BOOK REVIEW

RICHARD H. WILSON, JR.
California Institute for Integral Studies, San Francisco, California, USA

Gazzaniga, Michael S., *Who’s in Charge? Free Will and the Science of the Brain*. New York: Ecco (imprint of HarperCollins), 2011. (272 pp., $27.99, hardcover, ISBN-10: 0061906107; $16.99, paperback, ISBN-10: 0061906115; $10.67, Kindle edition, ASIN: B005UD1EVG)

A review, with reflections, of Michael S. Gazzaniga’s (2011) book, *Who’s in Charge? Free Will and the Science of the Brain*. Gazzaniga, a distinguished neuroscientist, wishes to connect contemporary understandings of the functioning of the human brain to the proper functioning of the American courtroom. What effect, if any, should these current understandings (and current technologies) have on legal conceptions of personal responsibility, guilt, and punishment? If, as many neuroscientists hold, the functioning of the brain wholly determines the functioning of the mind, can people rightly be held responsible for their actions? Gazzaniga argues that they can.

*Keywords: Determinism, freedom, free will, insanity plea, legal, neuroscience, responsibility.*

In 2009, Michael S. Gazzaniga—director of the SAGE Center for the Study of the Mind at the University of California, Santa Barbara, founding director of the MacArthur Foundation’s Law and Neuroscience Project, president of the Cognitive Neuroscience Institute, and more—delivered a series of lectures as part of the venerable and distinguished Gifford Lecture Series at the University of Edinburgh. The series, begun in 1888, is devoted, according to the terms of the will of its founder, Adam Lord Gifford, to “natural theology”—a consideration of religion “without reference to or reliance upon any supposedly special exceptional or so-called miraculous revelation. I wish it considered,” he continued, “just as astronomy or chemistry is” (quoted in *Who’s in Charge?* 47–48). This book grew out of those lectures.

Gazzaniga used his opportunity to relate some of the fascinating history of neurological studies—a history to which he himself has contributed with his groundbreaking split-brain experiments. But the storytelling, always absorbing,
is also always wrapped in a larger framework bordering on the philosophical. Following each stage of his historical overview, he considers the implications of the knowledge presented so far for our understanding of the bedrock human concepts of personal responsibility, of “free will,” and, more, to evolving conceptions of fairness, justice, and punishment in the U.S. legal system. By book’s end he is ready to draw some conclusions. This is heady, cutting-edge stuff, engagingly and articulately presented. While I am anything but a legal expert, it strikes me as a significant contribution to dialogue on the value and appropriate role of contemporary understandings of the brain on the shaping and implementation of American legal practices.

THE BASIC ARGUMENT

Gazzaniga wishes to connect contemporary understandings of the functioning of the human brain to the proper functioning of the American courtroom. What effect, if any, should these current understandings have on legal conceptions of personal responsibility, guilt, and punishment? If, as many neuroscientists hold, the functioning of the brain wholly determines the functioning of the mind, can people rightly be held responsible for their actions?

Gazzaniga presents neuroscience’s persuasive case against the existence of an inherent, truly existing “I” who’s “in charge” of the proper functioning of an individual, and which (or who) can therefore be held responsible for his or her actions. He does not, however, hold with the more radical determinists in his profession, arguing that (1) the mind is an emergent property of the brain, operating by its own rules and timetables, neither wholly bound by the brain nor wholly independent of it; (2) while the brain constrains the mind, the mind also constrains the brain; (3) just as multiple modules in the brain work together to create conditions for the emergence of consciousness and a unified sense of self, individual brains interacting together likewise create conditions for the emergence of a social mind; (4) freedom and personal responsibility are emergent properties of social mind; and (5) individuals, functioning in a social context, may therefore rightly be held responsible for their own actions.

He regards increased reliance on brain scans as evidence in courtroom trials as premature, but cautions that as neuroscience advances, there will very likely be valid implications for decisions about when, exactly, someone might properly be deemed incapable of responsible action.

