5

The Problem of Other People

The real challenge in the human environment throughout history that affected the evolution of the intellect was not climate, weather, food shortages, or parasites—not even predators. Rather, it was the necessity of dealing continually with our fellow humans in social circumstances that became ever more complex and unpredictable as the human line evolved.

—Richard D. Alexander, *How Did Humans Evolve?*

So this is hell. I'd never have believed it. You remember all we were told about the torture-chambers, the fire and brimstone, the "burning marl." Old wives' tales! There's no need for red-hot-pokers. Hell is—other people!

—Garcin, in Jean-Paul Sartre's *No Exit*

The Evolution of Big Brains

The human brain is adapted—of that much we can be sure. However, nobody is quite sure what it is adapted to. The usual evolutionary explanation—that the brain adapted to the natural environment in which humans developed—doesn't account for all of the facts that need to be explained. For one thing, humans and other hominids have developed and flourished in nearly every habitable environment on the earth. Even more perplexing, however, is the fact that human cognition seems to be so much more complex than it needs to be. There can be no question that some advanced cognitive abilities help considerably in finding food, securing mates, and avoiding predators. Yet nothing about the life of an ancestral human would seem to require the ability to play chess,

compose symphonies, write books, or compute differential equations. Not only do such activities lack a clear adaptive purpose, the sheer computational power necessary to perform them is beyond that required by any reasonable program of hunting and gathering, making the human brain seem, in the words of psychologist Geoffrey Miller, "like computational overkill."[1]

The cognitive differences between humans and other animals are not trivial. Evolutionary psychologist Robin Dunbar, who has written widely on brain size and social evolution, explains that the human brain is significantly larger—as a proportion of total body weight—than that the brains of any other species. Dunbar reports that the human brain is nine times larger than that of the average mammal, and twelve times larger than those of the mammal species generally considered the closest relatives to the ancestral mammals of the early Paleocene epoch. Since the human brain uses more energy than any other part of the body—brain tissue accounts for only 2 percent of the human body's weight but 20 percent of its energy use—large brains come with a very high price tag. Moreover, to accommodate their big brains, human infants must be born while their proportionally massive heads can still fit through the birth canal—meaning that human infants are born dangerously underdeveloped and completely vulnerable. Even so, an infant's head presents greater risks to its mother during childbirth than those faced by the females of almost all other mammalian species. "The fact that an organism has a large brain," Dunbar writes, "means that it really must need it very badly, otherwise the forces of natural selection will inexorably favour individuals with smaller brains simply because they are cheaper to produce."[2]

What value do huge brains add to the struggle for survival? Scientists are still debating this question. Miller, early in his career, argued that big brains were the result of runaway sexual selection for intelligence, though in his more recent book, *The Mating Mind*, he acknowledges some problems with the stronger interpretation of this hypothesis.[3] Other theories propose the need for solving adaptive problems involving such things as tool use, hunting, food processing, and the like.[4] None of these potential explanations for big brains has become widely accepted as a complete explanation; however, in 1990, University of Michigan biologist Richard D. Alexander, synthesizing the work of several other

researchers, proposed an answer that has gained steady support ever since and is now the most widely accepted of all the various hypotheses. Invoking the work of Nicholas K. Humphrey, Alexander proposes that early humans

> had in some unique fashion become so ecologically dominant that they in effect became their own principal hostile force of nature, explicitly in regard to evolutionary changes in the human psyche and social behavior. At some point in their evolution humans obviously began to cooperate to compete, specifically against like groups of conspecifics, this intergroup competition becoming increasingly elaborate, direct, and continuous until it achieved the ubiquity with which it has been exhibited in modern humans throughout recorded history across the entire face of the earth.[5]

According to Alexander's hypothesis, then, the environmental factor most responsible for the evolutionary expansion of the human brain was the presence of other human beings. Humans, in other words, are primarily adapted to each other.

