conferring credibility and status on it only when some information would tip the balance in its favor.”

Because the argument of my essay focuses on literary texts, I have so far dealt and will continue dealing with verbal or verbalizable metarepresentations, such as, “Eve says it is raining outside.” Torey’s emphasis on holding images in a “tentative” way—visual metarepresentation, if you will—reminds us that, as Sacks points out, “there is increasing evidence from neuroscience for the extraordinary rich interconnectedness and interactions of the sensory areas of the brain, and the difficulty, therefore, of saying that anything is purely visual or purely auditory, or purely anything.” In other words, whether we agree with Baron-Cohen and Sperber, who think that metarepresentational ability evolved primarily to model human minds, or with Cosmides and Tooby, who suggest that its gradual emergence must have responded to a broader variety of cognitive challenges faced by our ancestors, it seems that its functioning today informs our interaction with the world on more levels than we are immediately aware of.

~2~

METAREPRESENTATIONAL ABILITY
AND SCHIZOPHRENIA

I have considered above three conjectural instances of our taking in any new information as an architectural truth. Now it is time to ask what really happens when the cognitive mechanisms that allow us to store information under advisement are damaged. A number of neurological deficits, such as autism and schizophrenia, have been linked to the failure of metarepresentational capacity, as have several kinds of amnesia. To begin with the mildest functional instance of such a failure, children develop a mature Theory of Mind around the age of four, and it is suggestive that just before that (typically, from three to four), they can go through so-called childhood amnesia, that is, a tendency to “believe that they actually experienced events that never happened, if they are asked about these (fictitious) events repeatedly,” a consequence, perhaps, of having an immature “system for source tagging.” (This is not to say, of course, that as adults
we are immune to developing false memories through external suggestion. Unless we replace the evolutionary framework with the teleological, no such immunity can be expected when we have an immensely complex system, such as our metarepresentational capacity, functioning in an immensely complex world.)

Then there are also important studies of adult patients with amnesia induced by head trauma. Such patients, it turns out, often “experience highly selective memory loss, typically displaying intact semantic memory with impaired access to episodic memory.”2 Since it has been hypothesized that episodic memories are processed via metarepresentations (that is, by enabling people to form self-reflections, for example, “I thought that I would be afraid of the dog”), the study of such selective impairment may lead to new insights into our metarepresentational ability.

Furthermore, Christopher Frith has suggested that since “self-awareness cannot occur without metarepresentation,” that is, the “cognitive mechanism that enables us to be aware of our goals, our intentions, and the intentions of other people,” specific “features of schizophrenia might arise from specific abnormalities in metarepresentation.”4 The failure to monitor the source of a representation thus can lead to patients’ perceiving “their own thoughts, subvocal speech, or even vocal speech as emanating, not from their own intentions, but from some source that is not under their control,” whereas the “inability to monitor willed intentions can lead to delusions of alien control, certain auditory hallucinations, [and] thought insertion.”5

For example, the metarepresentation, “I intend to catch the bus,” could be perceived by a schizophrenic patient as “Catch the bus,” and “My boss wants of me ‘you must be on time’” as “you must be on time,” thus making the patient experience delusions of control or think that he/she hears disembodied voices talking to or about him/her. The latter, called “a third person hallucination,” can result from perceiving a metarepresentation, such as, “Eve believes ‘Chris drinks too much,’” as a “free floating notion ‘Chris drinks too much,’” and so forth.

Note that although people with autism also lack metarepresentational capacity (to the same degree to which they lack Theory of Mind), the above delusions associated with failure of source-monitoring are typical for patients with schizophrenia but not for those with autism. Frith and his colleagues explain it by the “markedly different ages of onset” for autism and schizophrenia. The former manifests itself in the first years of life, whereas the latter usually develops in the early twenties, when the
The majority of autistic children fail to develop [Theory of Mind]. They are unaware that other people have different beliefs and intentions from themselves. Even if they manage, with much effort and after a long time, to learn this surprising fact, they will be only able to infer the mental states of others with difficulty and in the simpler cases. As a consequence they cannot develop delusions about the intentions of others. Furthermore, they will know, over a lifetime of experience, that their inferences are likely to be wrong and will therefore be ready to accept the assurance of others as to the true state of affairs.

In contrast, schizophrenic patients know well from past experiences that it is useful and easy to infer the mental states of others. They will go on doing this even when the mechanism no longer works properly. For the first 20 years or so of life the schizophrenic has handled ‘theory of mind’ problems with ease. Inferring mental states has become routine in many situations and achieved the status of direct perception. If such a system goes wrong, then the patient will continue to “feel” and “know” the truth of such experiences and will not easily accept correction.