THE BOOK

I will present, chapter by chapter, the arc of his argument, followed by some responses to it.

Introduction

Here, Gazzaniga clearly and succinctly describes what he intends to do. “I do want to show that all of the spectacular advances of science still leave us an unshakable
fact. *We are personally responsible agents and are to be held accountable for our actions, even though we live in a determined universe*” (61–63) (italics in the original).

This is actually no small thing. What might be called mental determinism—the claim that brain activities fully determine the functioning of mind: understand the one, and you will understand the other—seems to leave no room for ideas like personal responsibility and freedom. The brain, after all, is a physical entity, and—the thinking goes—therefore functions in a deterministic way, like all physical entities. (Yes, the author does briefly consider quantum mechanics and the lack of determinism at the subatomic-particle level that it reveals, but quickly points out that that does not necessarily mean that at the more macro level, like that of neurons and brains, determinism fails to hold. Most neuroscientists, he points out, believe that it does hold.) How, then, if our brains really do function in a deterministic way, and mind is a outgrowth of that functioning, can freedom and responsibility even be posited?

Chapter One: The Way We Are

Are we different from other animals? Yes. Circumstantial evidence certainly tells us so—the success of the human species in so many ways, relative to other creatures. What of the scientific evidence? What do brain studies tell us is different in humans?

Early conjectures on the nature of the brain supposed that it was a “blank slate”—a *tabula rasa*, as it is often called. Neurons, in this model, were understood to be entirely plastic: any neuron could be trained to do anything. Gazzaniga quotes John Watson, director of the psychological laboratory at Johns Hopkins University and an “outspoken behaviorist”—behaviorism is probably the best-known school of thought embodying this point of view—who said, in 1930,

> Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I’ll guarantee to take anyone at random and train him to become any type of specialist I might select—doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief, regardless of his talents, penchants, tendencies, abilities, locations, and race of his ancestors. (199–203)

Gradually, however, the evidence suggested that this view was not correct. In the late 1940s Donald Hebb developed and found evidence for the idea of neural connectivity (“neurons that fire together, wire together,” producing what he called *cell assemblies*, unified components of memory and thought). Roger Sperry built on this discovery with his own of what he called neuro-specificity: that neurons are not (*à la tabula rasa*) general-purpose, trainable to do anything, but rather—at least in higher vertebrates—specialized for the particular task they perform.

But if the brain’s behavior is not entirely determined by experience, it is not entirely pre-wired, either. Peter Marler’s experiments in the 1960 with white-gowned sparrows showed that the song juveniles learn to sing could be any of various “dialects” of their own species’ song, but never that of another species, even a closely related one. Thus, pre-wiring, but with flexibility. On the one
hand, the brain possesses genetically predetermined possibilities for behavior, and learning is the act of sifting through them, finding the right combination of those in-built capacities to accomplish some task—this alone is a revolutionary insight. But the flexibility, the influence of the environment, is still deeply present. Experience can, it turns out, affect the actual growth and development of neurons themselves, building upon and extending those pre-existing structures. Exploiting in-built possibilities, in other words, often implies physical changes in neurons and their connections, through, as Gazzaniga writes, both “epigenetic factors and experience” (355–356).

Gazzaniga cites numerous examples supporting this view of pre-existent yet malleable (up to a point) capacities in the human brain. Many are from his own pioneering split-brain research. There is plenty of such in-built functionality there—modules, pre-wired circuits of neurons, poised to do all kinds of specialized tasks and to work together on complex ones. Furthermore, neurons and their structures often differ from other creatures’ in form and/or function; there are things in the human brain that have little or no analog elsewhere, even among primates. We are, indeed, in many ways, unique.

Finally, some intriguing claims and questions, which he aims to defend and address in later chapters. In the human brain multiple modules, operating in parallel, produce “our personal narrative” (676); there is no single, central “I” in control, despite all the subjective evidence to the contrary. Why is this apparent illusion so persistent? And if there is no “I” in charge, “who” is? And what are the implications of the answers to those questions for the whole concept of personal responsibility?

