HOW THE

Mind Works

STEVEN

PINKER

INTO THE CLASSROOM MEDIA

A VIDEO PRESENTATION FOR THE CLASSROOM

1. COMPUTATION

(. . . Romeo, Juliet, and intelligence.)

Steven Pinker begins his presentation by proposing that the mind is an engineering masterpiece, unparalleled by any artificial intelligence system.

Pinker illustrates that this is because the mundane mental activity that goes into activities like picking up a pencil or understanding a sentence involves formidable engineering tasks which we have not been able to duplicate in any man made system.

But the mind also has many apparent quirks. For example, why is the thought of eating worms disgusting? Why does the male of our species do insane deeds, like challenging others to duels? Or murdering their ex-wives? Why do people in all cultures believe in ghosts and spirits? Why do fools fall in love?

In order to understand why the mind is such an engineering masterpiece and why it is so quirky, Pinker will call upon three ideas: COMPUTATION, EVOLUTION, and SPECIALIZATION.

Pinker turns first to COMPUTATION. He argues that the primary function of the brain is information processing, and that this idea solves a two-fold problem; namely, what is intelligence and how can a "hunk of matter" (such as the brain) achieve it? Pinker then contrasts the attraction between Romeo and Juliet to the attraction between filings and a magnet in order to pinpoint intelligence as the pursuit of goals by inference; inference being a knowledge of logic, statistics, or cause and effect in the world.

For example, if the goal is to touch Juliette's lips, an inference capable of attaining that goal might be something like: If C is between A and B they cannot touch. If A goes over C, C is no longer between A and B. Therefore, to touch Juliette's lips, go over the wall.

In turn, he explains that in a computational system, goals and knowledge are represented as patterns of bits in matter in the system called "representations."

Having defined "intelligence," Pinker next proposes that the idea of COMPUTATION solves the problem of how beliefs and desires cause behavior.

For example, let's say you wanted to explain why Bill just got on a bus. You wouldn't have to put his head in a brain scanner. You wouldn't have to take samples and do DNA testing. The best answer can be obtained just by asking him. He would be likely to say something like "I want to visit my grandmother. I know the bus will take me there." No other explanation will do as good a job at predicting his behavior as that one. If Bill hated the sight of his grandmother, or if he knew that the route had changed, his body would not be on that bus.

That leads to the mind-body problem — namely, how do these colorless, odorless, tasteless, invisible little nothings called "beliefs" cause behavior in the way that one billiard ball clacking into another can cause a physical event to happen? I think the COMPUTATIONAL THEORY OF MIND, which says that ideas, knowledge, goals, and beliefs are implemented as information, as patterns in bits of matter, can dispel that paradox.

Here, Pinker comments on a common misconception about the COMPUTATIONAL THEORY OF MIND — namely, that it implies a computer is a good metaphor for the mind.

The idea is not that the commercially available computer is a good metaphor for the mind, but just that the explanation for what makes computers smart, and the explanation for what makes brains smart, is the same — namely, that they are both information processors.

An analogy would be that we invoke the same laws of aerodynamics to explain what keeps airplanes in the air and what keeps birds in the air. But it doesn't commit us to an airplane metaphor for the bird and cause us to check to see whether birds have complimentary beverage service, in-flight movies, and so on.

Finally, Pinker explains that the COMPUTATIONAL THEORY OF MIND holds that the lifeblood of the mind is information, running counter to our most common way of explaining the mind, according to which that lifeblood is energy or pressure.

For example, you are apt to hear the following kind of explanation of behavior in ordinary conversation or in the newspapers: "If only Fred had an outlet so he could let off steam, vent his hostility, and channel his rage rather than bottling it up, he wouldn't have exploded last Tuesday."

This is an example of the HYDRAULIC THEORY OF THE MIND. The idea that the mind is animated by some vessel of over-heated steam or pressure, which has to be properly channeled, lest it burst out and cause damage.

Now, there is no doubt that the hydraulic metaphor does capture something about human emotion. But that only leads to an interesting scientific question – namely, why is the brain going to so much trouble to simulate energy or pressure, given that we know that it doesn't literally work that way? That is a question that I will return to later.

Discussion Topics Related To This Section:

Pinker proposes that the idea of COMPUTATION holds the answer to the question of what intelligence is. Discuss his example of Romeo's attraction to Juliet, and explain his definition of intelligence within a computational framework.

Pinker argues that the idea of COMPUTATION solves the problem of how beliefs and desires cause behavior. Explain.

Pinker argues that the store-bought computer does not provide an appropriate metaphor for the mind. Why?

THE PUZZLE OF THE MIND

The human mind is an engineering masterpiece.
No robot can:

- See
- Move
- Speak and understand
- Use common sense as well as a person.

Intelligence is the pursuit of goals by inference (knowledge of logic, statistics, and cause-and-effect)

goal = "touch Juliet's lips"

BUT the mind has many apparent quirks, e.g.,

- Why is the thought of eating worms disgusting?
- Why do men challenge each other to duels and murder their ex-wives?
- Why do people believe in ghosts and spirits?
- Why do fools fall in love?

inference =

"If C is between A and B, they cannot touch.

If A goes over C, C is no longer between A & B.

Therefore, to touch Juliet's lips, go over wall."

IDEA #1. COMPUTATION.

The function of the brain is information-processing, or computation.

How a computational system works:

Goals & knowledge are information: they are represented as patterns in bits of matter in the system.

System is designed so that one representation causes another, and the changes mirror the laws of logical or statistical inference.

PROBLEM:

What is intelligence, and how can a hunk of matter (such as a brain) achieve it?

2. EVOLUTION AND SPECIALIZATION

(. . . minds, olive pitters, and other complex devices.)

In this section, Pinker moves to his next two key ideas: EVOLUTION and SPECIALIZATION. He begins with a simple question: How do we understand complex devices? Pinker then holds out an unfamiliar device, demonstrating its complexity and its parts. He argues that one could puzzle over it for a long time until someone revealed what its function was. In the case of the device at hand, it is an olive pitter.

Here, Pinker demonstrates the concept of "reverse-engineering," that is, beginning with a complex device and trying to uncover what it was designed to do.

In nature, Pinker explains that the only explanation we have for design (or for the illusion of design) is Darwin's Theory of Natural Selection. Therein lies the answer to the question of what the mind was designed to do.

Now, the first question you have to ask in reverse-engineering is "What is the function of the device?" Since, in the case of the mind, we appeal to Darwin's Theory of Natural Selection, we have an answer to that question — namely, the ultimate function of the mind is survival and reproduction in the environment in which the mind evolved. That is, the tribal, or foraging, or hunter-gatherer lifestyle, in which our ancestors spent the vast majority of our species' evolutionary history until the very recent invention of agriculture and civilizations, only 10,000 years ago.

Pinker now turns to his third key idea, SPECIALIZATION. Arguing that the brain is not a singular, homogeneous mass of "wonder tissue," Pinker asks us to consider that the mind has to solve many kinds of problems — from seeing in three dimensions, to moving arms and legs, to finding and securing mates. Because these are very different kinds of problems, Pinker asserts that the tools for solving them are bound to be different as well.

We know that specialization is ubiquitous in biology. The heart doesn't look like the kidney, because the heart is an organ for pumping blood, and the kidney is an organ for filtering it. Any arrangement of tissue that is good for one, is bound to be bad at the other. That's why we have both.

This SPECIALIZATION goes all the way down. Heart tissue is different from kidney tissue, heart cells are different from kidney cells, and so on. So the suggestion is that the mind, like the body, is organized into the mental systems, organs, tissues, and so on.

Putting his three ideas together, Pinker proposes that the mind is "a system of organs of COMPUTATION that allowed our ancestors to understand and outsmart objects, animals, plants and one another."

Discussion Topics Related To This Section:

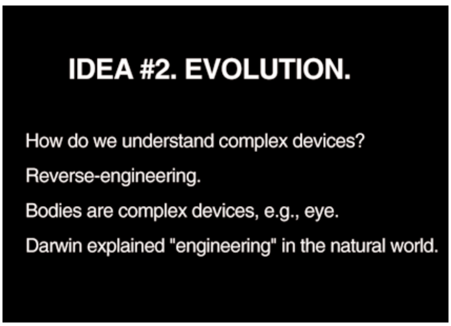
What is "reverse-engineering?"

