

Art and the Limits of Neuroscience

What is art? Why does it matter to us? What does it tell us about ourselves? These are our questions. In chapter 8 I expressed skepticism about the neuroscientific turn, the program of using the methods and tools of neuroscience to illuminate problems of art. I return to this theme in this chapter. It is critical that I do so for two reasons. The first is that we can learn a lot about art and its importance to us by better understanding why the neuroart program fails. Second, despite obvious shortcomings and the absence of anything resembling interesting findings or results, the neural reductionist approach to art enjoys remarkable prestige and popularity. This fact itself reveals something disturbing about our culture and our values.

Part of the problem is that neuroscience is straitjacketed, not by the methods of science, to be sure, but by unacknowledged philosophical assumptions, not so much by a theory as by an ideology about what we are. Each of us is a brain in a vat of flesh and bone, or, to change the image, we are like submariners in a windowless craft (the body) afloat in a dark ocean of energy (the world). We know nothing of what there is around us except what shows up on our internal screens.

Crucially, this model—*you are your brain*, the body is the brain's vessel, the world, including other people, are unknowable stimuli, sources of irradiation of the nervous system—is not one of neuroscience's findings; it is rather a raft of assumptions that have been taken for granted by neuroscience from the start; it is Descartes's conception but given a materialist makeover.

Of course not every neuroscientist subscribes to the you-are-your-brain ideology. In fact, there are some thinkers, Francisco Varela and Antonio Damasio, for example, who have tried to shape a viable alternative. My own research in philosophical cognitive science, including joint work with Evan Thompson, Susan Hurley, and Kevin O'Regan, is meant to be not so much an attack on neuroscience as an effort to reform it from within by weening it away from its Cartesian default assumptions.

In fact, the problem of fitting the study of brain into a more comprehensive study of mind has been recognized to be a central challenge for cognitive science since the birth of the discipline in the last half of the twentieth century. David Marr's 1982 book *Vision* is a critical text here. David Hubel and Torsten Wiesel won the Nobel Prize for Physiology and Medicine the year Marr's book was published. They were awarded the prize for their work on the mammalian visual system, based on devising techniques for recording from individual cells. This research held out the promise that it might be possible to explain how we see from the bottom up, so to speak, cell by cell. What we see, they considered, is created by the activation of cells each of which is specialized, let's say, for shape, color, movement, orientation, and so on. Marr resisted this optimism. Trying to understand vision at the level of the individual cell, he warned, would be like trying to understand bird flight by studying individual feathers. The question ought not to be what the cells are doing individually. Instead we should ask what problem or problems the visual system itself is trying to solve. Or to put it in the terms Marr favored: What function is vision computing? Once you've figured that out, Marr reasoned, then you can ask the further question: How does a mechanism like the human brain—in contrast, for example, to a digital computer—implement the information-processing laws or algorithms that constitute vision?

It's tempting to view Marr as arguing that we needn't focus on hardware; to understand the mind, it is the brain's software that is primary. This would be half right. Marr's deeper point, I think, is not that we shouldn't be focused on the brain; the point is that insofar as we focus on the brain, we need to look to it not as a piece of meat, or an electrochemical system, but as, in effect, a computer.

No one has done more to sharpen this idea for cognitive science

than Daniel Dennett. The brain, he argues, is a *synactic* engine (that is, an engine that makes use of purely formal, physical, meaningless properties of symbols). But it functions as a semantic engine (that is, as if it were sensitive to meaning and significance). The question is, *How?* How do the mere causal processes unfolding in the body's tissue come to have the world-referring, computation-crunching cognitive significance that they do have? Dennett urges that the way we get from mere neurophysiology to an understanding of the brain as subserving and enabling intelligent life is by framing questions about what the brain does in relation to the whole life of the animal. We shift back and forth from what the animal does and achieves to what happens in the brain, and so we tell a story about how what is going on in the brain belongs to and is a part of the story of the animal's life. Importantly, we don't reduce that life to the brain.