Chapter Two: The Parallel and Distributed Brain

The author becomes here for the first time a central player in his own narrative. Studies he helped pioneer in split-brain patients—patients whose corpus callosum had been cut (generally as a last-ditch effort to control chronic grand mal seizures), thereby severing a primary communication pathway between the two hemispheres of the brain—sought to address, among many other things,

the problem of whether each consciousness [i.e., each hemisphere] had its own protagonist: were there then two selves? Were there also two free wills? Why aren’t the two halves of the brain conflicting over which half is in charge? Is one half in charge? Were the two selves of the brain trapped in a body that could only be at one place at one time? Which half decided where the body would be? WHY WHY WHY was there this apparent feeling of unity? Was [sic] consciousness and a sense of self actually located in one half of the brain? (973–977)

Gradually, the evidence emerging from the combined efforts of considerable numbers of brilliant researchers began to shape the current model of the brain as a collection of specialized modules—a large collection—each doing its own thing, often competing with other modules to become the focus of conscious attention, sometimes achieving it, sometimes not. As such evidence increased, evidence for a localized “I” module correspondingly decreased. It just did not seem to be there.
Well, then: if that’s so, how is the brain able to function in such a unified and coherent way? Does there not have to be “something” in charge of it all to keep it humming so smoothly?

“The answer to [that last] question may be that our brain functions as a complex system” (1,146). The notion of a complex system has emerged from the contemporary field of complexity theory. Complex systems function just fine in an organized way without a central processor. (Thus, our sense of self need not be localized, and in fact, appears not to be.) They also have a habit of unpredictably and spontaneously giving rise to emergent properties, properties that may themselves be complex systems, operating according to rules that could not have been predicted prior to their emergence, rules functioning at a different level from those of the system that produced it. Thus, we may posit along with Gazzaniga, mind emerges from brain, and a sense of self emerges from one or both of them. He will have more to say about all of this in chapter 4.

Still, still, still: that pesky sense of a unified self remains mighty persistent, even for those who have convinced themselves that it must not be so. At chapter’s end, Gazzaniga wonders, again, from whence such a persistent illusion derives. The answer, he says, is a narrative-building module called the interpreter module, the subject of chapter 3.

Chapter Three: The Interpreter

Here Gazzaniga gives many examples, drawn largely from his own work, that demonstrate and support the idea that the brain is a supervisor-less complex system and that our conscious experience of a “unified self,” and of “free will”—the freedom of that self to act and choose—are actually the output of the interpreter module.

“Consciousness is an emergent property,” begins Gazzaniga, in as clear and concise a description of that elusive phenomenon as I have seen.

From moment to moment, different modules or systems compete for attention and the winner emerges as the neural system underlying that moment’s conscious experience. Our conscious experience is assembled on the fly, as our brains respond to constantly changing inputs, calculate potential courses of action, and execute responses like a streetwise kid. (1,639–1,642)

The assembling of “conscious experience” into “a single, unified, and coherent narrative” is the responsibility of the interpreter module. Located in the dominant left hemisphere, it is a *post hoc* narrative producer. It constructs

the story that ties together all the disparate aspects of our conscious experience into a coherent whole: order from chaos. *The interpreter module appears to be uniquely human* and specialized to the left hemisphere. *Its drive to generate hypotheses is the trigger for human beliefs, which, in turn, constrain our brain.* (1,647–1,649) (italics added)
The implications of all of this are enormous. The narrative constructed by the interpreter module is built after the fact, a story about what has already happened. There is no free will here, no consciously produced driver for moment to moment activities, because the self that we perceive to be the locus of free will is a fiction, produced by the interpreter module.