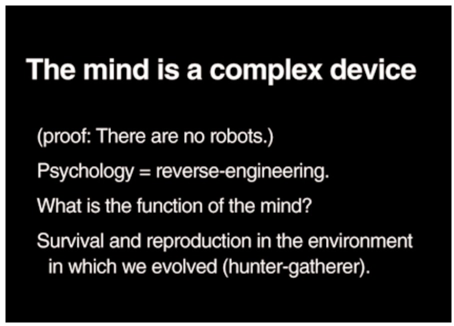
Why does Pinker appeal to Darwin's Theory of Natural Selection in order to explain what the mind was designed to do? How does natural selection create the illusion of design in nature?

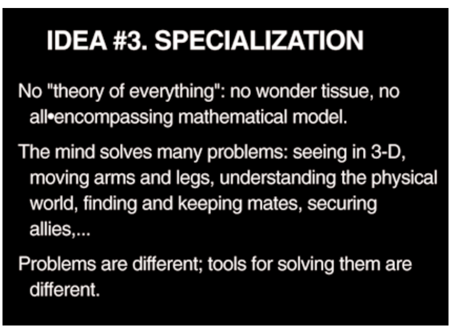
Explain the concept of SPECIALIZATION from biology.

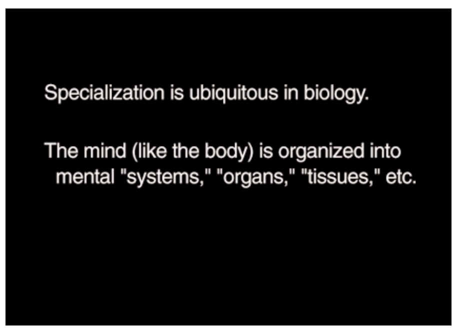
Pinker asserts that the mind, like the body itself, is specialized. Explain his reasoning.

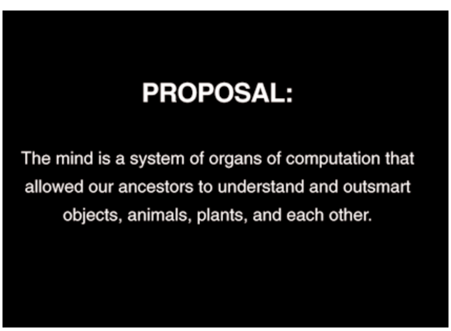
Pinker's proposal is that the mind is "a specialized system of organs of COMPUTATION that allowed our ancestors to understand and outsmart objects, animals, plants and one another." Using the ideas of COMPUTATION, EVOLUTION, and SPECIALIZATION, explain each component of this definition.











3. ADAPTATION AND MALADAPTATION

(. . . the problem of vision, and why we crave junk food.)

Having put forth his proposal for how the mind works, Pinker sets out to illustrate several examples. In this section, he turns to the problem of vision.

Pinker begins by asking what the brain is doing when we see. He demonstrates how, rather than a kind of video image, a brain's eye view of the world would look something like a massive spreadsheet with millions of numbers — each number representing the brightness of one spot on the visual field.

He then explains that the brain's job is to crunch those numbers. That is, to deduce what arrangements of three-dimensional objects in the world gave rise to that projection when light which was bounced off those objects was focused by the lens of the eye onto the retina, generating this massive spreadsheet.

The mind, Pinker explains, faces the additional problem of seeing in 3-D. Here, he illustrates one of the mechanisms by which the brain recovers three-dimensional shapes from a two-dimensional retinal image.

This trick is called "shape from shading." It relies on a basic law of physics, roughly speaking, the shallower the angle of a surface with respect to a light source, the less light it reflects back.

For example, if you hold a flashlight perpendicular to a piece of paper, it projects a nice, concentrated bright spot of light. But if you angle the paper with respect to the flashlight, the spot of light is smeared over a greater area, so any particular patch on the paper is bound to be dimmer.

Now, that has nothing to do with how the mind works, that's just how light works. But the mind has evolved a trick for exploiting that law of physics by, in a sense, running it backwards. It assumes that the dimmer the patch on a retina, the shallower the angle on the surface. In that way, the brain can reconstruct a shape from the angles of thousands of bits of surface, each of them deduced from the brightness value of that spot.

That is, for example, how you can tell the difference between a ping pong ball and a white poker chip. The ping pong ball has a complex pattern of shading, which the brain perceives as a spherical shape.

There is, however, a limitation to this trick, as Pinker next reveals.

It assumes that the world has a uniform pattern of pigmentation or coloration across the surface. It needs that assumption, because if there was a non-random or non uniform pattern of pigmentation, then it would be incorrect to deduce that any difference in brightness between one region and another comes from a difference in angle. It would simply come from there being more ink or coloration in one spot than another. That predicts that surfaces that are colored in clever ways should fool the shape-from-shading module into seeing things that aren't there.

In turn, Pinker explains the optical illusion involved with watching television. Because it is engineered to deliver a highly non-uniform pattern of shading across the surface of the glass, the brain (sticking with its assumption of uniform shading) interprets differences in lightness as differences in orientation. And so, we hallucinate a three-dimensional world behind the pane of glass.

Another example is make-up. People who are skilled at applying make-up know that if you put some blush on the sides of the nose, the eye of the beholder will perceive dimmer patch as a steeper surface. That causes the sides of the nose to appear more parallel, and it makes the nose look smaller.

Conversely, if you put some light powder on the upper lip, the eye of the beholder will interpret a brighter patch as a more perpendicular surface. That gives rise to that full lipped, pouty look that the super models all strive for.

Pinker uses these two examples to make a more general point. He explains that many apparent illusions, fallacies, and maladaptive behaviors come not from an inherent design defect in the mind, but from a mismatch between assumptions about an ancestral world built into our mental faculties (such as, that surfaces are uniformly shaded) and the structure of the current world.

I think this idea can make sense of a number of puzzles that have bothered evolutionary biologists about our species, such as; Why do we eat ourselves into an early grave with too much junk food. Why do we use contraception, which when you think of it is a kind of Darwinian suicide — you're taking steps to prevent your genes from propagating, rather than fostering them. Why do we gamble in casinos or buy state lottery tickets, sometimes referred to as the stupidity tax (because if the House makes a profit, the players, on average, must lose.)

Well, I think the resolution of these paradoxes is that our mental faculties make certain assumptions about the world: f or example, that sweet foods are nutritious — which was true of the world we evolved in, where the only source of sweet foods was ripe fruit (which is nutritious), and not true now that we have the technology to mass produce concentrated sources of sweet foods.

Furthermore, our mental faculties assume a world in which sex leads to babies, which again, was true until the very recent invention of reliable contraception — so that by installing sexual desire into organisms, the creation of babies took care of itself.

And finally, they assume a world in which statistical patterns have underlying causes. which This is true of virtually every part of the world, except for machines that have been specifically engineered to display statistical patterns without underlying causes (namely roulette wheels, slot machines, and other gambling devices.)

Discussion Topics Related To This Section:

Using the idea of "information processing," discuss Pinker's explanation of how the mind "sees."

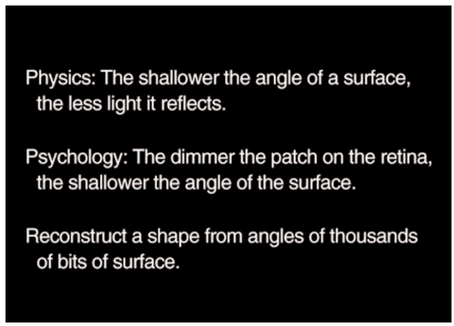
Pinker illustrates one method, called "shape from shading," by which the mind is able to recover three-dimensional shapes from a two-dimensional retinal image. Explain "shape from shading."