In books such as *Grooming, Gossip, and the Evolution of Language* and *The Human Story*, Dunbar expands substantially upon Alexander's thesis. His research suggests that human evolution occurred largely within groups of about 150 members (which remains the ideal size for many social organizations today). In order to interact successfully in such a group, a person would have had to keep track of 149 other minds. He or she would need to understand each person's motives, history, and temperament, and would also have to show some understanding of each person's relationships with each of the other members of the group. In this way, a person could attract mates, form alliances, reciprocate both slights and favors, and acquire status within the group. Keeping track of 149 minds and the millions of possible relationships among those minds requires much more computational power than playing a game of chess or composing "Claire de Lune." In *The Art Instinct*, Dennis Dutton points out that, while thirty-two chess pieces on sixty-four squares of a chessboard can produce trillions of possible moves, these possibilities "would be dwarfed in range by the contingencies served up by daily experience in a human lifetime."[6] This is true largely because the

pieces we must keep track of—other people and their relationships to each other—are so much more complex in even the simplest human community. Small cognitive disparities in such an environment would set off what evolutionary biologists call an "arms race"—a situation in which evolution occurs rapidly because small advantages possessed by some members of a population place intense selection pressure on the rest of the population to match and exceed those advantages.[7]

If this hypothesis is correct, then we would expect that the most interesting thing to human beings would be other human beings—specifically, the intimate details of their lives and relationships. We might expect people to devote an inordinate amount of time to gossiping about their acquaintances, reading tabloid reports of the private lives of high-status people, and, perhaps most importantly for the present study, reading certain kinds of fiction. Dunbar himself recognizes the potential of evolved social intelligence to respond to fictional accounts of other minds:

> Of all the books published each year, it is fiction that tops the list in terms of volume of sales. Take a glance around your local bookshop: university campus bookshops aside, two-thirds of the shelf space will contain fiction. Even then, it is not the rip-roaring adventure yarns that attract us, but the unfolding intimacies of the main characters. It is the way they handle their experiences that fascinates us, their reactions to the vagaries of life. . . . And out of all this fiction, it is not the writing of the acclaimed masters that tops the publishers' sales-lists, but romantic fiction.[8]

With only a few exceptions, fiction is about other people's minds, and other minds are inherently interesting to us for the same reason that trees are inherently interesting to woodpeckers: other people are the environmental backdrop against which we evolved. Other minds were the chief adaptive problem—and therefore the most significant source of anxiety—that early humans had to face.

~~*Sinners in the Hands of an Angry Paradox*~~

When the three principal characters in Jean-Paul Sartre's *No Exit* are escorted to their final destination, they are surprised by the relative timidity of hell. Each of them—Garcin the Brazilian deserter, Inez the

lesbian manipulator, and Estelle the society seductress—expects to find the hell of mythology—a gaping pit full of fire, brimstone, and horned demons with implements of torture. Instead, they are led, one at a time, into a reasonably comfortable drawing room in what appears to be a Victorian-style hotel. However, before they can take comfort in their good fortune, the cleverest of them, Inez, figures out the game. "Each of us," she tells the other two, "will be the torturer of the other two." Understanding the truth of Inez's statement, Garcin immediately makes what he thinks is a perfectly sound proposal:

> No, I shall never be your torturer. I wish neither of you any harm, and I've no concern with you. None at all. So the solution's easy enough; each of us stays put in his or her corner and takes no notice of the others. You here, you here, and I there. Like soldiers at our posts. Also, we mustn't speak. Not one word. That won't be difficult; each of us has plenty of material for self-communings. I think I could stay ten thousand years with only my thoughts for company.[9]

All three agree, but their agreement ends in a matter of seconds. The characters cannot mind their own business because they have no business of their own to mind. Each of them has lived a life of passive reflection, depending on their lovers, their organizations, and even their enemies to reflect them back to themselves. They have never created authentic existences for themselves—never, in Simone de Beauvoir's phrase, "cast themselves into the world."[10] Even in hell they must depend upon each other for any hope of self-definition. At first they each visit their former acquaintances in the vain hope that they will hear something about themselves. (Had he written today, I am quite sure that Sartre would have provided them with a computer and had them Google their own names.) As these impressions fade, they turn to the others for validation of their existence, first through sexual affirmation and then through sheer provocation. Inez sums up the dynamic that they all face when she says, "I can't get on without making people suffer. Like a live coal. A live coal in others' hearts. When I'm alone I flicker out."[11] On this point, Sartre's Hell actually corresponds quite closely to Dante's. Most of the sinners in the first five levels of *Inferno* either punish themselves or punish each

other with no need of either divine or demonic intervention, culminating in the torment of the wrathful on the banks of the River Styx who, with no compulsion beyond their inner passions, "thumped at one another in that slime / with hands and feet, and they butted, and they bit / as if each would tear the other limb from limb."[12]