How reliable are all these narratives, then? Is there any truth to them at all? I find myself drawn to pragmatist philosophers like William James and John Dewey to help address this question. They found the question of whether something is absolutely true to be fundamentally unanswerable, and therefore not something to dwell on overmuch. Rather, they suggested devoting our energies to discovering what beliefs actually work to create value, tangible or intangible, in human lives. (I oversimplify, but the point is nonetheless valid.) In a similar spirit, Gazzaniga reminds us that “all these modules”—including the interpreter—”are mental systems selected for over the course of evolution. The individuals who possessed them made choices that resulted in survival and reproduction. They became our ancestors” (1,665–1,666). The functioning of our brain, in other words, has proven its utility. The narratives created by the interpreter model have survival value. They work. They are not perfect; we all know what it’s like to embrace personal narratives that end up causing suffering for ourselves and for others. But we would not be here if our inherited capacity for these narratives were mostly unproductive. Apparently, post hoc or not, “true” or not, they function more for good than for ill.

Furthermore, in supporting the construction of beliefs that, as Gazzaniga wrote, “constrain the brain,” these narratives unexpectedly create space once again for valid, emergent conceptions of freedom and personal responsibility in human affairs. How and why this is so is largely the subject of the next chapter.

Chapter Four: Abandoning the Concept of Free Will

In chapter 4, Gazzaniga finally gets to the heart of the matter, the pivot-point for his entire argument: the implications of all of this for concepts of freedom and personal responsibility.

Most neuroscientists, Gazzaniga tells us, are last-ditch defenders, well after physicists have given up on it, of a strict causal determinism—in this case, in mental matters. The mind, they say, arises entirely from the brain, and therefore is “simply” the product of neuronal firings. Those neuronal firings, we now know, operate pretty much independently of our conscious awareness. Indeed, “conscious awareness” itself results from a never-ending competition among hundreds, thousands, perhaps millions of interdependent modules. They are humming along on their own pretty much all the time, and one or another of them manages to become the focus of our attention based on circumstances: our stomach rumbles, so we think of dinner; we smell smoke from our kitchen, so we run in and check out our stove. We think we have “free will,” but that’s as illusory as the little homunculus, the “I,” that we imagine is somewhere in our brain, exercising that will.

Gazzaniga, while agreeing with many of these basic premises, wants to push back on the conclusion. Calling on the ground-breaking work of Prirogine and others on dissipative systems, chaos theory, and emergence, he states his belief that
the mind is an emergent phenomenon, originating from the brain, but operating according to its own, separate set of rules. (That’s one of the key insights conveyed by the idea of emergence: spontaneous self-organization occurs regularly in nature, creating new systems whose rules of operation are perfectly coherent and meaningful, but which could never have been predicted from first principles of the ground of their emergence. Thus, for example, Newtonian physics is an emergent phenomenon. Its laws are perfectly coherent over a wide spectrum of phenomena, but they could never have been predicted, or derived, from quantum mechanics. Both sets of laws are true; they are somehow interconnected; but they operate on different levels. One is not more valid than the other. They offer different, complementary perspectives on reality.) Thus the mind has a reality of its own, connected to the world of firing neurons, in some sense deriving from it, but operating, in some sense, independent of it, too. Gazzaniga labels the constraining influence of the brain on the mind **upward causality**; this is the sort that determined determinists insist is all there is. But he believes that the mind also constrains the brain; this he terms **downward causality**. If we want to understand more deeply mind–brain relationships, it’s at the “layer” of the mind–brain complex where downward causality meets upward causality that we must look. That’s where the action is.

Toward the end of the chapter he begins to carry the argument forward to the focus of chapter 5. Responsibility and freedom, he believes, arise “in the space between brains” (2,181)—in social interactions, in other words. A “social mind”—his term—born of inter-mind relationships, is itself an emergent phenomenon, with its own character and rules of operation. He thus makes an important, and not entirely obvious, distinction between “free will,” understood as an illusory function of a non-existing entity, the homunculus in the machine, and “freedom,” an emergent reality at the level neither of the mind nor of the brain, but at the level of multiple minds interacting.