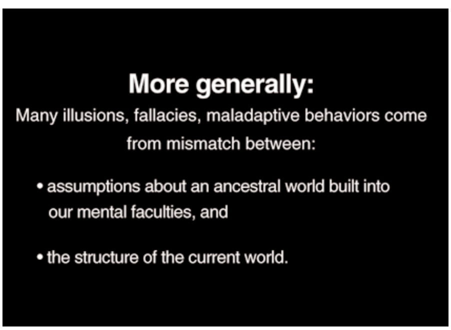
What assumption does the mind need to make in order to determine "shape from shading?"

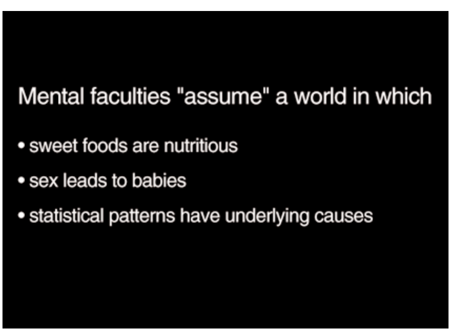
Pinker reveals that there is a limitation to the trick of determining "shape from shading," as illustrated by the optical illusions involved with television and cosmetic make-up. Explain.

What is the mismatching that Pinker refers to underlying these optical illusions?

How does this mismatching relate to other apparent maladaptive behaviors like eating junk food or gambling?







4. THINKING: THE COGNITIVE NICHE

(. . . babies, billiard balls, and "intuitive theories.")

Turning to "thinking," Pinker begins with the puzzle of what our ancestors did with their capacity for abstract intelligence.

Hunter-gatherers have all of the mental equipment necessary to do calculus, chess, law, and physics, given proper training. But in a hunter-gatherer lifestyle, there were no opportunities for those talents to be exercised, let alone be translated into a larger number of babies. So how can we explain what makes hunter-gatherers, including our ancestors, so smart?

In order to answer this question, Pinker illustrates the concept of a "biological arms race."

In nature, all organisms evolve at each other's expense . . . [and therefore] all organisms develop defenses against being eaten: animals run away, or they bite, or they develop armor or hide. Plants can't very well defend themselves by their behavior, so they resort to chemical warfare. Most plants have evolved a potent array of irritants, toxins, and bitter-tasting substances to deter creatures like us, who have designs on their flesh.

In nature, whenever you have an offensive weapon you have a defensive weapon evolving to counteract it, which leads to better offensive weapons and better defensive weapons and so on — it's the biological equivalent of an arms race.

In turn, Pinker argues that humans, using their capacity for abstract intelligence, entered this arms race with a decided advantage. By entering the "cognitive niche," humans were able to develop offensive weapons and defensive shields in the time span of a life-time, rather than in evolutionary time. That is, by understanding how the world works, humans were able to design techniques to penetrate the defenses of other plants and animals faster than they could evolve new defenses in return.

In all cultures, including those misleadingly called technologically primitive, there is an impressive array of tools, weapons, traps, snares, corals, poisons, and so on. There are ways of detoxifying plants by fermenting, cooking, leaching, and treating with other products. There is also cooperative action; the ability to bring about an effect that could not be brought about by one person acting alone by coordinating the behavior of several people.

Pinker then argues that, since the world is not a heterogeneous place, we have evolved a number of intuitive theories (each based on a core intuition) of how different parts of the world work.

The first and most basic, he suggests, is an intuitive version of physics, an understanding of how objects fall, roll, and bounce. Here, the core intuition is of stable, law abiding objects that have to pass through continuous space, and that continue to exist even when you don't look at them.

Pinker then points to how this core intuition can be demonstrated within the perceptual world of infants.

In order to do so, you rig up a world that disobeys the laws of physics by using cheap magic tricks — trap doors, mirrors, hidden compartments, and so on. Research shows that babies are visibly surprised when an object goes out of existence, or passes through the space belonging to another object, or goes from A to B without passing through all of the locations in between.

Pinker now turns to the idea of an "intuitive biology." Not all objects, he explains, seem to obey the laws of physics. Citing Richard Dawkins, Pinker points out that, if one were to throw a dead bird in the air, its trajectory would describe a graceful parabola and come to rest on the ground — just as it should, according to any physics text-book. But if one were to throw a live bird in the air, it would not describe a graceful parabola and come to rest on the ground. In fact, "it might not land this side of the county boundary."

Pinker points out that we don't view these seemingly law-defying objects as some kind of phenomenon. Instead, we interpret them as following a different kind of law — the laws of an intuitive biology.

Here, the core intuition is that living things contain an internal essence that supplies them with the renewable source of energy or oomph, that gives them their form, and that drives their growth and bodily functions.

I think that this is an explanation for how our species, everywhere, came to be such excellent intuitive biologists. When you go to any remote tribe, they will have hundreds of names for the local flora and fauna, which almost always correspond to the professional biologist's taxonomy of the genus or the species.

In turn, hunter-gatherers do a remarkable job of exploiting knowledge of how plants and animals work. From a few scratches on the ground they might infer the species and condition of the animal that left the track, and deduce where it's going so they can ambush it. They can remember the location of a flower in the spring and return to it the next fall to dig up the underground tuber that has grown in the interim. They also [experiment with] the juices and powders of plants and animals for use as medicines, poisons, and food additives.

Pinker now turns to our core intuitions regarding objects of our creation — namely, tools and artifacts. Using the example of a chair, he asks what all members of the chair category have in common. Here, he argues that we can't rely on physical shape for an answer, since chairs can assume any number of forms.

Instead, the only thing chairs have in common is that all of them are designed to hold up a human behind. The core intuition behind an artifact category, therefore, is its function.

Turning to the last of his four major "ways of knowing," Pinker argues for a kind of intuitive psychology.

When we interpret one another's behavior, we don't think of each other as some kind of mechanical wind-up doll. Rather, we impute beliefs and desires to one another. We assume that other people have minds. That is a kind of intuitive psychology we exhibit, for example, when explaining [from PART 1] why Bill got on the bus in terms of his desire to visit his grandmother and his knowledge that the bus would take him there.

Pinker now presents an array of evidence for these different intuitive theories, or "ways of knowing."

First, he points to how they appear in infancy.

At a very young age, babies make distinctions, for example, between living things and other moving objects.

If you show a baby a billiard ball clacking into another, very soon they will stop looking and get bored. But if you show them one billiard ball stopping short, and then the other one getting up and leaving, that really grabs their attention. But if you substitute people for billiard balls, you get the opposite intuition. Babies are not surprised if one person stops short and another one walks away. What surprises them is a collision.

As further evidence, Pinker points to cases of neurological disorder in which these different ways of knowing fractionate. For example, if a person suffers damage to certain parts of the brain because of a stroke, they may lose the ability to name living things like fruits, vegetables, and animals, but still may be able to name man-made objects, like tools and furniture. With damage to a different part of the brain, the opposite pattern might be observed, where the person can no longer name artifacts, but can still name living things.

The suggestion is that knowledge of tools and living things are stored in different ways in the brain.

Finally, Pinker offers one more piece of evidence from the misapplication of our intuitive psychology.

Perhaps the most dramatic evidence for the differences amongst these parts of psychology, comes from "misapplication" — cases where we take a part of the mind that evolved to mesh with one part of the world, and apply it to some other part that it wasn't designed for.

An example would be the belief in souls, ghosts, and spirits. This is an example of our intuitive psychology running overtime — interpreting minds that have no bodies, and are just floating free of the physical world.

Animistic beliefs do the opposite. They illicitly marry our intuitive psychology to our intuitive physics, biology or artifacts, and cause us to attribute minds to things that don't have them, such as trees, mountains, or idols.

Discussion Topics Related To This Section:

Pinker comments that "all organisms evolve at each other's expense." In turn, he describes what has been referred to as a "biological arms race." Explain.

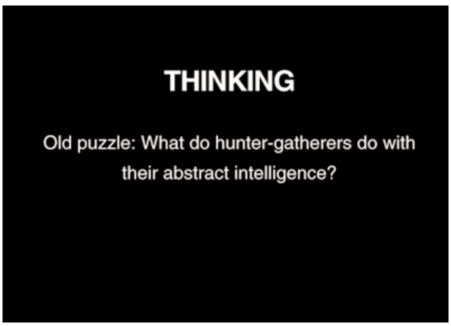
How does Pinker argue that our ancestors used their capacity for abstract intelligence? What evidence does he cite?