In *No Exit* (as in *Inferno*) the characters are involved in versions of a famous game-theory scenario known as the "prisoner's dilemma," which evolutionary psychologists often use to explain the prevalence of altruistic behavior among organisms in direct evolutionary competition with each other. The classic prisoner's dilemma game features two players who must choose either to "cooperate" or "defect." Mutual cooperation gives both players a modest reward; mutual defection gives each a modest penalty. When one player cooperates and one defects, the defector is richly rewarded and the cooperator is severely punished. In *The Prisoner's Dilemma*, William Poundstone describes the imaginary scenario upon which the exercise is based:

> Two members of a criminal gang are arrested and imprisoned. Each prisoner is in solitary confinement with no means of speaking to or exchanging messages with the other. The police admit they don't have enough evidence to convict the pair on the principal charge. They plan to sentence both to a year in prison on a lesser charge. Simultaneously, the police offer each prisoner a Faustian bargain. If he testifies against his partner, he will go free while the partner will get three years in prison on the main charge. Oh, yes, there is a catch. . . . If *both* prisoners testify against each other, both will be sentenced to three years in jail.[13]

A prisoner's dilemma is created whenever two (or in some variants more) people are in a situation with the following attributes: 1) the result of mutual cooperation is better than the result of mutual defection; 2) the worst ending position results from unilateral cooperation; and 3) the best ending position results from unilateral defection.[14] The prisoner's dilemma is, in the words of political scientist Robert M. Axelrod, "an abstract formulation of some very common and very interesting situations in which what is best for each person individually leads to mutual defection, whereas everyone would have been better off with mutual cooperation."[15]

In Sartre's version of the game—let's call it the "sinner's dilemma"—the sinners can cooperate by simply leaving each other alone to enjoy whatever comforts hell has to offer. This is precisely the agreement that Garcin proposes and the others in *No Exit* pretend to accept. However, for some of the players—such as Inez, who "can't get on without making people suffer"—the temptation to defect is too great, and, as in any prisoner's dilemma, if one person defects, the other(s) must defect as well to avoid the worst of the outcomes. The actual payoff matrix, assuming a two-person game, looks something like this:

	YOU COOPERATE	YOU DEFECT
I COOPERATE	Nobody bothers anybody else and we both live in peace and quiet for all of eternity, bored but uninjured.	I sit passively, not doing anything to you as you torture me with glee forever.
I DEFECT	I get to have all the fun of torturing you for eternity and don't even have to worry about you doing anything to me.	We spend eternity torturing each other. I suffer pain, but at least I have the fun of making you suffer, too.

If I can be assured that you will cooperate, then it makes sense for me to cooperate too. However, the whole point of the prisoner's dilemma is that I cannot know what you will do. Therefore, the only rational solution is for me to defect under the assumption that you will defect as well. That way, no matter what happens, I will not end up a chump. In a single round of the prisoner's dilemma, the only equilibrium point—the point at which each player has achieved the best possible result given the other player's strategy—occurs when both players defect. Though they can technically stop any time they want, Garcin, Inez, and Estelle are compelled by the logic of their situation to go on tormenting each other for all of eternity.

The prisoner's dilemma has become the most famous of all game-theory scenarios because it reveals a crucial tension at the core of the

human experience: we need other people, but we can't trust them. The guiding logic of the prisoner's dilemma, when projected on an entire society, takes us perilously close to Thomas Hobbes's "war of all against all" where life is "solitary, poor, nasty, brutish, and short." Any individual has a motive to cooperate only when everybody else cooperates too; when cooperation breaks down, it is in everybody's best interest to get what they can before somebody else gets it first. Yet many societies have managed to stay together long enough to create pyramids, highways, great cities, poetry, art, music, and even space stations. Until Robert M. Axelrod conducted his now famous computer tournaments in the 1970s, many people simply dismissed cooperative human societies as either irrational or bad at math. What Axelrod discovered, however, is that something remarkable happens when the prisoner's dilemma is played multiple times against the same opponent. In the game of "iterated prisoner's dilemma," players have a chance to remember what other players have done and to respond reciprocally to both cooperation and defection—and to anticipate how others will reciprocate in the future. The iterated prisoner's dilemma changes everything we know about the logic of the game.