Dennett's account has a novel consequence. Suppose that a woman in France reads in the French papers about a murder committed by a Russian in Trafalgar Square in London. We read about the same event in our papers as do the Russians back home in theirs. All of us come to believe that a Russian committed murder in Trafalgar Square. But notice, there is no reason to think that this shared fact about our mental states is realized in the co-occurrence, in each of us, of one single brain state. We are different people with different brains and different histories in different places learning about what is going on by reading about it in different languages. The reason we all have the same belief about the same crime is not that our brains are, in any specific respects, in the same neurophysiological condition. That we share beliefs isn't a fact about our brains. And this is so even though, for Dennett, the brain is crucial for all our cognitive achievements. Beliefs, for Dennett, are something more like functional states of the whole person, states that supervene not on internal goings-on alone but depend on the person or animal's relation to its environment.

For Dennett, meaning isn't inside the head. So nothing inside the head—not the brain, not anything else—could be the source of meaning.

Despite the work of Dennett, Marr, and others on the conceptual foundations of cognitive neuroscience, it remains the case that most neuroscientists—although not those named above—even those

not working on such grand issues as the nature of consciousness, art, and love—are committed to a single proposition that is in fact tantamount to the same Cartesian idea that they would be embarrassed to endorse outright. The momentous proposition is this: every thought, feeling, experience, impression, value, argument, emotion, attitude, inclination, belief, desire, ambition *is in your brain*. We may not know how the brain manages this feat, but, so it is said, we are beginning to understand. And this new knowledge of how the organization of bits of matter inside your head can *be* your personality, thoughts, understanding, wonderings, religious or material or sexual impulses is surely among the most exciting and important in all of science. Or so it is claimed.

The abiding idea here, then, is that our lived experience, our daily world, our everyday actions and reactions and feelings and concerns are events in the nervous system. The world itself is a domain of *we know not what* that acts on the nervous system and is screened off by its own effects. We find ourselves on *this* side of the wall of effects separating world and mind. We know things not as they are in themselves; we know only our brain's internal fabrications of them. Very few scientists openly assert such things—or if they do it is usually off the record or in popular books written away from the lab—but it follows ineluctably from the neural reductionism, the internalism, that is their starting point.

Some scientists try to escape this Cartesian vertigo by trying to have it both ways. They grant that we can't understand the value of money or the attachment between a parent and a child without taking up the standpoint of economics, and history, on the one hand, or the personal-level standpoint of love and caring, on the other. But this, it turns out, is simply a fact about us, about the kind of explanations we, owing to our cognitive limitations, find satisfactory. Love is a neural condition. The value we attach to money is a neurological fact about us and nothing more. Even if we find it hard to describe a mother's relation to her child without using the folk-psychological category of love, it would be possible to do so, at least in principle. If not to our satisfaction, then to the satisfaction of a better scientist than we ever manage to be.

I admit that the world acts on us, triggering events in the nervous

system. Of course it does! But it is also true that we act right back. Every movement of the eye, head, and body changes the character of our sensory coupling to the world around us. Objects are not triggers for internal events in the nervous system; they are opportunities or affordances for our ongoing transaction with them. The world shows up, in experience, not like a picture in the head. It shows up, rather, as the playing field or arena for our activity. Not the brain's activity. *Our* activity. Not activity inside our head, but activity in the world around us. Our concern is with the active life of the whole, embodied, environmentally and socially situated animal.

The brain is necessary for human life and consciousness. But it can't be the whole story. Our lives do not unfold in our brains. Instead of thinking of the creator brain that builds up the virtual world in which we find ourselves in our heads, think of the brain's job as enabling us to achieve access to the places where we find ourselves and the stuff we share those places with.

It is widely admitted that, for the present at least, we can't explain consciousness in terms of patterns of neural activation alone. In truth, we are not much beyond Descartes when it comes to trying to make sense of why or how we are conscious, as I asserted in chapter 8. I've been arguing for some time now that this is because we've been looking for consciousness in the wrong place. Consciousness is not a neural event inside us, although it depends causally on such events. Objects are not triggers for events in the nervous system, although of course they produce many such effects. And experiences are not effects, events set off as a result of the bombardment of the nervous system from outside. Rather, experiences are temporally extended patterns of active engagement between whole living beings and their worlds (including, I might add, their social world). As Francisco Varela and Evan Thompson write: brain, body, and world make consciousness happen.