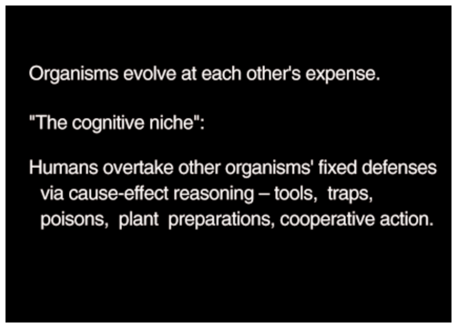
Pinker asserts that humans have evolved a number of “intuitive theories” — the most basic of which is an intuitive version of physics. Characterize this “intuitive physics” and discuss the evidence Pinker proposes from research on infants.

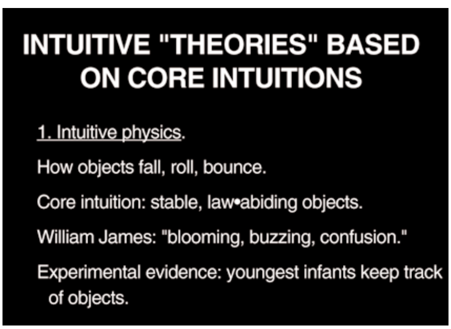
Explain the core intuition underlying our “intuitive biology.”

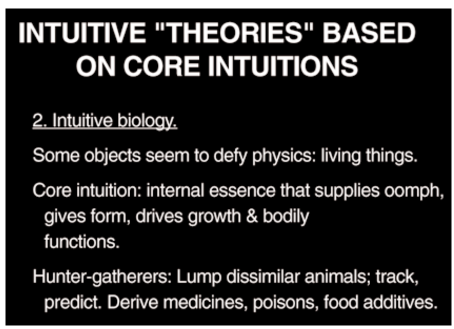
How does Pinker assert that humans exhibit an “intuitive psychology?”

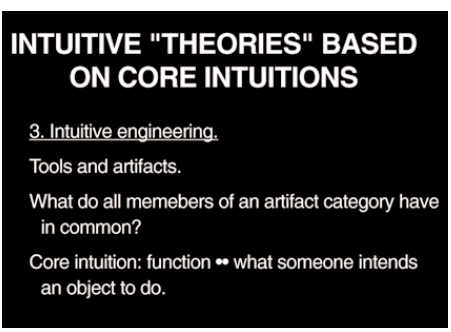
Pinker presents an array of evidence for these different intuitive theories. Discuss his evidence from both research on infants and cases of neurological disorder. What further evidence does Pinker argue is illustrated by our belief in ghosts and spirits?

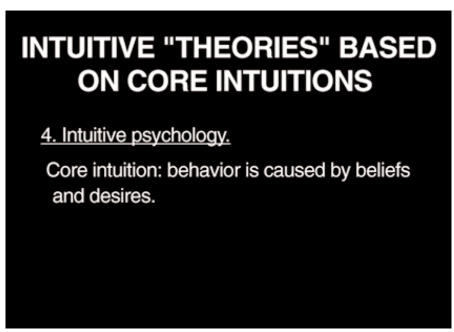












5. EMOTIONS ABOUT THINGS

(. . . reverse-engineering "disgust.")

In this section, Pinker reverse-engineers the emotion "disgust" as an example of our emotions about things.

Disgust, he explains, is triggered by certain objects in the world — as demonstrated by the following lyrics from a childhood camp song:

"Great green gobs of greasy, grimy, gopher guts, mutilated monkey meat, concentrated chicken feet, jars and jars of petrified porpoise pus, and I forgot my spoon."

What these (and similar) lyrics have in common, is that they names parts or products of animals. Accounting for cultural variations, Pinker asserts a general pattern, in that there are some animal substances that are always taboo and revolting. In turn, Pinker asks to consider whether disgust is rational: "After all, why discard useful potential sources of protein in the animal kingdom?"

However, the answer to that question may lie in the dilemma faced by omnivores: on the one hand, we are capable of handling a wide variety of plant and animal foods, on the other, some unknown subset of them are bound to be poisonous. Additionally, even ordinarily safe foods can become dangerous if they are subject to spoilage or other contamination.

While this subset will vary from environment to environment and will have to be learned, the suggestion is that disgust is a kind of deterrent against poisoning from contaminated animal products.

. . . that basically, we use our friends and parents the way kings used to use food tasters — namely, if they ate a bite and didn't keel over dead, then it's probably okay to eat. All other animal products are guilty until proven innocent.

Revealing another interesting aspect of disgust, Pinker points out how children's belief in "cooties" reflects a similar intuitive rationale.

Cooties, . . . as in "I don't want to drink your Coke. It's got your cooties on it."

The thing is, children are right. There are cooties. They're called micro-organisms. Micro-organisms have an interesting property that differentiates them from chemical contaminants, such as those found in plants — that is, germs multiply. What starts out as an undetectable trace, can quickly become twice as plentiful, then four times, then eight, then sixteen times, and so on.

So the intuition that an undetectable trace left behind by contact can saturate a substance of any size has some biological warrant, and suggests that disgust is a kind of intuitive microbiology.

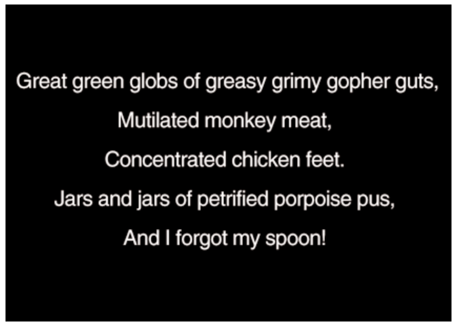
Discussion Topics Related To This Section:

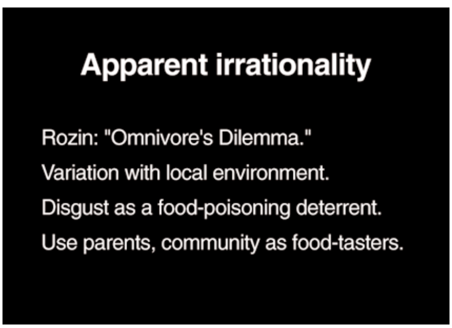
Why does "disgust" seem irrational at first glance?

What is the "omnivore's dilemma?"

What does Pinker reveal about disgust by reverse-engineering it?

Pinker argues that children's belief in "cooties" reflects a kind of intuitive microbiology. Explain.





6. EMOTIONS ABOUT PEOPLE

(. . . love, passion, and other apparent irrationalities.)

Pinker introduces this section with the question of why our emotions about others are so often passionate and seemingly irrational. He poses examples such as:

Pursuing vengeance until the day you die. Vowing not to rest until you slay the guy whose great-grandfather slew your great-grandfather. Defending your honor at all costs, by challenging someone to a duel if he besmirches your reputation with an unkind remark. Stabbing the guy who disses your brand of sneakers. Falling head-over-heels in love, and so on.

The standard account for passion, Pinker points out, is the romantic theory. This is the idea that all of us house a primal force which is part of our nature, and that is fundamentally irrational and maladaptive unless channeled into art and creativity.

Pinker proposes an alternative account of passion called the "strategic theory" rooted in the mathematical theory of games and the notion of "paradoxical tactics". By reverse-engineering romantic love and passionate vengeance, Pinker will demonstrate why a sacrifice of freedom and rational decision-making can actually be advantageous in situations of promises, threats, and bargains.

Turning first to "love," Pinker illustrates the idea that the realm of love is a kind of marketplace where we search for the most desirable person who will have us. In other words, love involves a certain element of smart shopping.

All of us at some point in our lives have been in search of the nicest, best looking, funniest, smartest, kindest person who will settle for us.

But your ideal match is a needle in a haystack, and you might die single if you wait forever for him or her to show up. So all of us have to trade off value against time, and at some point set up house with the best person that has come along so far.

This idea seems evidenced by a phenomenon referred to by social psychologists as "assortative mating by mate value," whereby, on average, partners within a couple tend to be closely matched in overall desirability as judged by third parties.