In 1979 Axelrod, a political science professor at the University of Michigan, invited prominent game theorists from around the world to submit, in the form of simple computer programs, decision rules for a tournament of iterated prisoner's dilemma. He describes the ground rules for the tournament in his book, *The Evolution of Cooperation*:

> It was structured as a round robin, meaning that each entry was paired with each other entry. As announced in the rules of the tournament, each entry was also paired with its own twin and with RANDOM, a program that randomly cooperates and defects with equal probability. Each game consisted of exactly two hundred moves. The payoff matrix for each move . . . awarded players 3 points for mutual cooperation, and 1 point for mutual defection. If one player defected while the other player cooperated, the defecting player received 5 points and the cooperating player received 0 points.[16]

Axelrod's experiments have been summarized and analyzed in well-known books by Poundstone, Matt Ridley, Richard Dawkins, and

others—as well as in Axelrod's own book-length treatment of the subject.[17]

In two separate tournaments guided by these rules, well-known theorists submitted programs with names such as TESTER (defects on the first round but cooperates as soon as another program defects), and TRANQUILIZER (cooperates at first and then tries to "get away" with defections later in the round). The winning strategy, submitted by Professor Antanol Rapoport of the University of Toronto, was called TIT FOR TAT and required only four lines of computer code. TIT FOR TAT always cooperated on the first move and thereafter did whatever its opponent did on the previous move. TIT FOR TAT cooperated with friendly programs, punished defectors, and immediately assumed a cooperative stance when defectors attempted to cooperate. When Axelrod created a natural-selection simulation—an environment in which each program was given representation in future "generations" according to its success in previous ones—TIT FOR TAT and other programs that cooperated initially and never defected first (and were therefore indistinguishable from TIT FOR TAT when they played each other) eventually drove out programs that defected initially or attempted to secure advantages through occasional defections during play.

TIT FOR TAT was so successful because it combined five different factors: 1) it was unselfish (Axelrod uses the phrase "not envious"), meaning it never tried to win more than its opponent in any single game, but was successful only to the extent that it enabled the other program to be just as successful; 2) it was nice, meaning it was never the first to defect; 3) it was reciprocal in both rewarding cooperation and punishing defection; 4) it was forgiving, meaning it only punished defections once and then reassumed cooperation as soon as the other program did; and 5) it was simple, meaning it followed clear, easy-to-understand rules that allowed other programs to respond to its logic. These five points correspond with many of the principles upon which the world's great religions are based: Christianity emphasizes forgiveness, Buddhism stresses simplicity, Jewish law ("an eye for an eye and a tooth for a tooth") encodes reciprocity, and Islamic law (through obligatory alms known as the *zakat*) mandates unselfishness. This should not be surprising; according to Axelrod's interpretation of his experiments, these are the rules that enable essentially selfish beings to form cooperative communities.

It would be shocking if religions—among the most important historical facilitators of community—had not discovered them.[18]

The real danger in cooperative environments—and therefore one of the primary adaptive problems for our human ancestors—comes from those who attempt to secure the advantages of cooperation without incurring any of the costs. In game theory, as in life, these people are called "freeloaders," "parasites," and, most often, "cheaters." As Leda Cosmides and John Tooby explain, "a cheater is an individual who illicitly benefits himself or herself by taking a benefit without having satisfied the requirement that the other party to the contract made the provision of that benefit contingent on." The more complicated the network, the easier it is to cheat. In a group of cooperators, therefore, the potential genetic benefits of deceptive non-cooperation are enormous, and the ability to detect non-cooperators appears to be hard-wired into human nature. Cosmides and Tooby argue that we have developed "cognitive adaptations that are specialized for reasoning about social contracts," and, specifically, "inference procedures specialized for cheater detection."[19] One of the more striking pieces of evidence for such a cognitive module involves a logical task that psychologists call the "Wason Selection Task," which involves selecting pieces of information that are important in proving or disproving a proposition. As science writer Matt Ridley explains, "people are surprisingly bad at the Wason test in some circumstances—for instance, if presented with it as an abstract piece of logic—but surprisingly good at it in others. In general, the more the puzzle is presented as a social contract to be policed, the easier people find it, even if the contract is deeply foreign and the social context unfamiliar."[20]