It may ultimately be an empirical question whether consciousness depends solely on neural processes or whether it depends also, constitutively, on the body and the world, including the social world, and their dynamic patterns of exchange. The possibility that maybe we need finally to get out of our heads to explain or make sense of human experience is not even considered as a live option. We just assume

that the membrane dividing brain and environment is somehow the causally critical division between self and world. This is, I think, to borrow a phrase of Dennett's, "a dead giveaway of vestigial Cartesian materialism" that still holds neuroscience captive.

This is why I insisted, in chapter 8, that we can't take neuroscience for granted as an intellectual readymade and apply it to the problems of art that we care about. This is why I insist that neuroscience has yet to frame an adequate conception of our biological nature.

Indeed, if I am right, far from its being the case that we can explain art from the standpoint of neuroscience, it may be that the order of explanation goes in the other direction. That is, it may be that a better understanding of art will allow us to forge the resources to articulate a more plausible conception of ourselves, one suitable, finally, for laying the ground for a better neuroscience. Indeed, going a step farther, art—as I try to show in this book—is not a *phenomenon* at all, like digestion or eyesight, that stands in need of explanation. It is itself a mode of investigation, a style of research, into the crucial questions that interest us, e.g., our human nature.

The idea that art is itself a setting in which we study ourselves opens up the possibility for a very different kind of collaboration between art and neuroscience.

If you are at all sympathetic with the skepticism about neuroscience that I am advancing here, then you will perhaps find it plausible that so-called neuroaesthetics is just another instance of neuroscience's intellectual imperialism, just another chapter in neuroscience's attempt to come up with a brain-based theory of everything. But in fact neuroscience's recent preoccupation with art reveals something deeper and more fundamental about the neuroscientific project.

This comes out clearly when we turn to visual neuroscience and the primacy accorded there to "neural representations." Visual consciousness—what we really see, it is said—is given by a picture in the brain. Marr called this the 2½-D sketch. It is not quite a full-fledged model of the 3-D world; it is, really, its picture. Technically, it is the array of point-light intensities on the retina but transformed, in effect, into a line drawing, in which boundaries between objects are

marked out. Vision may enable us to gain knowledge of a three-dimensional world, but it does so, it is widely stated in the vision literature, by giving us two-dimensional, or $2^{1/2}$ -D, experiences.

Seeing a cube from this standpoint is having a picture of the cube—in effect a line drawing—in your brain (or in your mind). And this, in turn, explains why we experience a cube when we look at a picture of one. For the seeing brain, a line drawing is the cube's equivalent. This follows from the inexorable logic that the cube itself, in its three-dimensionality, is screened off by the optics of projection. The best we can have, in sight, is how things look from here, that is, the best we can have, in vision, is the cube's 2-D (or $2^{1/2}$ -D) projection.

Neuroaesthetics—or at least the theory of pictures, pictoriality, depiction—is in play for visual neuroscience from the very beginning. So now consider the claim advanced by the Harvard vision scientist Patrick Cavanagh that artists, insofar as they are in the business of making effective pictures, “act as research neuroscientists”; their pictures are brain experiments. If I am right, the headline here that art is a kind of neuroscience or that artists act as neuroscientists is really a statement of first principles and marks no discovery. After all, it is axiomatic for visual neuroscience that the fact that we see a cube when presented with a line drawing of one shows that, for the visual brain, at least, a line drawing of a cube is the cube's equivalent. And so the fact that we don't immediately notice mistakes in a painter's rendering of shadows shows that the brain is not governed by a knowledge of physics in constructing its visual world. Cavanagh does not need to adduce evidence to support his claim that if we really saw the world in 3-D, then pictures would look jarring and inaccurate when we move. That pictures don't shock us in this way shows that we don't see the world in 3-D after all.

To repeat, the would-be discovery that artists are in effect carrying out neuroscientific experiments when they build pictures is not really a discovery at all. It follows from the familiar starting point idea that we are confined, in our experience, not to things themselves but to neural pictures of them. Cavanagh is reading his neuroaesthetics off his philosophical starting points.

One noteworthy upshot of the neural equivalence of representations with what they represent is that objects themselves, with all

their specific, tangible substantiality, play almost no role in neuroscience. If you want to study object perception, it suffices to work with photographs of objects. The power or meaning of a picture of an object, according to the equivalency thesis, is exhausted by the power and meaning of what is depicted; and vice versa, there is nothing more to an object's significance (at least to vision) than is carried by a suitable pictorial rendition of it.