It’s worth quoting in full his refutation of Benjamin Libet’s famous assertion that, since, as Libet showed by directly measuring the activity of the brain, “free choices” originate in brain activity that can precede conscious awareness by many seconds, freedom is illusory:

> It is this sort of analysis that finds me realizing the reasoning trap we can all too easily fall into when we look to Benjamin Libet’s kind of fact, that the brain does something before we are consciously aware of it. With the arrow of time all moving in one direction, with the notion that everything is caused by something before it, we lose a grip on the concept of complementarity. What difference does it make if brain activity goes on before we are consciously aware of something? **Consciousness is its own abstraction on its own timescale** and that timescale is current with respect to it. Thus, Libet’s thinking is not correct. That is not where the action is, any more than a transistor is where the software action is. (2,248–2,253) (italics added)

**Chapter Five: The Social Mind**

This is a rich and fascinating chapter, in which Gazzaniga argues convincingly for the thoroughly and deeply social nature of human beings. There is not space here,
any more than there has been in any of the other chapter summaries, to trace in
detail the many experimental examples he gives, but as his conclusion is important
to the thrust of his overall thesis for this book, I will summarize the flow of his
argument here.

1. While learned behaviors cannot be inherited directly (as Lamarck famously
thought), James Mark Baldwin, an American psychologist, observed in 1896
that a tendency to acquire certain useful traits could be.

2. The Baldwin Effect, obviously named for him, is therefore “a mechanism that
explains the [genetic] evolution of phenotypic (observable trait) plasticity”
(2,436)—in other words, the evolution of a capacity for flexibility in adapting
to varying environments.

3. Social banding gave early humans an advantage in survival. As a result, those
with a genetically acquired superior capacity for social adaptability had a
survival advantage over others. Their traits were evolutionarily selected for
(and still are). This is an example of the Baldwin effect.

4. Societies and individuals thus invariably engage in a mutually causal relation-
ship, as changes in individuals like those just mentioned obviously affect the
nature of societies, and societies, in turn, work to select for individuals best
able to adapt to its needs.

5. One of the traits evolved by humans—so far discovered almost nowhere else
(just a bit in chimpanzees)—is what psychologist David Premack in 1978
labeled theory of mind. This is the capacity most of us have “to understand
that others have minds with different desires, intentions, beliefs, and mental
states, and the ability to form theories, with some degree of accuracy, about
what those desires, intentions, beliefs, and mental states are” (2,530–2,532).

6. Another trait, highly developed in human beings, is the presence of what
are called mirror neurons—neurons that fire when we watch someone else
do something. They are the same neurons that would fire if we ourselves
were doing it. “They are thought to be the neural basis of not only action
understanding, but emotional understanding as well” (2,562–2,563).

7. We have even, as a species, evolved universal moral systems, defined in this
way by psychologist Jonathan Haidt: “Moral systems are interlocking sets
of values, virtues, norms, practices, identities, institutions, technologies, and
evolved psychological mechanisms that work together to suppress or regulate
selfishness and make social life possible” (2,641–2,643).

8. Through these and other evolved/acquired traits, and due to the survival advant-
geges that complex societies impart to their members, we have become social
to our core.

9. The clear conclusion: “our brain has evolved neural circuitry that enables us to
thrive in a social context” (2,836–2,837).

Chapter Six: We Are the Law

Here the conversation veers toward one of our most fundamental social institutions,
law. Gazzaniga credits philosopher Gary Watson with the idea that as we come to
observe and know ourselves and each other, “we shape the rules that we decide to live by” (2,899). Neuroscience, by replacing the ego-centered notion of free will with—in Gazzaniga’s carefully worked out view—a more complex model of brain and (emergent) mind interacting, along with multiple minds interacting together, may give us pause. It’s possible that we may want to rethink some of those rules, particularly as manifested in our legal system. Strict determinists make the case that, because mind is solely a function of brain, which operates pretty much on its own as a complex physical machine, beyond the influence of quaint notions like individual will and purpose, people cannot, ever, logically be held responsible for their actions. Gazzaniga, as we have seen, does not adopt this view. But the current legal system is based in deep ways on “old” notions of free will. Are our laws, then, based on illusion? If so, how might neuroscience’s discoveries guide us toward something better?