Yet, this economic analysis of love seems to miss something fundamental to the experience.

. . . [namely,] the involuntariness. The fact that you can't will yourself to fall in love with someone. The caprice. The fact that you can't predict exactly who will end up with whom.

[I'm sure many of you] have been in a situation in which you have been fixed up with someone who looks perfect on paper. You can tick off all of the desirable properties. But then when you got together, somehow sparks didn't fly. Cupid didn't strike. The earth didn't move. Why not?

Is this, he asks, any way to build an intelligent organism? Paradoxically, it just might be because of a particular problem which economists call "the commitment problem."

Romance, Pinker explains, is a kind of promise; to forgo opportunities to be with other people, to have children together, and so on. But there is a problem inherent to a promise — namely, how do you make it credible in the eyes of the person to whom you're making the promise, particularly when it may be to your advantage to break the promise later on?

In the case of love, the law of averages dictates that someone more desirable will surely come along at some point. And at that point, a hypothetical rational actor, following the narrow "smart shopping' strategy, would "dump their mate like a hot potato."

But in this hypothetical world of rational agents, the partner can anticipate that that day would come. Therefore, they would have been crazy to have entered the relationship to begin with — knowing that it would happen sooner or later — because there are certain irretrievable costs that are involved in starting a relationship, such as the opportunity costs of being with other people . . . [not to mention] giving up your apartment, selling your stereo, and so on.

So you have a situation that ought to be to the advantage of both parties, yet neither party can agree to it because neither has grounds for trusting the other.

Pinker cites economist Robert Frank, suggesting that one solution to this problem may be that if you don't choose to fall in love for rational reasons, than you can't choose to fall out of love for rational reasons. Thus, the very irrationality of romantic love acts as the guarantor of the implicit promise that is being made.

As evidence for this analysis, Pinker argues that romantic love is a universal human emotion; that something approximating our own notion of romantic love can be found in all the world's cultures.

Additionally, he cites the "logic of courtship:"

Another bit of evidence comes from the logic of courtship — that is, what you say or do to try to make someone fall in love with you.

If you were to whisper into your lover's ear "You're the nicest, funniest, best looking, richest person I've been able to find so far" it would probably kill the romantic mood. The way to a person's heart is to declare the exact opposite: "I can't help the fact that I'm in love. I'm attracted to you, the unique, quirky, irreplaceable individual, and I have no control over that emotion. I can't help falling in love with you. I like the way you walk. I like the way you talk. I want you so bad it's driving me mad," etc., etc., etc.

Finally, Pinker proposes that the theory explains a puzzle pointed out by William James and others, which is that the "emotions tie up the body as much as the mind."

When we're in the throws of passion, we show it. We tremble, we blush, we sweat. We blanch. Our voice croaks. We get goofy expressions on our face, and so on. One interpretation is that we're broadcasting the fact that our current course of action is under the control of the involuntary division of the central nervous system, the limbic system, which is in charge of housekeeping and physical plant functions, like heart rate and blood circulation — and that it is not under the control of the rational decision making division of the central nervous system, *the cortex and the frontal lobes; that since it is not a course of action that we have talked ourselves into, it is not a course of action we're likely to talk ourselves out of.*

If passionate love and loyalty are guarantors of our promises, then by symmetrical logic, Pinker now proposes that passionate vengeance and honor is a guarantor that our threats are not bluffs.

Humans, he explains, all rely on a set of implicit threats to deter possible antagonists. Yet, the inherent problem with a threat is that it may not actually be to one's advantage to back it up; in other words, it's real value is only as a deterrent.

Once it's time to carry it out, there is nothing in it for you. You could get hurt if you actually back up your words and try to beat someone up. Since the target of your threat knows that, they can threaten you right back by defying the threat — as they say, calling your bluff — because there is nothing to be gained when it comes to the point of enforcing it.

Well, one way around this is to be so constituted that it would be an intolerable insult to be crossed or insulted, so that you can't help pursuing vengeance even if it hurts you in the short term. The long-term advantage is that it makes you the kind of person that other people don't want to mess with.

This logic is often elegantly demonstrated in fiction and popular culture, as illustrated by a scene between the characters Casper Gutman (Sidney Greenstreet) and Sam Spade (Humphrey Bogart) from the film *The Maltese Falcon.*

Sam Spade dares Gutman to kill him, knowing that Gutman needs him to retrieve the falcon. Gutman says to Spade:

"That's an attitude, sir, that calls for the most delicate judgment on both sides. Because as you know sir, in the heat of action, men are likely to forget where their best interests lie, and let their emotions carry them away."

Discussion Topics Related To This Section:

Using the concept of "assortative mating by mate value," explain how love is a kind of marketplace. What does this calculating, economic analysis seem to miss according to Pinker?

What is the "commitment problem" inherent to promises? Apply the "commitment problem" to the realm of love.

Explain Pinker's proposal that the "apparent irrationality" of romantic love may hold the solution to this commitment problem.

Discuss Pinker's evidence for his analysis of romantic love.

If passionate love is a guarantor of our promises, then by symmetrical logic, Pinker argues that passionate vengeance and honor is a guarantor that our threats are not bluffs. Discuss Pinker's reverse-engineering of passionate vengeance.

EMOTIONS ABOUT PEOPLE

Why are our emotions about people passionate and seemingly irrational?

Pursuing vengeance; defending honor; falling head-over-heels in love.

Strategic Theory:

Paradoxical tactics. Sacrifice of freedom and rationality give an advantage in promises, threats, bargaining.

REVERSE-ENGINEERING ROMANTIC LOVE

The rational part of love: smart shopping.

Value versus time.

Assortative mating.

The irrational part of love: involuntariness, caprice.

Problem for the rational strategy: commitment problem.

Romance is a promise.

A rational agent might want to break the promise.

The promise is not credible.

Solution:

If you don't decide to fall in love for rational reasons, you can't decide to fall out of love for rational reasons.

Romantic love as guarantor of promise.

Some evidence:

- Universality of romantic love.
- Courtship strategies: Orientation to individual; involuntariness; costs.
- Involuntary signals.

Passionate love and loyalty are guarantors that promises are not double-crosses.

Passionate vengeance, "honor" is a guarantor that threats are not bluffs.

7. AN OPTIMISTIC VIEW

(. . . even if evolution isn't guaranteed to produce niceness.)

In this concluding section, Pinker acknowledges that most of us would rather not view ourselves as a system of computers, designed by natural selection, to promote survival and reproduction.

At the same time, Pinker maintains that there are some undeniable facts behind this view. He asserts that no scientifically literate person can deny that the mind is a product of the brain, that the brain is a product of evolution, and that evolution is not guaranteed to produce "niceness."

On the other hand, Pinker finds reason for optimism.

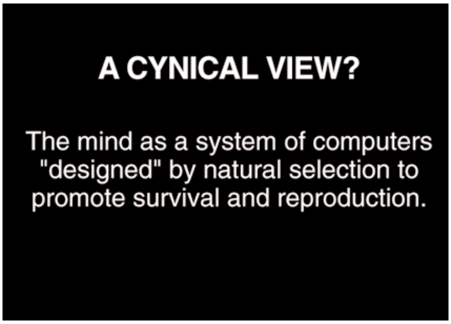
The idea of COMPUTATION suggests that the human mind is not just a bundle of crude drives and reflexes, but is composed of intricate, ingenious, and powerful software.

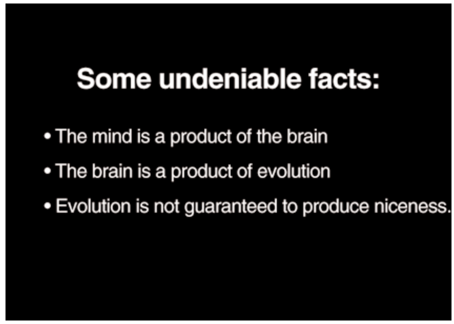
The idea of EVOLUTION suggests that our legacy from the natural world is not just the nasty, brutish emotions, like greed, aggression, lust, a thirst for blood, a territorial imperative, and so on, but that the kinder, gentler emotions, like love, friendship, and a sense of justice, are every bit as much a part of our legacy from natural selection.