Theories of Mind and Theories of Fiction

The great French satirist John-Baptiste Poquelin, best known by the stage name Molière, captures the problem of other people admirably in his 1664 satire, *Tartuffe*. The title character of *Tartuffe* is a predatory religious hypocrite—a pious con man who has insinuated himself into the household of the wealthy-but-naïve Orgon. We join the characters *in media res*, after Tartuffe has moved in and become the de facto master of the house. Orgon's wife, Elmire, and his brother in law, Cléante, have not been deceived by Tartuffe's hypocrisy, but Orgon will not listen to their warnings. He becomes more and more enamored with his hypocritical

friend and contracts a marriage between Tartuffe and his daughter Mariane—even though she is already in love with, and legally engaged to, an honorable young man named Valère. Not content with taking almost everything from Orgon, Tartuffe makes a clumsy attempt to seduce Elmire, who rebuffs him and tries to report the incident to Orgon.

Orgon, however, will have none of it. In a stunningly short-sighted display of loyalty, he signs over house and property to Tartuffe, giving him, as part of the bargain, a chest of potentially treasonous documents (written by an exiled friend). Not long afterward, Elmire forces her husband to confront the truth by pretending to give in to Tartuffe's advances while Orgon hides underneath a table. By the time that Orgon finally does obtain incontrovertible evidence of his false friend's treachery, Tartuffe owns his house, his estate, and a chest of documents that could send Orgon to prison. Unable to reflect critically on his own errors, Orgon simply moves from extreme credulity to extreme skepticism. "Enough, by God! I'm through with pious men," he tells his brother-in-law. "Henceforth I'll hate the whole false brotherhood / And persecute them worse than Satan could." At this point Cléante steps in and gives what many consider the moral of the story:

> Ah, there you go—extravagant as ever!
> Why can you not be rational? You never
> Manage to take the middle course, it seems,
> But jump, instead, between absurd extremes.
> You've recognized your recent grave mistake
> In falling victim to a pious fake;—
> Now, to correct that error, must you embrace
> An even greater error in its place,
> And judge our worthy neighbors as a whole
> By what you've learned of one corrupted soul?
> Come, just because one rascal made you swallow
> A show of zeal which turned out to be hollow,
> Shall you conclude that all men are deceivers,
> And that there are today no true believers?[21]

Orgon's reversal is often taken as the defining satirical moment of the play. David F. Maas, for example, presents Cléante's speech as a

culminating moment in the author's goal to "assist individuals to make rational adjustments, and therefore to find a wider choice of strategies than those available from either-or evaluations."[22] In *Mimesis and the Human Animal*, one of the early pioneering works of adaptionist literary criticism, Robert F. Storey uses this same scene to reject overly theorized explanations of human behavior in favor of relatively simple ones:

> The example of Orgon should give us pause. He's obsessed, it is true, but his obsession is a symptom, not a cause, of his absurdity. When his eyes are finally opened to Tartuffe's hypocrisy, his reaction is not to throw off his shackles, but to chain himself blindly to another illusion. . . . Because he seems such a vital creature, modern readers have been eager to account for his weaknesses in psychologically determinate ways: he a repressed . . . homosexual, he's a "narcissist" . . . he's a tyrant who is punishing his family. . . . All of this may be true, but the most plausible explanation is that he is a fool, and that fools are incorrigibly stupid.[23]

Orgon is certainly stupid, but his stupidity is part of an overall satirical strategy. Any argument for *Tartuffe* as a corrective satire must focus on Orgon, as Tartuffe simply lacks the good qualities that might make it worthwhile to correct his bad ones. Molière is not telling us to avoid being predatory hypocrites; rather, he is telling us to avoid 1) being taken in by hypocrites, and 2) to avoid using the existence of hypocrites as an excuse to discount genuine virtue. Phrased this way, Orgon's central problem in *Tartuffe* bears a striking similarity to the central problem in the prisoner's dilemma, and to the much more important problem of living in a world made up of both cooperators and defectors: how can we tell them apart?

