ANNALS OF THOUGHT APRIL 2, 2018 ISSUE

THE MIND-EXPANDING IDEAS OF ANDY CLARK

The tools we use to help us think—from language to smartphones—may be part of thought itself.

By Larissa MacFarquhar

March 26, 2018

Clark says that our minds extend out into the world, incorporating tools and other minds in order to think. **Photo-Illustration by Alma Haser**

Where does the mind end and the world begin? Is the mind locked inside its skull, sealed in with skin, or does it expand outward, merging with things and places and other minds that it thinks with? What if there are objects outside—a pen and paper, a phone—that serve the same function as parts of the brain, enabling it to calculate or remember? You might say that those are obviously not part of the mind, because they aren't in the head, but that would be to beg the question. So are they or aren't they?

Consider a woman named Inga, who wants to go to the Museum of Modern Art in New York City. She consults her memory, recalls that the museum is on Fifty-third Street, and off she goes. Now consider Otto, an Alzheimer's patient. Otto carries a notebook with him everywhere, in which he writes down information that he thinks he'll need. His memory is quite bad now, so he uses the notebook constantly, looking up facts or jotting down new ones. One day, he, too, decides to go to MOMA, and, knowing that his notebook contains the address, he looks it up.

Before Inga consulted her memory or Otto his notebook, neither one of them had the address "Fifty-third Street" consciously in mind; but both would have said, if asked, that they knew where the museum was—in the way that if you ask someone if she knows the time she will say yes, and then look at her watch. So what's the difference? You might say that, whereas Inga always has access to her memory, Otto doesn't always have access to his notebook. He doesn't bring it into the shower, and can't read it in the dark. But Inga doesn't always have access to her memory, either—she doesn't when she's asleep, or drunk.

Andy Clark, a philosopher and cognitive scientist at the University of Edinburgh, believes that there is no important difference between Inga and Otto, memory and notebook. He believes that the mind extends into the world and is regularly entangled with a whole range of devices. But this isn't really a factual claim; clearly, you can make a case either way. No, it's more a way of thinking about what sort of creature a human is. Clark rejects the idea that a person is complete in himself, shut in against the outside, in no need of help.

How is it that human thought is so deeply different from that of other animals, even though our brains can be quite similar? The difference is due, he believes, to our heightened ability to incorporate props and tools into our thinking, to use them to think thoughts we could never have otherwise. If we do not see this, he writes, it is only because we are in the grip of a prejudice —"that whatever matters about *my* mind must depend solely on what goes on inside my own biological skin-bag, inside the ancient fortress of skin and skull."

One problem with his Otto example, Clark thinks, is that it can suggest that a mind becomes extended only when the ordinary brain isn't working as it should and needs a supplement —something like a hearing aid for cognition. This in turn suggests that a person whose mind is deeply linked to devices must be a medical patient or else a rare, strange, hybrid creature out of science fiction—a cyborg. But in fact, he thinks, we are all cyborgs, in the most natural way. Without the stimulus of the world, an infant could not learn to hear or see, and a brain develops and rewires itself in response to its environment throughout its life. Any human who uses language to think with has already incorporated an external device into his most intimate self, and the connections only proliferate from there.

In Clark's opinion, this is an excellent thing. The more devices and objects there are available to foster better ways of thinking, the happier he is. He loves, for instance, the uncanny cleverness of online-shopping algorithms that propose future purchases. He was the last fan of Google Glass. He dreams of a future in which his refrigerator will order milk, his shirt will monitor his mood and heart rate, and some kind of neurophone connected to his cochlear nerve and a microphone implanted in his jaw will make calling people as easy as saying hello. One day, he lost his laptop, and felt so disoriented and enfeebled that it was as if he'd had a stroke. But this didn't make him regret his reliance on devices, any more than he regretted having a frontal lobe because it could possibly be damaged.

The idea of an extended mind has itself extended far beyond philosophy, which is why Clark is now, in his early sixties, one of the most-cited philosophers alive. His idea has inspired research in the various disciplines in the area of cognitive science (neuroscience, psychology, linguistics, A.I., robotics) and in distant fields beyond. Some archeologists now say that when they dig up the

remains of lost civilizations they are not just reconstructing objects but reconstructing minds. Some musicologists say that playing an instrument involves incorporating an object into thought and emotion, and that to listen to music is to enter into a larger cognitive system comprised of many objects and many people.

Clark not only rejects the idea of a sealed-off self—he dislikes it. He is a social animal: an eager collaborator, a convener of groups. The story he tells of his thinking life is crowded with other people: talks he's been to, papers he's read, colleagues he's met, talks they've been to, papers they've read. Their lives and ideas are inextricable from his. His doors are open, his borders undefended. It is perhaps because he is this sort of person that he both welcomed the extended mind and perceived it in the first place. It is clear to him that the way you understand yourself and your relation to the world is not just a matter of arguments: your life's experiences construct what you expect and want to be true.