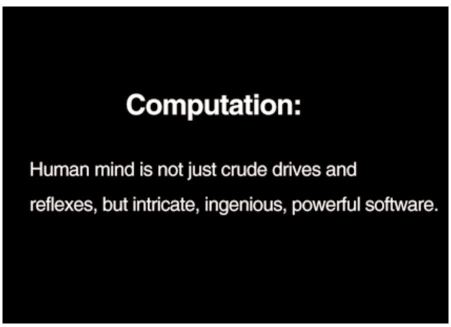
Finally, the idea of SPECIALIZATION — that the mind is composed of many interacting parts — holds out the hope that some parts of the mind — those with the longest view of the future — can figure out ways of out-smarting the other parts.

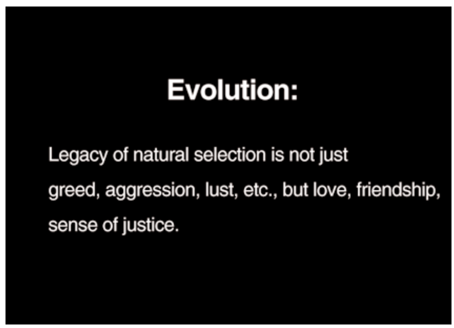
Discussion Topics Related To This Section:

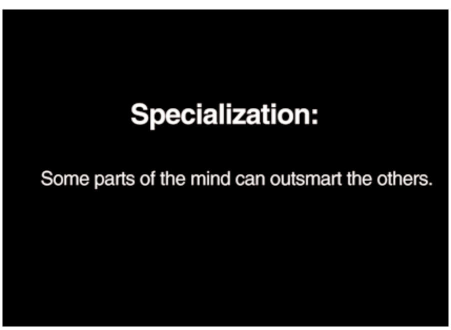
While evolution is not guaranteed to produce niceness, Pinker finds reason to be optimistic about the nature of the human mind. Using the three ideas of COMPUTATION, EVOLUTION, and SPECIALIZATION. Explain.











COULD A COMPUTER EVER BE CONSCIOUS?

Steven Pinker

Steven Pinker is Professor and Director of the Center for Cognitive Neuroscience of the Massachusetts Institute of Technology and author of *The Language Instinct*. This article is adapted from his forthcoming book *How the Mind Work*s (Norton, October).

In one of the first episodes of *The Twilight Zone*, a man named James Corry is serving a fifty-year sentence in solitary confinement on a barren asteroid. Allenby, the captain of a supply ship, takes pity on him and leaves behind a crate containing Alicia, a robot that looks and acts like a woman. Corry, of course, soon falls deeply in love. A year later, Allenby returns with the news that Corry has been pardoned and that he has come to get him and a maximum of fifteen pounds of gear. Alicia, unfortunately, weighs more than that. When Corry refuses to leave, Allenby shoots Alicia in the face, exposing a tangle of smoking wires. He tells a devastated Corry, "All you're leaving behind is loneliness." (Note 1)

The horrifying climax raises two vexing questions. Could a mechanical device ever duplicate human intelligence — the ultimate test being whether it could cause a human being to fall in love with it? And if a human-like machine could be built, would it actually be conscious; would dismantling it be the snuffing out of a sentient being?

Pose the first question to experts in Artificial Intelligence and you'll get one of two answers: 1) life-like robots are just around the corner, or 2) it will never happen. (Note 2) Don't believe either one. These are the kinds of "experts" who a few decades ago predicted that nuclear-powered vacuum cleaners were in our future and that man would never reach the moon. (Note 3) Certainly computers will continue to get smarter, as the recent defeat of the world chess champion, Gary Kasparov, by IBM's Deep Blue reminds us. Today's computers can converse in English on restricted topics, control mechanical arms that weld and spray-paint, and duplicate human expertise in dozens of areas, from prescribing drugs to diagnosing equipment breakdowns. Artificial Intelligence has jumped from the laboratory to everyday life. Most people today have had their speech recognized by telephone directory assistance systems, and many have used intelligent search engines on the World Wide Web, own appliances controlled by fuzzy logic chips, or hold mutual fund portfolios selected by artificial neural networks. (Note 4)

Still, today's computers are not even close to a four-year-old human in their ability to see, talk, move, or use common sense. One reason for this disparity is sheer computing power. It has been estimated that the information processing capacity of even the most powerful supercomputer is equal to the nervous system of a snail — a tiny fraction of the power available to the supercomputer inside the bloated human skull.(Note 5) The kinds of processing are different, too. Computers find it easy to remember a twenty-five-digit number, but difficult to summarize the gist of Little Red Riding Hood; humans find it hard to remember the number but easy to summarize the story. One reason for this difference is that computers have a single, reliable processor (or a small number of them) working very, very fast; the brain's processors are slower and noisier, but there are hundreds of billions of them, each connected to thousands of others. That allows the human brain to recognize complicated patterns in an instant, whereas computers have to reason out every niggling detail one step at a time. Human brains also have the advantage of sitting inside human beings, and can soak up terabytes of information over the years as the humans interact with with other humans and with the environment. Brains also have the benefit of a billion-year R&D effort in which evolution equipped them with cheat sheets for figuring out how to outmaneuver objects, plants, animals, and other humans.

So how well will tomorrow's machines perform? Technological progress is notoriously unpredictable. When it comes to replacement parts for the body, who knew that artificial hips would become commonplace and artificial hearts elusive? When it comes to the performance of duplicates of the mind, the most reasonable answer is that computers will probably do a lot better than they do now for some kinds of thinking, and they will probably not do as well as a human being for other kinds.

But let's return to science fiction and assume that someday we really will have Alicia-class robots. Will they be "conscious"? It all depends on what you mean by the word. Woody Allen once wrote a hypothetical course catalogue with a listing for Introductory Psychology that read, "Special consideration is given to a study of consciousness as opposed to unconsciousness, with many helpful hints on how to remain conscious." (Note 6) We laugh because we realize that the word "consciousness" has at least two meanings. (Note 7)

One meaning is illustrated by Freud's famous distinction between the conscious and unconscious mind. I ask, "A penny for your thoughts?" You reply by telling me the content of your daydreams, your plans for the day, your aches and itches, and the colors, shapes, and sounds you perceive. But you cannot tell me about the enzymes secreted by your stomach, the current settings of your heart and breathing rate, the projections on your retinas, the rules of syntax that order words as you speak, or the sequence of muscle contractions that allow you to pick up a glass. This shows that information processing in the nervous system falls into two pools. One pool can be accessed by the brain modules behind verbal reports, rational thought, and deliberate decision-making. The other pool, which includes gut responses, the brain's calculations for vision, language, and movement, and repressed desires or memories (if there are any), cannot be accessed by those modules. Sometimes information can pass from one pool to the other. For instance, when we first learn how to use a stick shift, every motion has to be thought out, but with practice the skill becomes automatic (conscious processes becomes unconscious). With intense concentration and biofeedback, we can focus on a hidden sensation like our heartbeat (unconscious processes become conscious).

Will computers ever become conscious in this sense of access to a subset of the information in the whole system? In a way, they already are. The operating system of your computer is designed so that certain kinds of information are available to the programmer or user — opening and saving files, sending messages to the printer, displaying directories — and others kinds are not — such as the movements of the disk drive head or the codes sent by the keyboard. That's because any information system, computer or brain, has to work in real time. A device in which every morsel of information has to be easily available at all times to every process would be perpetually lost in thought. It would have to calculate whether the price of tea in China was relevant to which foot should be put in front of the other one next. Only some kinds of information are relevant to what the system is doing at a given time, and only that information should be routed in to the system's main processors. Even robots of the future, with their thousands of processors, will need some kind of control system that limits what goes into and out of the individual processors. Otherwise the whole robot would lurch and zigzag as the processors fight for control, like Steve Martin in *All of Me* when his right side was controlled by the ghost of Lily Tomlin. So in that sense, computers, now and in the future, are built with a distinction between "conscious" and "unconscious" processing. (Note 8)

But it's a very different sense of the word consciousness that people find particularly fascinating. That sense is sentience]: pure being, subjective experience, raw feelings, first person present tense, what it is like to see red or feel pain or taste salt. When asked to define consciousness in this sense, we have no better answer than Louis Armstrong's when a reporter asked him to define jazz: "Lady, if you have to ask, you'll never know." (Note 9)

How can we ever know whether Alicia is conscious in this sense — whether there's "anyone home" seeing the world through her camera-eyes and feeling the signals from her pressure sensors? No matter how smart she acts, no matter how responsive, no matter how vehemently she says she is conscious, an Allenby can always insist that she's just a very fancy stimulus-response machine programmed to act as if she were sentient. Try as hard as you like, you will not come up with an experimental test that will refute him.