In Sections 8 and 9 below, I focus on fictional protagonists failing to keep track of themselves as sources of their representations of other people’s minds and thus “feeling” the truth of their (wrong) mind-attributions. I show that such failures could be used by the authors wishing to tease their readers by making them unsure of what is really going on in the story and which representations originating in the characters’ minds they could trust. However, before I get to the narratives that cultivate this kind of conceptual vertigo in their readers, let us consider a more manageable example of a character clearly marked off by the author as mentally unstable.

Fedor Dostoyevski’s novels feature many self-deceiving sufferers. Prominent among them, however, is Katerina Ivanovna Marmeladova (Crime and Punishment), a gentlewoman by birth and education, now a desperately poor widow dying of consumption among her starving children. Katerina Ivanovna repeatedly invents stories that enhance her past and future and immediately starts believing in these fantasies herself, to the raucous delight of cruel onlookers. For example, at the funeral of her alcoholic second husband, she comes up with the idea that she will soon receive a pension for him (which can never happen), and she decides to use that pension to open a boarding school for refined young ladies. Some
of her listeners are simply amused by such ravings, but others, such as her landlady, find her plans as to how to run the school, which county to locate it in, and whom to hire so convincing (for Katerina Ivanovna herself believes in them) that they begin seriously advising her on how to ensure the hygiene and good morals of her pupils (405).

Katerina Ivanovna does not like the thought of accepting advice from her landlady (whom she considers infinitely beneath herself), and she lets it show. The disagreement between the two women escalates into an ugly fight. At this moment, a temporary lodger enters the room, a respectable well-to-do lawyer Petr Petrovich Luzhin. Earlier, Katerina Ivanovna had told everyone that Luzhin was a friend of her first husband, a protégé of her father, and the very man who would use his significant connections to secure her the pension (all of which is, of course, her invention). Now Katerina Ivanovna turns to this near-stranger for support:

“Petr Petrovich!” cried she, “at least you protect me! Impress upon that stupid beast that she cannot treat this way a gentlewoman in distress, that there is court and justice... I will to the Governor-General... She will answer for it... In the memory of my father's past hospitality, protect us orphans!”

“Excuse me, Madam... I beg your pardon, excuse me, Madam,” Petr Petrovich was trying to get past her. “I've never had an honor of meeting your dear father, as you well know yourself... beg your pardon, Madam!” (Someone in the room roared with laughter.) “And I have not the least intention to participate in your endless squabbles with [your landlady...].”

Katerina Ivanovna stood still, unable to move, as if struck by lightning. She could not comprehend how Petr Petrovich could disavow the hospitality of her dear father. Having once invented that hospitality, she now completely believed it herself. . . . (407–8; translation mine)

I have no intention of “diagnosing” the poor Katerina Ivanovna with selective amnesia or schizophrenia, but I do want to point out that her delusions clearly stem from the failure to monitor properly the source of her representations. Katerina Ivanovna’s “I wish I could get a pension for my husband” changes to “I get a pension for my husband,” and her “I wish this respectable and influential man (i.e., Petr Petrovich) were a friend of my first husband and a protégé of my father” registers in her mind as “This respectable and influential man was a friend of my first husband and a protégé of my father.” Note that because these representations
are allowed to circulate freely, that is, without “tags” pointing to herself as their source, in Katerina Ivanovna’s mind they produce inferences that can corrupt the already existing stores of knowledge. After all, Katerina Ivanovna’s late father had been a socially prominent figure, and Petr Petrovich could have been, in principle, welcomed in his house, if the two men had ever had a chance to meet. What happens here is that Katerina Ivanovna’s original memory of her father’s house is now corrupted by the conviction that Petr Petrovich used to be a frequent guest there. (Compare it to the hypothetical situation above, in which the information that it is raining gold, when assimilated without a source-specifying tag, such as, “It was Eve who told me,” begins to impact our other knowledge stores and results in harmful behavior, such as canceling classes, quitting the job, maxing out on credit cards, etc.).

3

EVERYDAY FAILURES OF SOURCE-MONITORING

Of course, it is not just the hapless Katerina Ivanovna who invents stories about the state of affairs in the world and begins to act upon them as if they were real. We all do it. In many cases, such self-deception is quite beneficial—as one of the more level-headed (or just differently insane) characters from Crime and Punishment observes, “Best lives he who dupes himself the best” (502). But generally, especially if we consider the closely related issue of personal memories, it makes sense to think of our partial failures to keep track of some of the sources of our representations as part of the normal functioning of the metarepresenting brain. When I say “normal,” I mean to contrast it both with the sustained, pathological pattern of such failures typical for schizophrenic patients and with the deliberately planned and carefully highlighted instances of such failures in the works of fiction.

I was reminded some time ago about everyday failures of our source-monitoring—failures that we do not even register consciously unless pressed by circumstances—while reading the account of Martha Stewart’s trial in The New Yorker (Stewart had been accused of insider trading and subsequent lying to federal agents). The author, Jeffrey Toobin, refers to a “curious” testimony by one of Stewart’s close friends, Mariana Pasternak, who, at one point, could not identify the source of one of her memories: