage) does not. In keeping with this, the novel also presents a complex case of epistemic and motivational misrepresentation, due to the narrator’s repressed memories and un-self-conscious confabulation of a false memory. Finally, it presents a good example of how thematic explication may result from effortful interpretation of gaps that remain even after problems of unreliability have been resolved. In this case, the gap in question bears on a particular aspect of the relation between the story and a theoretical category (psychoanalysis) drawn from the world outside the story.