Perhaps it is some consolation to know that our befuddlement here is not just a technological puzzle but is of a piece with some of the deepest problems in philosophy. If I can't know whether Alicia is sentient, how can I know whether [you] are sentient? I [think] you are, and I'm not so sure about Alicia, but maybe I'm just chauvinistic about creatures that are made out of meat rather than metal. How can I be so confident that consciousness is secreted by the brain tissue in my skull, rather than lurking in the software that my brain is running — software that Alicia's computer could run just as well? (Note 10)

Lest you think that the answer is obvious, ponder these thought experiments. Suppose surgeons replaced one of your hundred billion neurons with a microchip. Presumably you would feel and behave exactly as before. Then they replace a second one, and a third one, and so on, until more and more of your brain becomes silicon. The chips do what the neurons did, so your behavior and memory never change. Do you even notice the difference? Does it feel like dying? Is some other conscious entity moving in with you? Suppose that the transporter in *Star Trek* works as follows. It scans in a blueprint of Kirk's body, destroying it in the process, and assembles an exact duplicate out of new molecules on the planet below. When Kirk is beamed down, is he taking a nap or committing suicide?

The head spins in confusion; it's hard to imagine what a satisfying answer to these questions would even look like. But they are not just brain-teasers for late-night college dorm-room bull sessions. The imponderables also drive our intuitions about right and wrong. Was Allenby guilty of destruction of property, or of murder? Does a newborn boy feel pain when he is circumcised, or is his crying just a reflex? What about a lobster boiled alive, or a worm impaled on a fishhook?

These problems won't be solved any time soon, so don't expect someone to tell you with certainty whether a computer will ever be sentient. Perhaps it is a meaningless question, and we have been deluded by misleading verbiage into taking it seriously. Perhaps some unborn genius will have a thunderbolt of insight and we will slap our foreheads and wonder why the problem took so long to be solved. But perhaps the problem never will be solved. Perhaps the human mind, a mere product of the evolution of one species on this planet, is biologically incapable of understanding the solution. If so, our invention, the computer, would present us with the ultimate tease. Never mind whether a computer can be conscious; our own consciousness, the most obvious thing there is, may be forever beyond our conceptual grasp. (Note 11)

Notes

(1) Zicree, M. S. 1989. *The Twilight Zone Companion*. 2d ed. Hollywood: Silman-James Press.

(2) Crevier, D. 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence.* New York: Basic Books.

(3) Cerf, C., & Navasky, V. 1984. *The Experts Speak.* New York: Pantheon.

(4) Hendler, J. 1994. "High-performance Artificial Intelligence." *Science*, 265, 891—892. Crevier, D. 1993.

(5) Crevier, *AI: The Tumultuous History of the Search for Artificial Intelligence*, op cit.

(6) Allen, W. 1983. *Without Feathers.* New York: Ballantine.

(7) Block, N., & commentators. 1995. "On a Confusion about a Function of Consciousness." *Behavioral and Brain Sciences*, 18, 227—287. Jackendoff, R. 1987. *Consciousness and the Computational Mind.* Cambridge, Mass.: MIT Press.

(8) Baars, B. 1988. *A Cognitive Theory of Consciousness.* New York: Cambridge University Press.

(9) Block, N. 1978. "Troubles with Functionalism." In C. W. Savage (Ed.), *Perception and Cognition: Issues in the Foundations of Psychology. Minnesota Studies in the Philosophy of Science*, Vol. 9. Minneapolis: University of Minnesota.

(10) Dennett, D. C. 1991. *Consciousness Explained.* Boston: Little, Brown.

(11) McGinn, C. 1993. *Problems in Philosophy: The Limits of Inquiry.* Cambridge, Mass.: Blackwell.

WHY WE TELL OUR GENES TO GET LOST

Steven Pinker

There is an old song by Tom Paxton, later made famous by Peter, Paul and Mary, in which an adult reminisces about a childhood toy:

> *A wonder to behold it was, With many colors bright.*
> *And the moment I laid eyes on it, It became my heart's delight.*
> *It went ZIP! when it moved, And POP! when it stopped, and WHIRRR! when it stood still.*
> *I never knew just what it was and I guess I never will.*

The whimsy of the song comes from the child-like pleasure in a complicated object with an inscrutable function. When we grow up, we demand to know what an artifact is designed to do. When we come across a contraption in an antique store, we ask what it is, and when we are told that it is a cherry pitter, the springs, hinges, and levers all suddenly make sense in a satisfying rush of insight.

This is called reverse-engineering. In forward-engineering, one designs a machine to do something; in reverse-engineering, one figures out what a machine was designed to do. Reverse-engineering is what the boffins at Sony do when a new product is announced by Panasonic. They bring one back to the lab and try to figure out what all the parts are for and how they combine to make the device work.

The human body is a complicated object, and since the seventeenth century, when William Harvey deduced that the valves in veins are there to make the blood circulate, we have understood the body by reverse-engineering it. The body is a wonderfully complex assembly of struts, springs, pulleys, hinges, sockets, tanks, pipes, pumps, and filters. Even today we can be delighted to learn what mysterious parts are for. Why do we have our wrinkled, asymmetrical ears? Because they filter sound waves coming from different directions in different ways. The sound shadow tells the brain whether the source of the sound is above or below, in front of or behind us.

The rationale for reverse-engineering living things comes, of course, from Charles Darwin. He showed how "organs of extreme perfection and complication, which justly excite our admiration" arise not from God's foresight but from natural selection operating over immense spans of time. Organisms vary, and in each generation the lucky variants that are better adapted to survival take up a larger proportion of the population. The complicated machinery of plants and animals thus appears to have been engineered to allow them to survive and reproduce.

The human mind is a product of the brain, another complex object shaped by natural selection, and we should be able to reverse-engineer it, too. And so we have for many parts of our psychology. Perception scientists have long realized that our sense of sight is not there to entertain us with pretty patterns and colors. It is contrived to grant us an awareness of the true forms and materials in the world. The selective advantage is obvious: animals that know where the food, the predators, and the cliffs are can put the food in their stomachs, keep themselves out of the stomachs of others, and stay on the right side of the clifftop. Many of our emotions are also products of natural engineering. Fear keeps us away from heights and dangerous animals; disgust deters us from eating bodily wastes, putrefying flesh, and other contaminants.

But reverse-engineering is possible only when you have an inkling of what the device was designed to accomplish. We don't understand the cherry pitter until we catch on that it was designed as a machine for pitting cherries rather than as a paperweight or wrist exerciser. The same is true in biological reverse-engineering. Through the 1950s, many biologists worried about why organisms seem to have body parts that did them no good. Why do bees have a barbed stinger that pulls the bee's body apart when dislodged? Why do mammals have mammary glands, organs that skim nutrients from the mother's blood and package it as milk for the benefit of another animal?

Today we know that these are pseudo-problems that come from a faulty idea of what the bodies of organisms are for. The ultimate goal of a body is not to benefit the body or the species or the ecosystem, but to maximize the number of copies of the genes that made the body. Natural selection is about replicators, entities that keep a stable identity across many generations of copying. Replicators that enhance the probability of their own replication come to predominate — crucially, regardless of whose body the replicated copies sit in. Genes for barbed stingers can predominate because copies of those genes sit in the body of the queen and are protected when the worker suicidally repels an invader. Genes for mammary glands can predominate because copies of those genes sit in the young bodies that are being nourished by the milk.