Clark seeks fusion with the world in everything he does. Most of his cars—a 1965 Triumph Herald, a 1968 Ford Thunderbird, a 1971 MG Midget, among others—have been convertibles. "On a sunny day, or just a non-rainy day, I feel trapped in a car if I can't get rid of the roof," he says. "Though I fear that you always look a bit of a plonker with the top down, so it's important to choose cars that are quirky rather than flashy." He loves electronic music, and one of his favorite things to do is go dancing. "I love the steamy, sweaty vibe of a hard-techno club," he says, "the way you can get totally lost in a sea of light, flesh, and music." Anyone who has gone clubbing with him can see that he feels the line between himself and everything else to be very thin. "After a few drinks, Andy's personality totally opens up," David Chalmers, a philosopher at N.Y.U., says. "In that moment, he is just *so* sweet and *so* lovable, and he does kind of merge with the world—everything is wonderful, everything is great! I think of that as his genuine nature, and the sober, reserved version during the day is just a proto version that is waiting for this true essence to be unlocked."

Clark is tall and spindly and moves in a hoppy, twitchy way, like a shorebird. His hair is a kind of punk mullet—spiky and gray on top, pink and a bit longer in the back. He likes costumes—he

recently appeared at a birthday party as David Bowie from the "Space Oddity" period. Even at the office, his shirts are heroic, psychedelic, the shirts of a man who trusts the world, their effect muted only slightly by his black hoodie, black jeans, and black boots. When he agreed, somewhat reluctantly, to take on the administrative role of department chair, ten years ago, he made up for it by treating himself to a large, comic-book-style, undersea-themed tattoo.

Cognitive science addresses philosophical questions—What is a mind? What is the mind's relationship to the body? How do we perceive and make sense of the outside world?—but through empirical research rather than through reasoning alone. Clark was drawn to it because he's not the sort of philosopher who just stays in his office and contemplates; he likes to visit labs and think about experiments. He doesn't conduct experiments himself; he sees his role as gathering ideas from different places and coming up with a larger theoretical framework in which they all fit together. In physics, there are both experimental and theoretical physicists, but there are fewer theoretical neuroscientists or psychologists—you have to do experiments, for the most part, or you can't get a job. So in cognitive science this is a role that philosophers can play.

Most people, he realizes, tend to identify their selves with their conscious minds. That's reasonable enough; after all, that is the self they know about. But there is so much more to cognition than that: the vast, silent cavern of underground mental machinery, with its tubes and synapses and electric impulses, so many unconscious systems and connections and tricks and deeply grooved pathways that form the pulsing substrate of the self. It is those primal mechanisms, the wiring and plumbing of cognition, that he has spent most of his career investigating. When you think about all that fundamental stuff—some ancient and shared with other mammals and distant ancestors, some idiosyncratic and new—consciousness can seem like a merely surface phenomenon, a user interface that obscures the real works below.

Thirty years ago, Clark heard about the work of a Soviet psychologist named Lev Vygotsky. Vygotsky had written about how children learn with the help of various kinds of scaffolding from the world outside—the help of a teacher, the physical support of a parent. Clark started musing about the ways in which even adult thought was often scaffolded by things outside the head. There were many kinds of thinking that weren't possible without a pen and paper, or the

digital equivalent—complex mathematical calculations, for instance. Writing prose was usually a matter of looping back and forth between screen or paper and mind: writing something down, reading it over, thinking again, writing again. The process of drawing a picture was similar. The more he thought about these examples, the more it seemed to him that to call such external devices "scaffolding" was to underestimate their importance. They were, in fact, integral components of certain kinds of thought. And so, if thinking extended outside the brain, then the mind did, too.

He wrote a paper titled "Mind & World: Breaching the Plastic Frontier," and gave it to David Chalmers, who was then a young postdoctoral fellow. Chalmers was taken with the idea, and gave the paper back scribbled all over with notes, pushing Clark, among other things, to expand his notion of cognition not only to inanimate objects but to people as well. "You need a nifty name for your position," Chalmers wrote. " 'Coupled externalism'? Or 'The Extended Mind' . . . or something along those lines." Clark liked Chalmers's comments, and they decided to rewrite the article together. They worked so closely that the finished product was, they both felt, a nice example of extended cognition in itself. They called it "The Extended Mind," by Andy Clark and David Chalmers; a note explained that the authors were listed in order of degree of belief in the paper's thesis.

When the paper first circulated, in 1995, many found it outlandish. But, as the years passed, and better devices became available, and people started relying on their smartphones to bolster or replace more and more mental functions, Clark noticed that the idea of an extended mind had come to seem almost obvious. The paper became the most-cited philosophy paper of its decade. The philosopher Ned Block likes to say that the extended-mind thesis was false in 1995 but is true now.