The discussion is a fascinating one. Gazzaniga makes the case that people ought rightly to be held responsible for their actions, because freedom and personal responsibility are real: arising as they do from social mind, they escape the determinist trap, centered as it is on single minds alone. But he also shines his spotlight on us, the socially responsible ones, not just the bad guys. He points out that physical environment and culture shape individual brains and minds every bit as much as brains and minds shape culture and environment, in an ongoing, reciprocal, mutually causative way.

An important root of legal systems is our collective sense of morality, and Gazzaniga argues that ongoing brain studies are showing that the roots of morality—at least, some of them—lie in the brain. Even infants and young children expect their playmates to share and play nicely; when they perceive intentional transgressors, they opt for punishment for them, even if (at least for four-year-olds) it means they will need to suffer themselves. (This is the one chapter in which Gazzaniga acknowledges the powerful effects of culture on the functioning of the brain, citing Richard Nisbett’s observation of striking differences in the way Westerners and East Asians process reality. This is welcome, but alas, all too briefly considered and acknowledged. Gazzaniga presents many of the outcomes of brain studies as “facts,” seemingly true for all people. But the large majority of the studies, as far as I can tell, seem to have been done in the United States, or, if not, in the West. The idea of an independently existing ego-in-charge, that he argues so relentlessly against, may be much more of an issue here than, say, in Japan. I would like to have seen a greater cross-cultural reach in his fascinating collection of studies. What might be learned by comparing test results generated in Japan to those generated in the United States? From this book, alas, there’s no way to tell. Gazzaniga is free, of course, to focus only on Western models, but I would have preferred that that choice, and potential limitations to any universal claims made for the results, be far more explicitly acknowledged than they are.)

Gazzaniga goes into quite a bit of detail to make the case that neuroscience’s gradually increasing presence in courtrooms—in the form of brain scans, mostly, so far—is probably premature, ascribing greater powers of prediction to those scans than is warranted by the current state of the science. At the same time,
he warns us to get ready: future advances in neuroscience, and progress is rapid these days, may indeed offer significant challenges to some of the fundamental assumptions of American law.

Chapter Seven: An Afterword

In this brief afterword, Gazzaniga summarizes the points and claims he has made in his book, acknowledges his own struggles to make sure that he is offering justified and accurate explanations for the data he is seeing, and writes for us this poignant conclusion:

While reviewing material for this book, I realized that a unique language, which has yet to be developed, is needed to capture the thing that happens when mental processes constrain the brain and vice versa. The action is at the interface of those layers. In one kind of vocabulary it is where downward causation meets upward causation. In another vocabulary it is not there at all but in the space between brains that are interacting with each other. It is what happens at the interface of our layered hierarchical existence that holds the answer to our quest for understanding mind/brain relationships. How are we to describe that? . . . Understanding how to develop a vocabulary for these layered interactions, for me, constitutes the scientific problem of this century. (3,484–3,492)

REFLECTIONS

On Mind

It’s a fascinating fact, although not a surprising one, that Gazzaniga says essentially nothing about what he means by “mind,” let alone “human mind.” No one else has succeeded in defining those terms very well, either. There are signs that he means “consciousness,” but no one’s done any better with that one. He wryly quotes psychologist Stuart Sutherland’s definition of consciousness from the 1989 International Dictionary of Psychology:

The having of perceptions, thoughts, and feelings: awareness. The term is impossible to define except in terms that are unintelligible without a grasp of what consciousness means. Consciousness is a fascinating but elusive phenomenon; it is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written about it. (982)

Well then.

He certainly is focusing on that aspect of the human mind—the interpreter module—that produces our unified sense of self, of an “I” who’s in charge of things; one of his main goals is to convince us that that “I” is an illusion (he convinces me). But he situates the interpreter module in the context of a large number of mental modules, so it seems that his sense of “mind” encompasses those things, too.

I would like to step into this breach to consider briefly some conceptualizations of “mind” that have emerged from complexity studies. There are certainly those who question the scientific rigor of complexity studies, not without reason. But
since Gazzaniga himself invokes it in support of his ideas (emergence of individual and social mind), it seems only fair.