The rational "middle course" that Cléante advocates—to avoid hypocrites and trust those whose virtue is genuine—is much more difficult than Molière makes it seem. Most hypocrites aren't as easy to detect as Tartuffe, and most of the people deceived aren't as credulous as Orgon. In the real world, people's motives are usually a convoluted mix of altruism and self-interest, and detecting non-cooperators requires enormous skill.

Symptoms of anxiety caused by interactions with other people have

been observed in all human cultures as well as in some of our nearest primate relatives.[24] Significant anxiety disorders lie at both horns of the social version of the prisoner's dilemma. Those who cannot trust other people at all often develop paranoia and persecutory delusions, which have both been linked to the need to keep track of and respond to legitimate social threats.[25] Those who do not feel competent to engage other people and build cooperative relationships are often diagnosed with social anxiety disorder (SAD), or social phobia—the most commonly diagnosed of all anxiety disorders.[26]

When it comes to social interaction, the stakes are high and the consequences for failure are great. Many psychologists now believe that human beings have evolved a specialized cognitive mechanism for inferring the mental and emotional states of others. According to this hypothesis, selection pressures faced by early hominids led to an increase in the size of social groups, which also required an increase in cognitive abilities. Larger groups meant more potential allies, sexual partners, rivals, and enemies to keep track of, plus more reciprocal relationships to process. (Keep in mind that every new person added to a social group increases the possible relationships among group members exponentially, so even a small growth in average group size required a massive increase in cognitive power.)[27] At some point it became advantageous to evolve a cognitive process for inferring the thoughts of others by combining what we know about them with what we gather from their tone of voice, hand movements, facial gestures, and other contextual cues that signal one's frame of mind. However, as we got better at detecting the thoughts of others, we also got better at concealing our own thoughts, setting off an evolutionary arms race between detection and concealment that lead to the extremely sophisticated cognitive mechanism that psychologists now call the "theory of mind."

Most people today exercise their theory of mind automatically without realizing that it is an extremely complicated process that evolved over hundreds of thousands of years. For most of us it is second nature to infer other people's intentions through contextual clues such as their mannerisms, their tone of voice, or their "body language." Most people cannot even remember a time when they did not understand that other people's beliefs could be different than their own. It is not until we encounter people with difficulties forming a theory of mind that we realize what a complicated cognitive mechanism it really is. In the late

1980s, autism researchers Simon Baron-Cohen and Uta Frith theorized that individuals with autism spectrum disorders have difficulty "reading minds," or correctly attributing to other people states of mind that differ from their own. Together, Baron-Cohen and Frith developed the "Sally/Anne" test, which they used to diagnose autism in children. In *Mindblindness: An Essay on Autism and Theory of Mind*, Baron-Cohen describes the Sally/Anne test as follows:

> The test involves seeing that Sally puts a marble in one place and that later, while Sally is away, Anne puts the marble somewhere else. The child needs to appreciate that, since Sally was absent when her marble was moved from its original position, she won't know it was moved, and therefore must still believe it is in the original location. . . . On the test question "Where will Sally look for her marble?" the vast majority of normal children and children with Down's syndrome passed the test, indicating the original location. But only a small minority of the children with autism did so. Instead, most of them indicated where the marble really was. Since the children with autism were older and had a higher "mental age" than the children in either of the two control groups, this study supports the notion that, in autism, the mental state of belief is poorly understood.[28]

Reading minds requires us to use the same kinds of metarepresentational tags that we use to process contradictory propositions. For example, most people would have no trouble processing the sentence "John believes that Mary is in Jamaica" even if they knew that Mary was sitting right next to them at a restaurant in Duluth. The brain stores knowledge of Mary's actual location as a representation about the world we live in, while it stores knowledge of John's state of mind as a metarepresentation. We can use this same strategy to keep track of much more complicated propositions, such as: "John believes that Mary wrote a book about Pam's belief that Jerry was kidnapped by pirates." Such a statement requires us to keep track of four propositions—what John believes, what Mary wrote, what Pam believes, and who kidnapped Jerry—three of which may be recognizably false without altering the truth value of the overall statement.