After the paper was published, Clark began thinking that the extended mind had ethical dimensions as well. If a person's thought was intimately linked to her surroundings, then destroying a person's surroundings could be as damaging and reprehensible as a bodily attack. If certain kinds of thought required devices like paper and pens, then the kind of poverty that precluded them looked as debilitating as a brain lesion. Moreover, by emphasizing how thoroughly

everyone was dependent on the structure of his or her world, it showed how disabled people who were dependent on things like ramps were no different from anybody else. Some theorists had argued that disability was often a feature less of a person than of a built environment that failed to take some needs into account; the extended-mind thesis showed how clearly this was so.

Clark recognized that there could be problems with a cyborg existence. The same algorithms that were so helpful in recommending music could intrude in creepy ways, and a world in which minds were constantly merging was also one that threatened to destroy privacy altogether. But maybe that would be a good thing, he thought—maybe privacy was mostly secrecy, and the airing of secrets would make human variety so visible that it would come to be more accepted. "As the lives of the populace become more visible, our work-a-day morals and expectations need to change and shift," he wrote. "As the realm of the truly private contracts, as I think it must, the public space in any truly democratic country needs to become more liberal and open-hearted." He was optimistic that things would work out in the end. "Where some fear disembodiment and social isolation," he wrote, "I anticipate *multiple* embodiment and social *complexity*."

He did not feel the need to become a cyborg in a literal way—for the moment, he was content with detachable, non-penetrative devices. What mattered for the merging of self and world was the incorporation of a thing into cognition, not into a body. But he was fascinated by Kevin Warwick, a professor in the Department of Cybernetics at the University of Reading, who had acquired the nickname Captain Cyborg. Warwick had implanted a silicon chip in his left arm which emitted radio signals that caused doors in his office to open and close and lights and heaters to switch on and off as he moved around. It felt to Warwick that he had become one with his small world, part of a harmoniously synchronized larger system, and the feeling was so pleasant that when it came time to remove the implant he found it hard to let go.

Later, in New York, by means of another, more complex implant in his arm, Warwick connected the nerve fibres in his wrist and hand to a computer. Over the Internet, he was able to control a robot hand back in Reading, and even to feel things that the robot hand was touching. Encouraged, Warwick persuaded his wife, Irena, to have an implant put in her arm as well, creating the first purely electronic communication between two human nervous systems. It could

work over the Internet. It was the first step, he claimed, toward telepathy.

Many people found Warwick extremely annoying, a buffoonish publicity seeker, but Clark loved his cyborgian ambition, his desire to merge inside and out, even more profoundly than they were merged already. He was particularly drawn to Warwick's idea of electronically mediated intimacy. How much farther could it be taken? he wondered. How intimate could two people get? Could two brains be connected in such a way as to coördinate some joint activity, such as dancing? It seemed distinctly possible. After all, the brain already consisted of two hemispheres linked by a dense bridge of neurons. And brains were known to be amazingly plastic, even late in life. "Who knows," he wrote, "what new skilled forms of interpersonal and neuroelectronic harmony may emerge?"

Clark lives on two upper stories of a big old Edinburgh row house with his partner, Alexa Morcom, a cognitive neuroscientist who studies memory. He was delighted to discover, when he first met Morcom's parents, that her great-uncle was Christopher Morcom, the first love of Alan Turing, one of the founders of computer science. Clark and Morcom have filled their apartment with a riot of small plastic objects—"Star Trek" action figures, action figures in tiny tutus, mini robots, bigger robots, Daleks from "Doctor Who," dolls, manikin torsos, as well as shelves and shelves of old records and DVDs. In the hall is a grandfather clock with a Barbie sitting behind glass where the clock face used to be, and on either side of the television in the living room is a pair of manikin legs. Behind the sofa stands a nearly life-size palm tree, made of strings of green lights, which used to stand next to Clark's beloved Jacuzzi when he lived in Bloomington, teaching at Indiana University. He knew that he wouldn't have a Jacuzzi in Edinburgh, so he brought the tree home to remind him of it.

On the stair landing between the two floors is an incongruous gesture to emptiness—a Buddha head and a few stones. Morcom meditates regularly and goes on meditation retreats. Clark has tried meditation a couple of times, but he finds that he just sits there and doesn't get much out of it.

He is not really the emptiness type. He loves stuff—he welcomes it into his mind and into his

house. He loves technology, and he loves old things, and he loves old technology most of all. His favorite movie is "Brazil"—a romantic tale set in a future automated by such endearingly retrograde technology as pneumatic tubes and mechanical breakfast-makers. Once, years ago, he gave a talk in Los Alamos and was taken to the Black Hole, a store that sold defunct scientific equipment that had been bought in bulk from the National Laboratory by the store's owner, a nuclear-weapons technician turned peace activist. Clark was dazzled by the merchandise —"heavyweight first-generation calculating machinery . . . cathode-ray tubes . . . gray, heavy, metal boxes (rather like office filing cabinets) with enormous single red buttons, labeled EMERGENCY." He bought as much as he could carry, including, he remembered later, "two black boxes full of inscrutable, but wisely glowing, valve electronics."

Fortunately for Clark, Morcom shares his taste in home décor, and she, too, is an adventurous dresser. But their work personalities are quite different.

"I tend to be a bit of a critic," Morcom said.

