



From your computer ...



CURRENT ISSUE

[complete contents](#)

FEATURES

[new democracy](#)

[forum](#)

[new fiction forum](#)

[poetry](#)

[fiction](#)

[film](#)

[archives](#)

ABOUT US

[masthead](#)

[mission](#)

[rave reviews](#)

[contests](#)

[writers' guidelines](#)

[we're hiring](#)

[advertising](#)

SERVICES

[bookstore locator](#)

[literary links](#)

[subscribe](#)

Reading Your Mind

How our brains help us understand other people

Rebecca Saxe

► False beliefs are among the familiar and awkward facts of life. You fail to show up on a Friday night because you thought the party was on Saturday. A friend overdraws her checking account because she thought there was more money in it. Sometimes the intrigue caused by false beliefs becomes limitlessly complex, as when a secretly married girl takes a sleeping potion to avoid being forced to marry another man, only to wake to find her true husband has killed himself because he thought she was dead.

But false beliefs are not only a source of mundane embarrassments and Shakespearean plots. Our ability to recognize when other people have false beliefs, and to consider these beliefs in explaining their behavior, provides a window on basic features of the human mind.

Consider what cognitive scientists call the false-belief task. A preschooler is presented with a junior analogue of a Shakespearean plot, with two main characters, Sally and Anne. He is told that Sally has a ball, that she has put the ball in a brown basket and gone outside; that Anne takes the ball from the basket and plays with it inside the house and then puts it in a green box; and that Sally is now coming back inside to get her ball. Where, he is asked, will Sally look for her ball?

We know that Sally will look for the ball in the brown basket: that is where she put it, and she thinks it is still there. Five-year-olds see it the same way: they breeze through the false-belief task. Not so three-year-olds. The younger children consistently predict the opposite: they expect Sally to look for her ball in the green box, where the ball really is. It's as if the three-year-olds cannot take Sally's (false) belief about the ball into account in predicting her behavior. To make sense of Sally's search, or Romeo's suicide, we

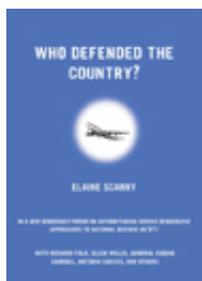
Search this site or [the web](#)
Powered by [FreeFind](#)

Find!

Site

Web

New Book From BR



Who Defended the Country?by Elaine Scarry
(Beacon Press)

need to understand that people's actions are caused by their own beliefs about the world (which may be mistaken), and not by the world itself. The three-year-old effectively predicts that Romeo will not kill himself, because Juliet is not really dead.

Children's early understanding of what makes people do the things they do appears to develop in two stages. In the first stage, children understand that people act in order to get the things they want: that human beings are agents whose actions are directed to goals. At 18 months, a child already understands that different people can have different desires or preferences—that for instance an adult experimenter may prefer broccoli to crackers, even though the infant herself much prefers crackers. Toddlers not yet two years old talk spontaneously about the contrast between what they wanted and what happened. Even nine-month-old infants expect an adult to reach for an object at which she had previously looked and smiled.

Children in the first stage are missing something very specific: the notion of *belief*. Until sometime between their third and fourth birthdays, young children seem not to understand that the relationship between a person's goals and her actions depends on the person's beliefs about the current state of the world. Two-year-olds really do not understand why, if Sally wants the ball, she goes to the basket, even though the ball is in the box. They do not talk spontaneously about thoughts or beliefs, and have trouble understanding that two people could ever have different beliefs. Similarly, while a five-year-old knows that she has to see a ball to be able to tell whether its red, a three-year-old believes he could tell if the ball is red just by feeling it. In the first stage, children think that the mind has direct access to the way the world is; they have no room in their conception for the way a person just believes it to be.

The limitations of a stage-one understanding of the mind apply even to the child's own past or future beliefs. If you show a child a crayon box and ask her what she thinks is inside, all children will say that the box contains crayons. But if you open the box to show that it actually contains ribbons, re-close the box, and then ask the child what she thought was in the box *before* it was opened, the three-year-old children claim they thought all along that the box contained ribbons.