And so it should not be surprising that the psychologist and experimental neuroscientist Vilayanur S. Ramachandran offers an explanation of why we value an eleventh-century bronze statuette of the goddess Parvati by showing that, for neuroevolutionary-psychological reasons, "we" supposedly respond to full-figured women. And this notwithstanding the fact that, whatever else is true of the statuette, it is clear that it is not a full-figured woman. The treatment of the statue, the depiction, as if it were a transparency through which we encounter what it depicts, is justified by the equivalency thesis, or, in this case specifically, by the idea that the statuette is, neurologically speaking, the equivalent of a full-figured woman.

The equivalency of a picture and what it depicts, however, is deeply implausible. As I argue later on, it is one thing to see a thing and another to see its depiction. That is, whatever we say about why pictures have representational powers, it better not commit us to the idea that pictures give us the experiences we would have if we were encountering in the flesh whatever it is that they depict. When you see a picture of something, you have a sense of its presence despite the manifest and perceptually salient fact of its nonpresence. You may see a full-figured woman in the statuette. But you are manifestly not in the same state you would be in if you were actually in the presence of a full-figured woman.

A clever neuroscientist will object. If there is a difference between seeing a thing and seeing it in a picture, indeed, if the picture shows you something while at the time showing you that it isn't there, then this distinctive structure in the experience of pictures will itself be neurally realized. Neuroscience is not committed to ignoring this experientially important difference; it is committed only to the idea that there is no significant difference that is not, itself, ultimately, a neurological difference.

Well, we've already seen that the dogma that mental states of whatever kind—love, monetary value, conscious experience, etc.—are identical to brain events is not one that has yet been made good on. And anyway, the neural equivalency thesis—that seeing something and seeing its picture are states of the same kind—is implicated by Cavanagh's and Ramachandran and William Hirstein's analyses of representation in art. If you give it up, then we have to give up Cavanagh's claim that painters are neuroscientists, and Ramachandran and Hirstein's claim that the statuette is a super-stimulus, perfectly designed to titillate the brain just as a woman with large breasts and buttocks would.

In any case, there is a more pressing issue. Ramachandran and Hirstein's story leaves the *artwork*, with its distinctive values and qualities, out of the story. Everything we need to know about a work of art is clarified, for Ramachandran and Hirstein, once we do our real work, which is, they believe, explaining the sources of our titillation. And what titillates is not the picture but the world (or, in this case, women). Or rather, the picture (the statuette) titillates us only insofar as it is the world's equivalent. Ramachandran and Hirstein don't go so far as to offer an account of the specific work of art—with its distinctive look, patina, size, weight, religious significance, and so on. And this shouldn't be surprising. Just as the object falls out of consideration within the scope of neuroscience, so the *art object*, the work itself, plays no role, is never brought into focus, in the neuroscience of art (neuroaesthetics).

Indeed, it isn't clear how the artwork can ever make an appearance in neuroaesthetics. After all, for neuroscience, objects themselves are only triggers that are functionally equivalent to their neural representations. So the same is true of art objects. At best they are triggers for events in the nervous system; the object itself screens itself off thanks to its effect on the nervous system. And so it shouldn't be surprising that researchers looking at our aesthetic response to sculptural form conduct their studies in the absence of sculptures, confining themselves to pictures of sculptures, or that others identify aesthetic experiences as those triggered by the inspection, inside the magnetic resonance imaging machine, of digital photographs of paintings.

If the equivalency thesis is right—if a picture is the neurological equivalent of what it depicts—then the disappearance of the artwork itself from the study of art can be chalked up to good scientific method.

But if we grant, first, that it is one thing to see a picture of something and another to encounter it in the flesh, and second, that when it comes to art, we are interested, in part, in creations that capture our attention directly and in their own material person, so to speak, and not merely as transparencies through which to encounter whatever it is that they represent, then we've got a problem on our hands. Neuroaesthetics seems unable to bring its own subject matter, art, into focus as a subject of investigation.