"You're more of a critic than me."

"In my field, there's a lot of big ideas, but I'm more the person that comes along and wants to test them and see if they're useful."

"I think I'm more of a synergizer," Clark said. "I like to see a bunch of things and see how they might fit together into a story, and the more bits of human experience that story can touch the more I'm going to like it. But I think that's how science works: some people need to run with a thing to see where they can take it; other people need to be skeptical and push back against them. I'm the one who picks it up and runs with it." He tends to get along with people who criticized his ideas. After all, he's grateful that they were writing about his work.

"Without your critics, you've not got a career," he said.

"Exactly," Morcom said. "It means nobody's paying attention to you. Whereas in science there's a whole row going on about criticizing people in public. The number of times that I've seen people give talks and people are thinking, That's bollocks, absolute shit data, and no one brings it up."

Clark grew up in a working-class neighborhood in South London. His father was a policeman who loved mathematics; his mother was a housewife who wrote poems and articles for the local paper. Clark was the first in his family to go to university. The idea came up only because a priest suggested it; his father thought customs and excise would be a sound career choice. As a kid, he spent most of his time reading Marvel comics. He was less interested in reading ordinary fiction; he didn't find, as some people do, that writing called up images in his mind, whereas with comics all the bright pictures he could want were right out there on the page.

When he went to university, at Stirling, in Scotland, he planned to study French literature—he'd quite enjoyed reading Sartre and Camus in high school—but once there he got drawn into philosophy. He found that he was good at logic, and, when it dawned on him that philosophy was something you could actually do for a living, he went on to get his Ph.D., in philosophy of mind, while living in London, with a couple of people who sold the *Socialist Worker*, in a grotty flat on the Isle of Dogs.

After he finished his Ph.D., in the early nineteen-eighties, he got a job as a temporary lecturer at Glasgow University, where he taught arguments for the existence of God. It wasn't really his thing, the existence of God, but that was the opening there was. Meanwhile, he taught a night class on the mind and artificial intelligence, and began to read about what became known later as GOFAI—Good Old Fashioned A.I. GOFAI created a kind of machine intelligence by programming computers with a knowledge base of symbols, and algorithms to manipulate them. GOFAI had proved quite successful at solving certain sorts of problems—problems requiring logic and precision, the kind that humans tended to find difficult. But it was very distant from the cognition you might find in a real animal. Humans could do logic problems, but usually only with the help of tools, like pen and paper. He began to wonder whether GOFAI had made a fundamental error, mistaking what a tool-using mind could do for the cognition of a brain alone.

At the time, the discrepancy between symbolic A.I. and animal cognition didn't necessarily seem like an issue. GOFAI people weren't trying to build animals—they were trying to build intelligence. The thought was that the mind was a kind of software program, and the body and the brain were just hardware, so there was no reason in principle that cognition couldn't be reproduced on a

different kind of hardware—on a silicon-based machine, say, rather than on carbon-based flesh. For this purpose, you didn't need all the other equipment that came with animals—arms, legs, lungs, heart. Lurking behind this thesis was the mostly unspoken hope that if you could upload a mind onto a computer then that mind could be preserved and its owner would not die.

Clark found it liberating to imagine minds freed from their ordinary, meaty bodies, but GOFAI felt a bit too intellectual, a bit too high up. The symbolic A.I. systems were powerful, but they were also quite brittle—if some small thing went wrong, they didn't work at all. Then, a couple of years later, in the mid-eighties, he heard about a new approach to A.I. called connectionism. Connectionists took a different tack, by attempting to simulate the way that millions of neurons, each of which was very simple and responded only to its immediate neighbors, combined in the brain to produce complex cognition. Instead of programming an artificial neural network with symbolic knowledge, a language that was complete from the get-go, the idea was to see if artificial networks could learn, a little at a time, building on very simple beginnings. And, indeed, the new neural networks appeared to be much more flexible and robust than the symbolic systems had been—they could survive damage and noise. And because they worked simultaneously along multiple parallel paths, instead of in one orderly serial, they were much faster.

The artificial networks seemed closer to human cognition than GOFAI was, and at first Clark found that very exciting. But despite the early hopes of the connectionist programmers the results were disappointing. "Where are the artificial minds promised by 1950s science fiction and 1960s science journalism?" he wrote. "Why are even the best of our 'intelligent' artifacts still so unspeakably, terminally dumb? One possibility is that we simply misconstrued the nature of intelligence itself. We imagined mind as a kind of logical reasoning device coupled with a store of explicit data—a kind of combination logic machine and filing cabinet." He suspected that much of A.I. was marshalling the increasing power and abilities of computers and steering them determinedly in the wrong direction.

He came to believe that if you were going to figure out how intelligence worked you had always to remember the particular tasks for which it had evolved in the first place: running away from predators and toward mates and food. A mind's first task, in other words, was to control a body.