An impressive conceptual change occurs in the three- or four-year-old child. From American and Japanese urban centers to an African hunter-gatherer society, children make a similar transition from

the first stage of reasoning about human behavior, based mainly on goals or desires, to the richer second stage, based on both desires and beliefs. What explains the change? How do children acquire the idea that people have beliefs about the world, that some of the beliefs are false, and that different people have different beliefs about the same world? Between three and five, children mature in so many ways: their vocabulary increases by orders of magnitude, their memory improves, they just know more facts about the world. Each of these changes might account for the advantages of a five-year-old over a three-year-old in solving the false-belief task.

But more than just an accumulation of knowledge is at issue. Rather, we seem to be equipped by evolution with a special mental mechanism—a special faculty or module in our minds—dedicated to understanding why people do the things they do. The maturation of this special mechanism between three and four, in addition to all the other changes happening around the same time, makes the difference between a child who simply doesn't get Romeo's decision and one who does.

A Mental Module

The idea that human beings are endowed with a special faculty for reasoning about other minds fits into a much wider and older tradition of debate about the origin of all concepts, especially relatively complex ones. Most psychologists would grant that some basic perceptual primitives—for example, color, sound, and depth—are derived from the physical world by dedicated innate mechanisms in the mind. But where do more abstract concepts come from—concepts such as *house*, *belief*, or *justice*? How, for example, does a child originally learn that other people have beliefs?

One answer is that the mind uses powerful general learning mechanisms—the same mechanisms that help us learn about any other subject matter—to detect correlations (or other, more complex statistical relationships) between occurrences of the primitives, and then builds abstract complex concepts out of patterns of simpler perceptual ones. John Locke, George Berkeley, and David Hume defended essentially this story in the 18th century, and many modern psychologists have found it compelling. To acquire the concept *person*, a child might join together her perceptual primitives corresponding to the visual appearance of people, the sounds people make, and the ways people move. The concept *belief* would require an even more complex

conglomeration of primitives.

An alternative answer is that the acquisition of certain concepts is like the acquisition of language. We do not master the grammar of our native language—for instance, learn how to form questions—by applying general learning mechanisms, but by using a special set of principles specific to language acquisition: a so-called language faculty. Similarly, some of our more complex concepts themselves, and special learning mechanisms devoted to acquiring those concepts, may be programmed into our minds from the start. In the case of the human ability to explain and predict action by attributing and reasoning about beliefs, three kinds of evidence suggest that a special mental mechanism is at work.

Privileged reasoning. Reasoning about beliefs seems to be relatively privileged in most people: it develops earlier and resists degradation longer than other, similarly structured kinds of logical reasoning.

For a four-year-old, the false-belief task described above is a very hard problem. And early on, researchers thought that young children might find it especially difficult to reason about beliefs because they are invisible and intangible. But an ingenious experiment by Debbie Zaitchik indicated otherwise. Zaitchik devised a version of the false-belief task which is the same in every respect, except that it uses a (concrete, tangible) out-of-date photograph. In the false-photograph story, after Sally puts the ball in its original location, the brown basket, she takes a Polaroid photograph. (The preschooler subject is given a chance to play with the camera before the experiment begins). Then Anne moves the ball from the basket to the second location, the green box. Before the child is allowed to see the picture, he is asked to predict: *where will the ball be in the picture?*

If invisible, intangible beliefs make the false-belief task especially difficult, then the false-photograph version ought to be easier. In fact, Zaitchik found the opposite: for young children the false-photograph version is significantly harder than the original false-belief task. Counterintuitively, the need to reason about beliefs—rather than other representations of the world, such as photographs—makes the false-belief task easier for children. The same is true for Alzheimer's patients: reasoning about beliefs resists degradation by encroaching dementia longer than other kinds of logical reasoning, including the false-photograph problem. Even healthy young adults respond faster and more accurately to

the false-belief version. Most people seem to have a natural fluency in thinking about beliefs, and this fluency helps to overcome the logical demands of a problem about the contents of another mind.