The process of creating source tags for information contained in other people's minds is inextricably connected to the process of generating narratives about other people's motives. In many cases we cannot create a tag without first creating a narrative to accompany it. If, for example, John tells me that Mary has been in Jamaica for three days, while I am sure that I had dinner with Mary two nights ago, I must construct a narrative into which I can set both propositions before I can assign one of them a source tag. Such narratives include, but are certainly not limited to:

> I have been extremely stressed at work and forgot that my dinner with Mary was four nights ago.
>
> John is lying to me because he is in love with Mary and doesn't want me spending time with her.
>
> Mary's twin sister, Terri, was secretly impersonating Mary at dinner so that she could pump me for information about their brother, Berry.
>
> Malevolent space aliens kidnapped Mary and cloned her, sending the real Mary to Jamaica and the phony Mary to dinner with me in order to infect me with a retrovirus that would reorganize my genome and help them produce their wretched young.

While these stories vary widely in probability, they each contain the basic components of any narrative. They also contain elements that most readers of fiction will immediately recognize: romance, intrigue, suspense, and secret plots. Without at least some of these elements, the narrative does not contain enough information to tell me how to deal with John. Simply determining that "John is lying" will not suffice. Without understanding the motive for the lie, I will not know how I should treat John, which is the reason that I need a theory of mind in the first place.

Once the theory of mind mechanism evolved, it became one of the principal reasons that we now are able to process, and even enjoy, fictional representations. People with a healthy theory of mind are capable of feeling like they "know" a literary character—whether it is a favorite friend like Elizabeth Bennett or an amoral scoundrel like Iago. We en-

counter and interpret these characters much in the same way that we encounter and interpret real people. This connection between theory of mind and fiction has recently been made clear in Lisa Zunshine's book, *Why We Read Fiction*. Zunshine argues that fiction, much like chase play, allows us to exercise and hone skills that are crucial to our survival. Dunbar similarly presents fictional narrative as a logical outgrowth of a theory of mind. As we become increasingly good at nesting thoughts in different minds, he argues, we become able "to imagine how someone who does not actually exist might respond to particular situations. In other words, we can begin to create literature."[29] When we encounter fictional minds, then, we get what Zunshine describes as a "work out" of "our ability to store representations under advisement and to reevaluate their truth-value once more information comes in."[30]

Zunshine's "literature-as-theory-of-mind-workout" proposition can account for a good deal of what we very broadly label "realistic fiction"—literature designed to reproduce the complex psychology of human interaction. I can almost feel my cognitive muscles flexing when I try to determine whether Estella really loves Pip or whether Mr. Darcy still wants to marry Elizabeth Bennett. It is more difficult, however, to see how something like *Tartuffe* might have the same effect. Molière does not present Tartuffe as a complex character whose intentions and motivations must be painstakingly deciphered. While it is true that Orgon makes tremendous mistakes because he misreads another mind, viewers are never encouraged to see Tartuffe through Orgon's eyes or to make a difficult judgment about the title character's true intentions. We know exactly what Tartuffe is from the play's full title, *Tartuffe, or the Hypocrite* (in French, *Tartuffe, ou l'Imposteur*), so our enjoyment of the play can hardly stem from the challenge of deciphering his state of mind. Clearly, something else must be going on.

William Flesch's *Comeuppance: Costly Signaling, Altruistic Punishment, and Other Biological Components of Fiction*—an interesting recent extension of Zunshine's "workout" argument—offers several good candidates for this something else. Flesch argues that fiction gives us practice, not just in deciphering the minds of other people, but also in responding to them with the appropriate approbation or outrage. After an extensive discussion of the prisoner's dilemma game, he argues that "humans cooperate, and continue cooperating, because we monitor

one another's cooperation vigilantly. To give us an incentive to monitor and ensure cooperation, nature endows us with a pleasing sense of outrage at defection and a concomitant sympathy for the victims of deception." The pleasure we derive from sympathizing with victims and feeling outraged at defectors transfers easily to fictional narratives:

> We are fitted to track one another and to track as well how others monitor one another and what they do when they monitor one another. What we wish to track is past behavior, including past tracking of past behavior, in order to respond in the present to that behavior. Fiction recruits this central capacity in human social cognition for taking pleasure in responding to the nonfactual. It gratifies the proximal or psychological aim of our interest in what some have done and how others have responded. That aim is the pleasure we take in strong reciprocation, especially punishment, a pleasure useful in nonliterary contexts as an incentive to altruistic punishment and presumably evolved for that reason.[31]

The importance of punishing defectors can, for Flesch, be illustrated by another game-theory exercise called the ultimatum game. In this game, two people are collectively offered a sum of money. "The person who receives the money," Flesch explains, "has to propose a split with the person who doesn't receive the money. . . . The other person can accept the split or veto it. If he or she accepts it, the money is divided as proposed. If the person vetoes it, neither of the players gets anything."[32] A rational proposer should propose the most uneven split allowed in the rules, and the rational responder should accept any proposed division. Even a ninety-nine-to-one split leaves both players better off than a veto. However, when the game is played under experimental conditions, most proposers offer a fifty-to-fifty split, and most responders reject any proposal that gives them less than 25 to 30 percent of the money.[33] Both proposers and responders, in other words, respond to a nonrational, altruistic view of fairness. Flesch's argument focuses specifically on the responders—those who give up what is essentially free money in order to punish those they perceive as unfair. The same cognitive predisposition that causes us to engage in altruistic punishment causes us to admire others who engage in altruistic punishment, including fic-

tional characters, which goes a long way toward explaining why good guys usually triumph over bad guys in fiction. We are wired to derive pleasure when we see defectors punished.

The end of *Tartuffe*—like the ends of the *Odyssey*, "Hansel and Gretel," and *Die Hard*—gives us exactly the comeuppance that Flesch describes. When Orgon discovers Tartuffe's treachery and tries to force him to leave, Tartuffe takes control of the house (that Orgon signed over to him) and demands that his former host leave. He attempts to enforce his eviction order by summoning a police officer and accusing Orgon of treason—a charged leveled on the strength of the documents that Tartuffe possesses. The officer comes to the house, appearing to have instructions to arrest Orgon; at the last minute, however, he arrests Tartuffe instead, with the explanation that the king has detected Tartuffe's treachery:

> His royal soul, though generous and human,
> Views all things with discernment and acumen;
> His sovereign reason is not lightly swayed,
> And all his judgments are discreetly weighed.
> He honors righteous men of every kind,
> And yet his zeal for virtue is not blind,
> Nor does his love of piety numb his wits
> And make him tolerant of hypocrites.
> 'Twas hardly likely that this man could cozen
> A King who's foiled such liars by the dozen.[34]

By the end of the final scene, Tartuffe has been exposed, humiliated, and condemned to prison—all by the actions of the wise king (King Louis XIV later intervened on Molière's behalf when the French Church denounced *Tartuffe* as heretical) who saves the day as both the detector and the punisher of Tartuffe's decidedly non-altruistic behavior.

Taken together, however, Zunshine's characterization of fiction as a workout of our mindreading abilities and Flesch's theory of fiction as an exercise in altruistic outrage offer a compelling way to understand the intense pleasure that we derive from the experiences of imaginary people. While it makes intuitive sense that fictional experiences would provide less pleasure than real experiences, the opposite may well be

true. Fictional minds give us more pleasure than real minds for the same reason that tigers in a zoo are fun while tigers in backyards are terrifying. Natural selection designed us to pay special attention to tigers. The captive tiger allows us to enjoy this extra focus on a purely aesthetic level, knowing that we are in no actual danger. Similarly, encountering a fictional mind, as Brian Boyd emphasizes, "removes the dangers of deceit or manipulation and offers the promise of interest . . . [and] therefore offers a win-win situation, a non-zero-sum game, an advantage for teller (benefit in attention and status, at a cost in imaginative effort), and for the listener (maximum cognitive interest at little cost except time)."[35] This is not to say, of course, that the theory of mind evolved to make stories possible—any more than the visual cortex evolved to help us appreciate fine paintings. These adaptations were designed to help us process information, avoid predators, locate food, find mates, and otherwise increase our evolutionary fitness. One of nature's greatest motivators is pleasure—pleasure from the sight of bright, contrastive colors or from the experience of encountering and deciphering another person's motives—and one of art's functions is to provide these pleasures without the corresponding threats to our lives or well being.