At root this inability to bring the artwork into focus stems from the doctrine that the artwork's significance is its effect on our perceptual (or emotional) systems. The fact that experimentalists rely on pictures of artworks rather than artworks themselves means that they must ignore the more powerful response that would be triggered by actual artworks, a response that is sensitive not only to pictorial or representational content but also to such factors as scale and installation. But there is a deeper, less technical point, one that is prior to all that: artworks are not response triggers, and aesthetic experiences are nothing like events switched on in the brain. No stimulus could ever stand in for the artwork. Not even the artwork itself! For the artwork is not a stimulus.

Actually, as we noticed in chapter 8, this problem of bringing the artwork itself into focus in neuroaesthetics was already in evidence with the programmatic claim, advanced by Semir Zeki, that it is brains that see art and that the laws of the brain constrain art. To support this view, Zeki adduces such facts as that Mondrian's paintings exploit intricate details of human color perception and that no artist makes art from ultraviolet light, precisely because such radiation is, after all, invisible.

The problem here is that the brain constrains our experience of art. Not because of anything distinctive to do with art, or with the neural representation of art, or with the experience of art, but because the brain constrains our visual experience of anything we see. Insofar as works of art are to be looked at and visually examined,

then it will be only insofar as we can see that we can engage with them. What we have here is not so much a false account of the neural underpinnings of artistic appreciation as one that leaves art and its distinctive neural underpinnings, if there are any, out of serious consideration. It never gets that far.

A more promising approach, perhaps, would be to turn attention to *aesthetic experience*—the intense delight we take in beholding the beautiful or the awe-inspiring (the sublime)—in order to investigate its neural correlates. This is the approach taken by Gabrielle Starr and her neuroscience colleagues Edward A. Vessel and Nava Rubin. But there are problems here as well.

Let us suppose that there are neural correlates of aesthetic experience. This doesn't mean that these correlates have any explanatory significance. Here's an example from a different domain. It has been claimed that there are differences in the brains of men and women; these differences are adduced as evidence that gender is hardwired. But the thing is, if there *are* differences between men and women, behavioral, cognitive, experiential—I am not assuming that there are, but *if* there are—then it stands to reason that these differences would make a neurological difference. How could they not? But that doesn't rule out the possibility that it is social, cultural, environmental factors that drive, cause, and so explain the differences. And so with aesthetic experience. If there are aesthetic experiences, then these experiences have neural correlates. Everything leaves a trace in the brain (as opposed to the brain's tracing everything!). That much is clear. There are neural correlates for everything. But it remains to be demonstrated that those correlates have anything to teach us about the experience that interests us. I am not aware of any evidence that they have done this.

Another problem has to do with the difficulty of laying out interesting or stable or necessary (let alone sufficient) hallmarks of aesthetic experience. Some aesthetic encounters are moving and some are not. Some are moving to me but not to you. Some are moving on one occasion but not on others. Most are not moving at all until we have, so to speak, learned to get moved. Can we pick out aesthetic experiences as those we rank as most moving, say, on a four-point scale? This is Starr's methodology. She and her colleagues ask subjects

to judge photos of canonical works of art as moving (or not) on a scale of 1 to 4. I doubt that we can operationalize aesthetic experience this way. And I suspect that this not an accident but rather itself a fundamental feature of aesthetic experience. I return to this below.

In any case, Starr and her colleagues treat the experiences of looking at digital reproductions of artworks in the fMRI scanner that are ranked 4 on the scale of degrees of movingness as “intense aesthetic experiences.” And they argue that there are, in fact, striking neural correlates of such experiences, across different artistic domains (music, painting, and poetry).

Specifically, they argue that “intense aesthetic experiences,” but not less intense ones produce activation in what has been called the “default mode network.” This is said to be a network of associated brain areas that are believed to be suppressed, relative to a resting state, by task- or world-oriented activity but that regain baseline rates when attention is turned away from the world, during rest, daydreaming, imagination, thought, etc. It has been argued that the default mode network is a neural system for specifying the self. This claim resonates in an intuitive sort of way with the idea that this system is active when we are strongly moved aesthetically. It is as if being moved aesthetically transports us, turns us inward, removes us from mere seeing, mere object-oriented inspection to something more like contemplation.