Asperger's and autism. There are some striking exceptions to this natural fluency, and the exceptions can be particularly informative. Children and adolescents with autism or Asperger's Syndrome struggle to think about other minds in ways that most people find effortless. Even when their logical abilities are otherwise fully intact, people suffering from autism and Asperger's present a selective impairment in their capacity to attribute beliefs to others, a dramatic reversal of the general pattern. This difficulty with understanding other's thoughts emerges clearly in the false-belief task. Autistic subjects pass the false-photograph task, but not the false-belief task—precisely the opposite pattern of performance from most four-year-olds. Whatever it is that makes four-year-olds, undergraduates, and Alzheimer's patients so fluent in reasoning about the mental causes of behavior is catastrophically absent in the autistic.

Four-year-olds and even adults who can successfully perform the false-belief task are therefore probably not using just the general logical problem-solving abilities that help with arithmetic and cooking and physics. The logical demands of the false-photograph version are very similar, but performance is different: worse among four-year-olds, Alzheimer's patients, and healthy adults, much better among children with autism. The idea that reasoning about beliefs is supported by a distinct, special-purpose mental mechanism is consistent with all of these results.

Still, we can easily find alternative explanations for differences in performance on such tasks. Photographs are less common than beliefs, both in real life and as subjects of conversation. So knowledge about photographs may take longer to acquire and erode more rapidly than knowledge about beliefs. Perhaps subjects with autism, who tend to be socially isolated, are simply less well-practiced at talking about the mind; or maybe they are uninterested in it, relative to other subjects. What we would really like is to be able to get inside the mind, to be able to see the stages of mental processing directly, and so to be able to see for ourselves whether there is a special mechanism for reasoning about beliefs that accounts for the differences between the false-belief and false-photograph versions of the task.

Neuroimaging. Functional neuroimaging purports to offer just such

a direct window on the mind's operations. It allows us to ask whether distinct brain regions in a healthy human adult are recruited just when the person must use attributed beliefs to reason about someone else's action.

The most popular technique is functional magnetic resonance imaging (fMRI). A basic MRI provides an amazingly fine-grained three-dimensional picture of the anatomy of soft tissues, like the grey and white matter (cell bodies and axons) of the brain, entirely invisible to X-rays. *Functional* MRI also gives the blood's oxygen content in each brain region, an indication of recent metabolic activity in the cells, and so an indirect measure of recent cell firing. The images produced by fMRI analyses show the brain regions in which the blood's oxygen content was significantly higher while the subject performed one task—a false-belief task, for example—than while the subject performed a different task—a false- photograph task.

So does the fMRI evidence show that distinct brain regions are recruited when people must, in the course of their reasoning, attribute beliefs to other people? It does.

Studies using fMRI adapt the false- belief and false-photograph tasks for adults. Subjects lying inside an fMRI scanner read short stories and answer questions about the stories. When subjects read stories in which the action hinges on a false belief, the oxygen content of the blood in a particular set of brain regions increases; when the same subjects read stories in which a physical representation—a photograph, drawing, or map—becomes false, the oxygen content in the same regions stays low. These brain regions include the temporo-parietal junction on both the left and the right—just above and behind your ear—and medial prefrontal cortex—a few centimeters back from the middle of your forehead.

An observed neural division of labor is not proof of distinct mental mechanisms, but additional evidence suggests that activity in the temporo-parietal junction and medial prefrontal cortex really does reflect the operation of a special mechanism dedicated to reasoning about beliefs.

First, in these experiments activity in these regions increases when, and only when, the storyline hinges on a character's beliefs. Stories about a person's physical appearance get no response from these brain regions. Even stories about or movies depicting simple goal-directed actions—a person moving toward a target—get very little

response. Concepts like goal and desire—the concepts that one- and two-year-olds use in the earliest stage of reasoning about the mind—are not the domain of these brain regions. At least with the stories that have been tested so far, activity in the temporo-parietal junction and the medial prefrontal cortex increases particularly and precisely for those stories that three-year-olds cannot figure out, but five-year-olds can.

Second, the same brain regions that are implicated in reasoning about beliefs in healthy adults show a reduced response in autistic adults. Fulvia Castelli and her colleagues have conducted experiments in which they show both healthy subjects and subjects with autism a set of short animations. In some of the animations, two triangles bounce around the screen, but in other animations the same triangles appear to interact with one another, chasing, harassing, and even flirting with each other. The two brain regions described above, the temporo-parietal junction and medial prefrontal cortex, were significantly more active during the interactive animations in the brains of healthy adults, but not in the brains of the autistic subjects.