The idea of pure thought was biologically incoherent: cognition was always embodied. In the early days of A.I., intelligence had for the most part been talked about as the ability to do things that A.I. researchers found hard, like proving theorems and playing chess. Things that small children found easy, such as walking around without bumping into walls, or telling the difference between a stuffed animal and a table, were not thought of as requiring any interesting sort of intelligence at all. But then some researchers started to build robots, and they discovered that programming childlike skills like walking was actually extremely difficult—harder than chess.

In the mid-nineteen-nineties, when he was teaching at Washington University in St. Louis, Clark decided that he, too, needed a few robots to think with. He had always loved robots—the uncanniness of a machine that behaved like something alive. The robots he had were very simple and easy to program: they looked like little doughnuts on wheels. But, when it came to robots, simplicity was not always a bad thing. In fact, the unexpected virtue of simplicity was one of the most important lessons that had emerged from robotics.

In St. Louis, Clark started reading around in robotics. He discovered an Australian roboticist at M.I.T. named Rodney Brooks who had been thinking along the same lines as he had: maybe trying to install a ready-made higher intelligence was misguided. Maybe the way to go was building an intelligence that developed gradually, as in children—seeing and walking first. Perhaps intelligence of many kinds, even the sort that solved theorems and played chess, emerged from the most basic skills—perception, motor control. While constructing a robot that he called Allen, Brooks decided that the best way to build its cognition box was to scrap it altogether. Allen was more complex than Elmer and Elsie. It was controlled by three objectives—avoid obstacles, wander randomly, seek distance—layered in a hierarchy, such that the higher could override the lower. But Allen would not know, as Shakey had known, what it was heading toward. It would make no plans. It would simply encounter the world and react.

Robots like Allen, and Elmer and Elsie before it, seemed to Clark to represent a fundamentally different idea of the mind. Watching them fumble about, pursuing their simple missions, he recognized that cognition was not the dictates of a high-level central planner perched in a skull

cockpit, directing the activities of the body below. Central planning was too cumbersome, too slow to respond to the body's emergencies. Cognition was a network of partly independent tricks and strategies that had evolved one by one to address various bodily needs. Movement, even in A.I., was not just a lower, practical function that could be grafted, at a later stage, onto abstract reason. The line between action and thought was more blurry than it seemed. A creature didn't think in order to move: it just moved, and by moving it discovered the world that then formed the content of its thoughts.

The world is a cacophony of screeches and honks and hums and stinks and sweetness and reds and grays and blues and yellows and rectangles and polyhedrons and weird irregular shapes of all sorts and cold surfaces and slippery, oily ones and soft, squishy ones and sharp points and edges; but somehow all of this resolves crisply into an orderly landscape of three-dimensional objects whose qualities we remember and whose uses we understand. How does this happen? The brain, after all, cannot see, or hear, or smell, or touch. It has a few remote devices—the eyes and ears and nose, the hands farther away, the skin—that bring it information from the world outside. But these devices by themselves only transmit the cacophony; they cannot make sense of it.

To some people, perception—the transmitting of all the sensory noise from the world—seemed the natural boundary between world and mind. Clark had already questioned this boundary with his theory of the extended mind. Then, in the early aughts, he heard about a theory of perception that seemed to him to describe how the mind, even as conventionally understood, did not stay passively distant from the world but reached out into it. It was called predictive processing.

Traditionally, perception was thought to work from the bottom up. The eyes, for instance, might take in a variety of visual signals, which resolved into shapes and colors and dimensions and distances, and this sensory information made its way up, reaching higher and higher levels of understanding, until the thing in front of you was determined by the brain to be a door, or a cup. This inductive account sounded very logical and sensible. But there were all sorts of perceptual oddities that it could not make sense of—common optical illusions that nearly everyone was prone to. Why, when you saw a hollow mask from the inner, concave side, did it nonetheless look convex, like a face? Or, when one image was placed in front of your right eye—a closeup face,

say—and a very different image, such as a house, was simultaneously placed in front of your left eye, why did you not perceive both images, since you were seeing both of them? Why, instead, did you perceive first one, then the other, as though the brain were so affronted by the preposterous, impossible sight of a face and a house that seemed to be the same size and exist in the same place at once that it made sense of the situation by offering up only one at a time?

It appeared that the brain had ideas of its own about what the world was like, and what made sense and what didn't, and those ideas could override what the eyes (and other sensory organs) were telling it. Perception did not, then, simply work from the bottom up; it worked first from the top down. What you saw was not just a signal from the eye, say, but a combination of that signal and the brain's own ideas about what it expected to see, and sometimes the brain's expectations took over altogether. How could it be that some people saw a dress as white and gold while others saw the same dress as blue and black? Brains did not perceive color straightforwardly: an experienced brain knew that an object would look darker and less vivid in shade than in the sun, and so adjusted its perception of the "true" color based on what it judged to be the object's situation. (Psychologists speculate that a brain's assumptions about color may be set by whether a person spends more time in daylight or artificial light.) Perception, then, was not passive and objective but active and subjective. It was, in a way, a brain-generated hallucination: one influenced by reality, but a hallucination nonetheless.