So when we ask questions like "who or what is supposed to benefit from an adaptation?" and "what is the design in living things a design for?," the theory of natural selection provides the answer: the long-term stable replicators — genes. This has become a commonplace in biology, summed up in Samuel Butler's famous quip that a hen is an egg's way of making another egg, in Richard Dawkins' book title *The Selfish Gene*, and in Stephen Jay Gould's remark that reproductive success "cannot be the passage of one's body into the next generation — for, truly, you can't take it with you in this sense above all!"

What difference does this make to reverse-engineering the mind? For many parts of the mind, not much. We can understand vision and fear with a vague sense that they benefit the perceiver and fearer. But when it comes to our social lives, where our actions often do not benefit us, it makes a big difference who or what we take to be the ultimate beneficiary. Mammary glands were demystified only when we realized that they benefit genes for making the mammary glands — not the copies in the mother, but the copies likely to be found in the milk-drinker. In the same way, kind acts towards our children and other relatives can be demystified when we realize that they benefit copies of the genes that build a brain that inclines a person toward kind acts towards relatives — not the copies in the kind actor, but the copies likely to be found in the beneficiaries. We nurture our children and favor our relatives because doing so has a good chance of helping copies of the genes for nurturance and nepotism inside the children and the relatives.

In the case of altruistic behavior toward non-relatives, a different explanation is needed, but it still hinges on an ultimate benefit to the genes for the altruistic behavior. People tend to be nice to those who are nice to them. Genes for trading favors with other favor-traders can prosper for the same reason that the partners in an economic trade can prosper: both parties are better off if what they gain is worth more to them than what they give up.

The theory that human social behavior is a product of natural engineering for gene propagation came to be known in the 1970s as sociobiology, and was summed up by saying that the brain is a fitness-maximizer, or that people strive to spread their genes. It offered a realization of Darwin's famous prediction in *The Origin of the Species* that "psychology will be based on a new foundation," fully integrated into our understanding of the natural world.

But there was one problem with the theory. When we look at human behavior around us, we discover that the brain-as-fitness-maximizer theory is obviously, crashingly, stunningly wrong. Much of human behavior is a recipe for genetic suicide, not propagation.

People use contraception. They adopt children who are unrelated to them. They take vows of celibacy. They watch pornography when they could be seeking a mate. They forgo food to buy heroin. In India, some people sell their blood to buy movie tickets. In our culture, people postpone childbearing to climb the corporate ladder and eat themselves into an early grave.

What are we to make of this Darwinian madness? One response is to look for subtle ways in which behavior really might aid fitness. Perhaps celibate people have more time to raise large broods of nieces and nephews and thereby propagate more copies of their genes than they would if they had their own children. Perhaps priests and executives in childless households make up for their lack of legitimate offspring by having many clandestine affairs. But these explanations feel strained, and less sympathetic observers have come to different conclusions: human behavior has nothing to do with biology and follows arbitrary cultural norms instead.

To anyone with a scientific curiosity, it would be disappointing if human behavior had to be permanently walled off from our understanding of the natural world. The founders of a new approach called evolutionary psychology have argued that it needn't be. The anthropologist Donald Symons, and the husband-and-wife team of John Tooby, another anthropologist, and Leda Cosmides, a psychologist, have shown that when you think it through, you find that the gene-centered theory of evolution does not predict that people are fitness-maximizers or gene-propagators.

First, natural selection is not a puppetmaster that pulls the strings of behavior directly. The target of selection, the genes buried in eggs and sperm, are not able to control behavior either, because they are obviously in no position to see the world or to move the muscles. Naturally selected genes can only design the generator of behavior: the package of neural information-processing and goal-pursuing mechanisms called the mind.

That is why it is wrong to say that the point of human striving is to spread our genes. With the exception of the fertility doctor who artificially inseminated patients with his own semen, the donors to the sperm bank for Nobel Prize winners, and other kooks, no human being (or animal) strives to spread his or her genes. The metaphor of the selfish gene must be taken seriously: people don't selfishly spread their genes; genes selfishly spread themselves. They do it by the way they build our brains. By making us enjoy life, health, sex, friends, and children, the genes buy a lottery ticket for representation in the next generation, with odds that were favorable in the environment in which we evolved (because healthy, long-lived, loving parents did tend, on average, to send more genes into the next generation). Our goals are subgoals of the ultimate goal of the genes, replicating themselves. But the two are different. Resist the temptation to think of the goals of our genes as our deepest, truest, most hidden, motives. Genes are a play within a play, not the interior monologue of the players. As far as we*are concerned, our goals, conscious or unconscious, are not about genes at all, but about health and lovers and children and friends.

Once you separate the goals of our minds from the metaphorical goals of our genes, many problems for a naturalistic understanding of human behavior evaporate. If altruism, according to biologists, is just helping kin or exchanging favors, both of which serve the interests of one's genes, wouldn't that make altruism just some kind of hypocrisy? Not at all. Just as blueprints don't necessarily specify blue buildings, selfish genes don't necessarily specify selfish organisms. Sometimes the most selfish thing a gene can do is to build a selfless brain — for example, one that gives rise to a loving parent or a loyal friend.

Take another example. Recently a major newsmagazine ran a cover story entitled, "Adultery: Is it In our Genes?" A skeptic wrote in to say that since adulterers take steps to prevent pregnancy, adultery has nothing to do with a strategy for spreading genes. Ah, but whose strategy we are talking about? Sexual desire is not people's strategy to propagate their genes; it's people's strategy to attain the pleasures of sex, and the pleasures of sex are the genes' strategy to propagate themselves. When the genes don't get propagated, it's because we are smarter than they are.

But if a desire for sex serves the interests of the genes, are we condemned us to an endless soap opera of marital treachery? Again, not if you remember that human behavior is the product of a complex brain with many components, which can be thought of as distinct circuits, modules, organs, or even little agents, as in Marvin Minsky's metaphor of The Society of Mind. Perhaps there is a component for sexual desire that serves the long-term interests of the genes by making more children, but there are, just as surely, other components that serve the interests of the genes in other ways. Among them are a desire for a trusting spouse (who will help bring up the copies of one's genes inside the children), and a desire not to see one's own body — genes included — come to an early end at the hands of a jealous rival.

There is a second reason that behavior should not and does not maximize fitness. Natural selection operates over thousands of generations. For ninety-nine percent of human existence, people lived as foragers in small nomadic bands. Our brains are adapted to that long-vanished way of life, not to brand-new agricultural and industrial civilizations. They are not wired to cope with anonymous crowds, written language, modern medicine, formal social institutions, high technology, and other newcomers to the human experience.

Since the modern mind is adapted to the Stone Age not the computer age, there is no need to strain for adaptive explanations for everything we do, such as pornography, drugs, movies, contraception, careerism, and junk food. Before there was photography, it was adaptive to receive visual images of attractive members of the opposite sex, because those images arose only from light reflecting off fertile bodies. Before opiates came in syringes, they were synthesized in the brain as natural analgesics. Before there were movies, it was adaptive to witness people's emotional struggles, because the only struggles you could witness were among people you had to psych out every day. Before there was effective contraception, children were difficult to postpone, and status and wealth could be converted into more children and healthier ones. Before there was a sugar bowl, salt shaker, and butter dish on every table, and when lean years were never far away, one could never get too much sweet, salty, and fatty food.

And, to come full circle, right now you and I are co-opting yet another part of our minds for an evolutionarilys novel activity. Our ancestors evolved faculties of intuitive engineering and intuitive science to master tools and make sense of their immediate physical surroundings. W e are using them today to make sense of understand the universe, life, and our own minds.

Reverse-engineering our minds — figuring out what they are designed to accomplish — could be the fulfillment of the ancient injunction to know ourselves, but only if we keep track of who is designed to accomplish what. People don't have the goal of propagating their genes; people have the goal of pursuing satisfying thoughts and feelings. Our genes have the metaphorical goal of building a complex brain in which the satisfying thoughts and feelings were linked to acts that tended to propagate those genes in the ancient environment in which we evolved. With that in mind, we might make better sense of the mysterious ways in which we humans pop, zip, and whir.