This is a suggestive idea, but one that may not be well supported. For one thing, I would question whether the self can be specified, represented, or activated only by acts of withdrawal from the world. The self shows up not only as the *me* of self-reflection—as the research team of Kalina Christoff, Diego Cosmelli, Dorothee Legrand, and Evan Thompson remind us, drawing on ideas of William James—but also as the *I* of agency. There is also evidence, cited in the notes, that in fact the so-called default mode network is frequently activated during engaged involvement with tasks and the environment. It isn't clear, then, what default mode activation means, and so it is unclear what we should make of the putative correlation between such activation and what is getting labeled, for the purposes of these studies, aesthetic experience.

Now, I have already stated that I don't think you should think of perceptual experiences in general as events that take place in your

head. So it isn't surprising that I also reject the idea that aesthetic experiences are a special species of event that happen in your head. But my worries here don't derive from my own "externalist" starting point.

A striking feature of aesthetic responses—I addressed this also in chapter 8—is that they are cognitive achievements, comparable, if not identical, to getting a joke. "Getting it" requires wit, insight, understanding. Crucially, the relevant scope of understanding concerns what is communicated and expected, and in particular what is expected about ways of communicating. Aesthetic responses are like this.

Here is another feature of aesthetic responses. We have them not only, as it were, in isolated encounters with works of art (as we do when we are in the brain scanner). We frequently *learn* to have them, and our responses are informed by what teachers, critics, friends, and family say and think about the work, by what works we have seen before, and also by what we do or are interested in doing in relation to the work. Aesthetic responses are not fixed data points, but are more like positions staked out in an ongoing conversation, ongoing in our day, in our lives, and also in the historical time in our culture. Aesthetic responses are cultivated and nourished and they are also challenged. Aesthetic responses are themselves the question art throws up for us, not something we can take for granted in making sense of art itself.

This brings me to a final feature (also mentioned in chapter 8). Aesthetic responses are judgments. We take stands on works of art. We don't just "like" them; we like them, as Kant said, in the "universal voice." That is, we expect others to like them, too, and if they don't, then we expect that there is a lot to be said about our disagreement. As the philosopher Alexander Nehamas suggests: aesthetic judgments are the beginning of conversations and not their conclusions. Art is experienced in the setting of argument, criticism, and persuasion. It was Kant's insight to realize that all this is compatible with the fact that it may be essential to the art encounter that there is no way of adjudicating disputes in this area, that there are no decision procedures, no rules, no way of proving who's right and wrong. But art nonetheless raises the question: *Who's right?*

Aesthetic responses, then, are not symptoms or reactions or sta-

ble quantities. They are actions. They are modes of participation. They are moments of conversation. There is nothing about which we can even ask: What are its neural correlates? And moreover, to look for neural correlates is already to have turned away from what deserves to be called aesthetic experience, to events or even feelings in our minds.

John Dewey made the paradoxical claim—quoted as this book's epigraph—that it is the very existence of art objects that stands in the way of our framing a plausible aesthetic theory. Exactly. It isn't about the objects. And it isn't about the effect the object triggers in us. Dewey said: art is experience. We can explain it like this: it's about what we do with the art objects. In that sense it is about the *work* of art.

To wrap this up, then: neuroscience is too individualistic, too internalistic, too representationalistic, too idealistic, and too antirealistic to be a suitable technique for studying art.

But there's a deeper point to be made. Neuroscience, as we have seen, treats art as a phenomenon and so it studies art the way it studies any phenomenon. Art objects are triggers distinguished by the distinctively "aesthetic" character of the experiences to which they give rise. I have argued that the trigger conception of art and aesthetic experience actually fails to bring either of these entities into the scope of investigation. Art isn't explained; it is explained away.

And now we get to the heart of the matter. Art isn't a phenomenon to be explained. It is, rather, a mode or activity of trying to explain. This idea is central to the argument of this book. I turn to it now in the next chapter, but it is only in later chapters that I'll be able to make good on the full significance of the claim. Art and philosophy are both species of a common genus: they are, as I put it, reorganizational practices. This statement turns out to have an important and surprising implication about the relation of both art and philosophy to biology, as we noticed in chapter 2. But it also helps us appreciate that the idea that we might reduce the aesthetic to the neurological is no less wrongheaded than the idea that we might reduce philosophy itself to events in the brain.