This top-down account of perception had, in fact, been around for more than two hundred years. Immanuel Kant suggested that the mind made sense of the complicated sensory world by means of innate mental concepts. And an account similar to predictive processing was proposed in the eighteen-sixties by the Prussian physicist Hermann von Helmholtz. When Helmholtz was a child, in Potsdam, he walked past a church and saw tiny figures standing in the the belfry; he thought they were dolls, and asked his mother to reach up and get them for him: he did not yet understand the the concept of distance, and how it made things look smaller. When he was older, his brain incorporated that knowledge into its unconscious understanding of the the world—into a set of expectations, or "priors," distilled from its experience—an understanding so basic that it became a lens through which he couldn't help but see.

Being prey to some optical tricks—such as the hollow-mask illusion, or not noticing when a little word like "the" gets repeated, as it was three times in the previous paragraph—is a price worth paying for a brain whose controlling expectations make reliable sense of the world. Some schizophrenic and autistic people are strikingly less susceptible to the hollow-mask illusion: their brains do not so easily dismiss sensory information that is unlikely to be true. There are parallel differences with other senses as well. When neurotypical people touch themselves, it feels less forceful than an identical touch from another person, because the brain expects it—which is why it's hard to tickle yourself. Schizophrenics are better able to tickle themselves—and also more prone to delusions that their own actions are caused by outside forces.

One major difficulty with perception, Clark realized, was that there was far too much sensory signal continuously coming in to assimilate it all. The mind had to choose. And it was not in the business of gathering data for its own sake: the original point of perceiving the world was to help a creature survive in it. For the purpose of survival, what was needed was not a complete picture of the world but a *useful* one—one that guided action. A brain needed to know whether something was normal or strange, helpful or dangerous. The brain had to infer all that, and it had to do it very quickly, or its body would die—fall into a hole, walk into a fire, be eaten.

So what did the brain do? It focussed on the most urgent or worrying or puzzling facts: those which indicated something unexpected. Instead of taking in a whole scene afresh each moment, as if it had never encountered anything like it before, the brain focussed on the news: what was different, what had changed, what it didn't expect. The brain predicted that everything would remain as it was, or would change in foreseeable ways, and when that didn't happen error signals resulted. As long as the predictions were correct, there was no news. But if the signals appeared to contradict the predictions—there is a large dog on your sofa (you do not own a dog)—prediction-error signals arose, and the brain did its best to figure out, as quickly as possible, what was going on. (The dog is actually a crumpled blanket.) This process was not only fast but also cheap—it saved on neural bandwidth, because it took on only the information it needed—which made sense from the point of view of a creature trying to survive.

But figuring out how to combine top-down predictions and bottom-up signals was not always

easy. When prediction-error signals arose, the brain had to weigh two competing accounts of what was happening: the prediction and the new information. Which should it trust? Its priors, which had generated the prediction, had proved trustworthy in the past; and sometimes the information coming from the eyes wasn't reliable. Should it update its priors based on the new information? (There is a dog on the sofa—right there!) Or should it reject the information on the ground that it seemed highly likely to be wrong? (Dogs don't just appear out of nowhere inside apartments.) What the brain needed to do was figure out how probable it was that this particular prior was correct, and how probable it was that the new sensory information was correct, and crunch those two probabilities together to come up with an answer.

To Clark, predictive processing described how mind, body, and world were continuously interacting, in a way that was mostly so fluid and smoothly synchronized as to remain unconscious. He wrote a book on the subject titled "Surfing Uncertainty," and surfing was his metaphor for life: yes, the waves that the ocean threw up at you could be wild and cold and dangerous, but if you surfed over and over again, and went with the waves instead of resisting them, and trusted that you would be O.K., you could leave your self-conscious mind behind and feel a joyful sense of oneness with the world.

Clark saw the theory of predictive processing through the scrim of his optimistic personality. But it's not obvious that a theory emphasizing the uncertainty of perception—the way that the brain has to *infer* what is outside rather than straightforwardly taking it in—is a theory of oneness. To another philosopher who had taken an interest in predictive processing—Jakob Hohwy, who taught at Monash University, in Melbourne—the theory emphasized, on the contrary, how very difficult it was for the brain to understand things outside itself. Clark saw the brain as travelling light, taking in only the news, only what it needed for its next move; but Hohwy saw how much heavy mental equipment was necessary to process even the briefest glance or touch. He wrote an essay for a forthcoming book titled "Andy Clark and His Critics," in which he proposed a counter-metaphor to Clark's joyful surfer: Nosferatu. The brain was like a vampire, shut in a coffin.