Systems- and cybernetics-pioneer Gregory Bateson and Chilean biologist Humberto Maturana (working with Francisco Varela in Paris) developed independently, in the 1960s, a conception of mind radically different from any that had come before. In Fritjof Capra’s words,

> according to the theory of living systems, mind is not a thing but a process—the very process of life. In other words, the organizing activity of living systems, at all levels of life, is mental activity. The interactions of a living organism—plant, animal, or human—with its environment are cognitive, or mental interactions. Thus life and cognition become inseparably connected. Mind—or, more accurately, mental process—is immanent in matter at all levels of life. (Capra 1996, 172)

The brain, then, is a *structure*, in which the *process* of mind manifests—or, to use Gazzaniga’s term, from which it emerges. But the brain alone does not give rise to mind, because the brain itself functions as one part of a psychosomatic organism, which in turn functions as a part of a living and nonliving environment. (In fact, “the entire dissipative structure of the organism participates in the process of cognition, whether or not the organism has a brain and a higher nervous system” at all [Capra 1996, 176].) Even in humans, the entire nervous system (not just the brain) has a role to play, and not even just that: “recent research indicates strongly that in the human organism the nervous system, the immune system, and the endocrine system, which traditionally have been viewed as three separate systems, in fact form a single cognitive network” (Capra 1996, 176).

Mind, then, encompasses cognitive events, and cognitive events inhere in all living things. Ervin Laszlo, one of the pioneers of systems philosophy, goes farther. Mind, he tells us, is an intrinsic property of all systems, not just of living ones. “If [as I have argued] systemic properties [‘structure, self-direction, self-maintenance, control, non-summativity,’ etc.] serve as criteria for the attribution of ‘interiority’ to the modules of the micro-hierarchy [a hierarchy, devised by Laszlo, of system structures based on their size, from the atomic to the global], then any module physically describable in these terms is also describable in terms of mind-events from the *introspective* viewpoint” (Laszlo 1972, 169, emphasis in original). That, say, an atom can usefully be said to have “mind-events” and “interiority” is a striking, nay, revolutionary, and counterintuitive statement at first blush, but its validity nonetheless does seem to be a natural corollary of the systems view of phenomena.

All of these ideas help bring into focus just how daunting is the task that Gazzaniga and his fellow neuroscientists have set for themselves. Brain is structure, mind is process; brain is a physical, objective, “exterior” system, mind an invisible, subjective, “interior” one. How can one even conceive of describing with scientific rigor the interactions between the two? Certainly, re-conceptualizing the mind as also, in some sense, an externally observable structure, as the brain is, might help. Can the process of mentation give rise to self-maintaining mental
structures, patterns of electrical activity in the brain that are fluid, non-rigid, yet also self-maintaining in some coherent form? Certainly the very idea of pattern self-maintenance implies constraints being downwardly imposed on the brain by the emergent mind, exactly the point of view Gazzaniga puts forward.

Or perhaps it might help if, in seeking to connect like to like, we re-imagine the brain not simply as structure, but also as process—which of course it is. Cells come into being and pass away, but also cell structures, webs of connected neurons that constantly evolve and decay, as well as moment-to-moment interactions/neuronal firings between and within them. If we think of ourselves as trying to understand relationships between two processes—fluid, ever-changing brain, and fluid, ever-changing mind—might that lead to progress?

And one more conundrum, among many: What are the effective boundaries of mind? Where does brain end, and mind begin? For that matter, where does body end, where does environment end, and mind begin?

I am reminded, in thinking about all these things, of a scene from Joe Versus the Volcano, a 1990 Tom Hanks/Meg Ryan movie for which I have an inexplicable fondness. In it, a crooked doctor diagnoses the hypochondriac main character (Hanks) with a “brain cloud.” (Always fatal, of course.) So I cannot help it: the mind reminds me of a brain cloud. Minds, like clouds, are emergent, self-maintaining, constantly interacting with everything around them, continuously in flux, dependent on their surroundings while influencing them, existent, yet non-substantial, elusive, vague, nothing, yet something . . .