These, then, are the three main prongs of the case that the human mind contains a special mechanism dedicated to inferring what other people are thinking, and to using this information to explain and predict human action: the same logical problem about a false representation, when phrased in terms of a person's belief, is easier to solve for four-year-olds, undergraduates, and patients with Alzheimer's, harder to solve for adolescents with autism, and recruits activity in a distinct set of brain regions, including the temporo-parietal junction and medial prefrontal cortex. This case for a belief module is far from unassailable, and indeed every one of these prongs is still vigorously disputed, but the whole picture is compelling.

Human Nature

So far, we have avoided the questions about whether the capacity to reason about other minds is innate, universal (common to all members of the human species), and specific to the human species. But the very idea of an evolved special mechanism of the mind implies that this mechanism is part of the human genetic endowment, universal within our species and possibly unique to it. So troubles with any of these three ideas may mean trouble for the idea of a mental faculty dedicated to reasoning about other minds. And each of these three is the subject of intense current debate.

Rather than try to do justice to the enormous range and subtlety of these debates, I will defend just the possibility that the capacity to use attributed beliefs to explain and predict behavior is innate, universal, and species-specific by answering three narrower questions: (1) How can a capacity be innate if that capacity only begins to operate three to five years after birth? (2) How can we say that reasoning about other minds is universal, when the very notion of a mind changes dramatically across cultures and across time? (3) Based on what evidence do we accord or deny to other species the ability to reason about other minds?

Innateness. The notion of innateness is subject to much confusion. The simplest misconception is the idea that if a capacity is innate, it must be present from birth. But consider two unambiguously innate processes, adolescence and language acquisition. Changes may happen in the body or the mind many years after birth and yet unfold according to a set genetic plan—like a man growing a beard. The innate tendency to acquire a rich grammar is so powerful that children can sometimes create a new language in a single generation. Adults brought into a community of migrants and immigrants with no shared language create a broken and oversimplified system for verbal communication called a "pidgin" language. The children of these adults, though, spontaneously learn to speak a grammatically much richer version of their parents' language, called a "creole." Creolization is striking evidence for the innate tendency to acquire a language. These examples show that there is no problem in principle with an innate mechanism whose operations are triggered long after birth.

Language and adolescence are also good models of another truism of development: the role of the environment. Both language and adolescence, though innate, are dramatically influenced by the surrounding world of the child. With adolescence, the environment affects mainly the timing of an otherwise very stereotyped pattern of physical change. In the acquisition of language, on the other hand, the surrounding culture determines the whole content and character of what is acquired, though the basic principles that guide language acquisition and make it possible are innate and universal. The moral is that an innate mechanism may be present at birth or emerge any time thereafter; the local environment may influence just the timing of its emergence or its entire specific content. There is no simple signature of innateness.

Cultural variation. Still, if the human genetic endowment is responsible for the change around age three that lets children start

using inferred beliefs to explain behavior, we would expect something recognizably like this change to happen at some age in every culture. What then are we to make of the claim that some languages lack even a word that corresponds to "think/believe"? Or that in some cultures, the reasons offered for behavior are never personal mental states, but social norms, duties and obligations? Some cultures are very unwilling to talk about personal beliefs, desires, and emotions, while others, like our own, are more or less obsessed with them.

Clearly, cultures diverge enormously in how they choose to describe the relationship between mind and world. But such diversity is not necessarily incompatible with the operation of a single universal innate mechanism for reasoning about other minds.

First, some cultural differences may reflect distinctions not in the underlying concepts of the mind, but in what are called "display rules"—that is, not in what people think but in what a culture considers acceptable for conversation. Sensitive topics, like sex, disease, and heresy, wax and wane in their perceived suitability for public consumption, but the concepts themselves may remain the same. The term "display rule" itself comes from a related tradition of cross-cultural study of emotional expression. Cultures vary wildly in the behaviors they deem appropriate at the death of a loved one, from breast-beating and clothes-rending to serenity and acceptance, but no culture fails to recognize the mourner's grief. Open talk about personal beliefs and desires could be subject, in some cultures, to similar censorship.