"A lot of us feel that we are not very much in tune with the world," Hohwy says. "The world hits

us and we don't know what to do with the sensory input we get. We are constantly second-guessing ourselves, withdrawing, and trying to figure out what is happening. Something that is very familiar to a lot of people, certainly myself, is social anxiety. We are trying to infer hidden causes—other people's thoughts—from their behavior, but they are hidden inside other people's skulls, so the inference is very hard. A lot of us are constantly wondering, Did I offend that person? Do they like me? What are they thinking? Did I understand their intentions?" To Clark, the most noticeable thing about the mind was the way its understandings were so often swift and perfectly tailored to the body's needs; Hohwy noticed how often things went wrong. "I think a lot about mental illness," he says. "We forget what a high per cent of us have some mental illness or other, and they're all characterized by the internal model losing its robustness. One per cent of us have schizophrenia, ten per cent depression, and then there is autism. The server crashes more often than we think."

In 2008, Clark came across an article in *New Scientist* that described what purported to be a unified theory of the brain. The theory involved the predictive-processing ideas that he'd already been thinking about, but it was broader, explaining not just cognition and perception but also action with a single mechanism. Clark learned that this new theory had been conceived by a University College London professor, Karl Friston, the most-cited neuroscientist in the world. Friston had invented a statistical technique for analyzing brain activity in neuroimaging experiments, but he regarded neuroimaging as his day job: he spent his weekends contemplating theoretical neurobiology. Friston called his idea the free-energy principle. Free energy, as Friston defined it, was roughly equivalent to what Clark called prediction error; and the brain's need to minimize free energy, or minimize prediction error, Friston believed, drove everything the brain did.

Clark and Friston met and started talking. Previously, Friston had done most of his conceptual thinking on Sundays, alone in his office—a room on Queen Square, furnished in the manner of M's office in a James Bond film (a standing globe, a cocktail table with several champagne flutes on it, a hanging tapestry, a sofa draped with a shawl). He had no connection to the philosophical end of cognitive science. "Until I met Andy," he wrote later, "I did not really understand

philosophy. I knew it was a good thing; like the national parks, poetry, village fetes, history—and other nice things that enrich our life. However, I never really understood its (scientific) purpose."

But Friston had begun to realize that he was not very good at explaining himself. He had tried, but nobody understood him. Psychologists and neuroscientists couldn't understand him because they didn't have the mathematics—that couldn't be helped. But mathematics people didn't understand him, either. Reading groups and discussions had been organized in universities from New York to Melbourne with the mission of understanding Friston's free-energy principle, only to disband, inevitably, in failure. The impossibility of understanding Friston had become an online meme. An artificial-intelligence scholar who taught at Northwestern posted an article titled "How to Read Karl Friston (in the original Greek)." Somebody started a Twitter account, "Farl Kriston," that began, in its first few months, by tweeting impenetrable quotes from Friston himself—"In what follows, we assume that the imperative to maximise model evidence is a (possibly tautological) truism"—before degenerating, HAL-style, into desperate gibberish ("I am, whatever I think I am. If I wasn't, why would I think I am?").

Friston's free-energy principle was particularly exciting to Clark, however, because it seemed to link predictive processing to his earlier thinking about embodied cognition (the way that thinking had evolved for and with the body). Friston believed that minimizing prediction error—roughly the same as minimizing free energy—caused the body to act. True, this account of bodily action sounded a bit peculiar. How does the brain cause an arm to move? It *predicts* that the arm is moving. Proprioceptive sensors issue frantic error signals to the muscle telling it that the arm is not moving; the muscle resolves this uncomfortable situation by causing the arm to move, thus rendering the brain's prediction correct.

To Clark, the incorporation of bodily action into predictive processing's mind-world loops made sense. But he was leery of the theory's all-encompassing ambition. Friston was not content to formulate a theory of the human brain; he had applied his principle to animals, even plants. Ever since he was a child, he said, he had felt "an obsessional drive to integration and simplification": he was initially drawn not to neuroscience but to mathematics and physics. Clark, on the other hand, was attached to a view of the world, derived from evolutionary biology, that saw life as a messy,

ad-hoc business, patched together bit by bit over the eons, one system on top of another, with lots of redundancy and clutter along the way. Simplicity did not attract him. He was also suspicious of it—it didn't smell to him like the right answer.

The basic problem, Clark thought, was that he was a scruffy, while Friston was a neat. Clark loved variety and profusion and abundance. It wasn't just that he believed that it was *true* that living creatures were patched-together bags of tricks—he also *liked* things that way. Friston's arguments had been pulling him toward simplicity—he was now prepared to entertain the idea that predictive processing was a high-level neat system that orchestrated biological scruffiness below. But he was never going to like elegance the way that Friston did. Clark told Friston that Friston was, in temperament, like the austere philosopher W. V. O. Quine. Friston had never heard of Quine, so Clark explained that Quine had once said, of what he considered to be an unnecessarily complicated idea, that its "overpopulated universe is in many ways unlovely. It offends the aesthetic sense of us who have a taste for desert landscapes."

Friston's account of action—predicting that your arm will move in order to make it move—sounded less peculiar if the term "prediction" were taken less literally. A prediction in the Fristonian sense was not a guess about the future; it was something more like a projection—a concept through which the brain understood the world. This concept could be a hypothesis about what was going on; or it could be an imagined scenario—a fantasy. The brain imagined the arm moving, pictured the arm moving, and, through the force of that fantasy, it caused the arm to actually move. Of course, sometimes reality did not coöperate: sometimes an arm was paralyzed, or caught in the jaws of a bear. In that case, the brain would be forced to deal with the prediction errors by overriding its hopeful prediction and conceding that, in this case, at least, the sensory information was correct, and the arm really wasn't moving.

The idea of a fantasy sounded oddly Freudian in the context of neuroscience, but the connection between the free-energy principle and Freud was one that Friston himself had been pursuing of late, ever since Christoph Mathys, a young neuroscientist from Zurich, had arrived to work in his lab. After Mathys had been there for six months, Friston found out by accident that he was training as a psychoanalyst on the side. Mathys hadn't mentioned this—it wasn't the first thing you

brought up among neuroscientists. But it turned out that one reason Mathys had come to the lab in the first place was that he had perceived a deep resemblance between the Freudian model of the mind and Friston's free-energy principle, and had realized that there was a historical link between the two.

Mathys knew that Helmholtz's theory of unconscious inference—originating in his childhood experience of seeing doll-like figures in a belfry—was a precursor of Friston's theory of perception; and he knew that Freud had been influenced by Helmholtz, too. (Soon after arriving in London, Mathys visited Freud's final home, in Hampstead, and was thrilled to spot a copy of Helmholtz's handbook on physiological optics on a shelf right above the famous couch.) Mathys mentioned to Friston the similarity between his free-energy principle and Freud's model, and said that they had a common ancestor in Helmholtz. Friston became very interested. He sought out people who knew more about Freud than he did, and co-wrote several papers elaborating on the connection.

Freud's version of free energy (he used the same term) was similar to his notion of excitation: an uncomfortably stimulating psychic energy, which the nervous system sought to discharge. "Accumulation of excitement," he wrote in "The Interpretation of Dreams," "is perceived as pain and . . . the diminution of the excitement is perceived as pleasure." The urge to discharge the free energy was what drove a person to act—to move around, to seek sex, to work. Friston's version of free energy—prediction error—could sound at first as if it were all about cognition, just as Freud's version could sound at first as if it were all about sex, but at root they were both about survival. Minimizing prediction error, in other words, was much bigger than it sounded. When the brain strove to minimize prediction error, it was not just trying to reduce its uncertainty about what was going on in the world; it was struggling to resolve the contradictions between fantasy and reality—ideally by making reality more like fantasy. The brain had to do two things in order to survive: it had to impel its body to get what it needed, and it had to form an understanding of the world that was realistic enough to guide it in doing so. Free energy was the force that drove both.

Perhaps because Clark has been working so closely with a neuroscientist, he has moved quite far from where he started in cognitive science in the early nineteen-eighties, taking an interest in A.I. "I was very much on the machine-functionalism side back in those days," he says.

"I thought that mind and intelligence were quite high-level abstract achievements where having the right low-level structures in place didn't really matter." Each step he took, from symbolic A.I. to connectionism, from connectionism to embodied cognition, and now to predictive processing, took him farther away from the idea of cognition as a disembodied language and toward thinking of it as fundamentally shaped by the particular structure of its animal body, with its arms and its legs and its neuronal brain. He had come far enough that he had now to confront a question: If cognition was a deeply animal business, then how far could artificial intelligence go?

He knew that the roboticist Rodney Brooks had recently begun to question a core assumption of the whole A.I. project: that minds could be built of machines. Brooks speculated that one of the reasons A.I. systems and robots appeared to hit a ceiling at a certain level of complexity was that they were built of the wrong stuff—that maybe the fact that robots were not flesh made more of a difference than he'd realized. Clark couldn't decide what he thought about this. On the one hand, he was no longer a machine functionalist, exactly: he no longer believed that the mind was just a kind of software that could run on hardware of various sorts. On the other hand, he didn't believe, and didn't want to believe, that a mind could be constructed *only* out of soft biological tissue. He was too committed to the idea of the extended mind—to the prospect of brain-machine combinations, to the glorious cyborg future—to give it up.

In a way, though, the structure of the brain itself had some of the qualities that attracted him to the extended-mind view in the first place: it was not one indivisible thing but millions of quasi-independent things, which worked seamlessly together while each had a kind of existence of its own. "There's something very interesting about life," Clark says, "which is that we do seem to be built of system upon system upon system. The smallest systems are the individual cells, which have an awful lot of their own little intelligence, if you like—they take care of themselves, they have their own things to do. Maybe there's a great flexibility in being built out of all these little bits of stuff that have their own capacities to protect and organize themselves. I've become more and more open to the idea that some of the fundamental features of life really *are* important to understanding how our mind is possible. I didn't use to think that. I used to think that you could start about halfway up and get everything you needed." ♦

Published in the print edition of the April 2, 2018, issue, with the headline "Mind Expander."

Larissa MacFarquhar, a staff writer at The New Yorker, is the author of "Strangers Drowning: Impossible Idealism, Drastic Choices, and the Urge to Help."