Second, cultural differences may be variations on a theme, emerging out of a common underlying concept. As children become acculturated adults, the underlying distinctions provided by the innate mechanism are refined and expanded and made more precise, but never wholly abandoned. The innate category "agent," for example—a being that can change the world in order to achieve its desires and goals—may include in different cultures not just embodied people, but also witches, spirits, or dead ancestors. This difference will lead to other differences in what counts as a satisfying explanation of a behavior. In the modern West, claiming that the person was possessed by a witch is not satisfying, unlike at other places and times. Still, all of these differences are arguably relatively superficial, adult refinements of the child's abstract but universal notions of mind, agent, and goal-directed action.

To make the case for universality, however, we need to do more

than simply deny that cultural distinctions always reflect real conceptual differences; we also need some positive evidence for the common fundamental concepts themselves. So far, in every culture that has been tested, children below the age of three or four have failed the false-belief task described at the beginning, and older children have passed. Many of the cultures tested have been industrialised and urban, but not all. Jeremy Aisop and Paul Harris developed a version of the false-belief task for children of the Baka, a group of pygmies living in the rain forests of Southeast Cameroon. Baka children aged three to six were asked to hide a piece of food that an adult would need for cooking, and then to predict the reactions of the adult when she returned: where she would look, how she would feel and so on. Like their Western counterparts, the Baka three-year-olds predicted that she would look for the food where it really was; the older children knew better. Reasoning about beliefs may therefore be a universal human achievement, and the underlying capacity may well be part of our intrinsic nature.

Uniqueness. Some theories go one step further and argue that the ability to reason about other minds is not only universal among human beings, but unique to them, part of what marks us off from our nearest evolutionary neighbors. Many peculiarly human skills—language acquisition and use, cultural transmission of knowledge, and Machiavellian deception and counter-deception—depend on our ability to figure out what another person is thinking, whether this knowledge is then used to forward co-operative or competitive ends. What's more, experiment after experiment has failed to provide clear evidence that even our nearest relatives, chimpanzees, reason about the contents of other minds. The experimenters are getting more ingenious, though, and the question of species specificity remains open.

One important recent innovation, made by Brian Hare and his colleagues, is to test a chimpanzee's ability to reason about the mind of another chimpanzee, and to do so in a competitive setting where some natural benefit follows from knowing what the other chimpanzee believes. In Hare's design, a chimpanzee relatively low down in his troop's dominance hierarchy watches while two pieces of highly desirable food are hidden. The critical manipulation is that while one of the pieces of food is hidden another more dominant chimpanzee is also allowed to watch, and the subordinate chimpanzee is allowed to watch the dominant chimp watching. The chimpanzees are then allowed into the stage. The rules of the dominance hierarchy forbid the low-down chimpanzee from taking

a piece of food before the dominant one gets to it. What Hare and his colleagues measured was whether the subordinate chimp was more likely to go for the piece of food that only he saw hidden, and that the dominant chimpanzee didn't know about, in order to increase to increase his chances of both getting food and avoiding conflict. And indeed, this is just what the subordinate chimpanzees did, as if they could successfully keep track of where the dominant *thinks* there is food.

This elegant experiment doesn't quite resolve the question, though. There is another way for the subordinate chimp to solve the competitive problem, one that depends only on certain behavioral associations, and not on ideas about beliefs at all. The subordinate could predict that the dominant will head in the direction he last looked, or that he will go first for the piece of food that he saw. Knowing about the behavioral relationship between looking or seeing and acting is undeniably an important prerequisite of the ability to reason about other minds, a critical part of stage one, the knowledge about minds that two-year-olds already have.

What is still missing is definitive evidence that any non-human animal has ever gone beyond stage one, to make the three-year-old's impressive transition into a world of beliefs: a transition that enables us able to predict one another's conduct, coordinate for the common good, and suffer the sorrows of Romeo and Juliet when we get things wrong. ■

Rebecca Saxe is a junior fellow of the Harvard Society of Fellows, having completed her Ph.D. in cognitive science at MIT under Nancy Kanwisher.

Originally published in the [February/March 2004](#) issue of *Boston Review*.



Copyright [Boston Review](#), 1993–2004. All rights reserved. Please do not reproduce without permission.

■ [home](#) ■ [new democracy forum](#) ■ [fiction, film, poetry](#) ■ [archives](#) ■ [masthead](#) ■ [subscribe](#) ■