Two Glosses

I will leave these ideas now simmering in the background, as I return to Gazzaniga’s work, and present some ideas of my own regarding his. I want to gloss two passages, one from chapter 4, one from chapter 7, that I think summarize his position well. I hope to tease out some implications—some connections with the ideas just presented—that may not be obvious, but that I think are truly there. They are important, in my view, and I will defend that position following the glosses. Both passages are originally from single paragraphs, although my comments, in italics, appear to break them up.

The first, from the end of chapter 4, is from a section titled “Complementarity Si, Downward Causation No.” By complementarity Gazzaniga means what Joanna Macy (1991) has termed mutual causality, a reflexive, nonlinear causality in which a influences b, which in turn influences a, which . . . Here, a and b are brain and mind. I found this passage a bit confusing, another reason to gloss it. Gazzaniga has just finished arguing for the reality of simultaneous upward (brain → mind) and downward (mind → brain) causality, and that “I” emerges from the layer where “up” meets “down.” In the paragraph that follows he appears to be suddenly backing away from that position, returning to a more deterministic stance. In fact, though, he is not, as I hope my comments will make clear.

Setting a course of action is automatic, deterministic, modularized, and driven not by one physical system at any one time but by hundreds, thousands, and perhaps millions.
Yes, this we know from all the chapters thus far.
The course of action taken appears to us as a matter of choice, but the fact is, it is the result of a particular emergent mental state being selected by the complex interacting surrounding milieu. Action is made up of complementary components arising from within and without. That is how the machine (brain) works. Thus the idea of downward causation [as opposed to complementarity, which subsumes both downward and upward causality] may be confusing our understanding. As John Doyle says, “Where is the cause?”

There is not a cause you can point to to explain consciousness, our sense of self: it arises from the complementary interaction of mind and brain. These phenomena make sense only as expressions of a single, mutually causal, complex system: mind–brain.

What is going on is the match between ever present multiple mental states and the impinging contextual forces within which it functions.

Then should we not really say, mind–brain–environment is the real system of interest here?

Our interpreter then claims we freely made a choice. (2,254–2,260)

Yes. Clear from chapter 3.

The second passage:

We are that abstraction that occurs when a mind, which emerges from a brain, interacts with the brain (3,453–3,454) . . .

Again: mind–brain (or mind←→ brain). I take “brain” here to be a stand-in for the entire nervous system—in that sense, even for the entire body. But here’s a thought. Given complementarity, why do we keep insisting that mind emerges from brain? Might it not be more accurate to say that brain and mind co-emerge, initially through evolution (which of course involves unceasing environmental impingement and interaction) from some underlying physical–mental ground of potentialities—if not in the universe, as I, like Laszlo, Bateson, et al., actually think is the case, then at least on this planet? But more: is it not accurate to say that they continually co-emerge, moment to moment, each incessantly influencing and constraining the other, as long as they both shall live? Why is one member of that pair—brain or physicality, in the materialist–scientific view, mind or mentality in some spiritualist traditions—more basic than the other? Is it not an error to think that it is?

In the end, my argument is that all of life’s experiences, personal and social, impact our emergent mental system.

So again: mind–brain completes itself as mind–brain–environment.

These experiences are powerful forces modulating the mind.

I would say, without them, there is no mind.

They not only constrain our brains but also reveal that it is the interaction of the two layers of brain and mind that provides our conscious reality, our moment in real time. (3,262–3,264)

I would say, following Gazzaniga’s own logic, that it’s the moment-to-moment interactions of the three layers of brain, mind, and environment, both human/social and non-human, “that provides our conscious reality, our moment in real time.”

So What?

Those are the passages. Now: Why do I care about all of this?
I am personally interested in the mystery of self-directed, inner transformation—the type that might be termed “spiritual.” Is there room, in Gazzaniga’s model, for such a possibility?

In a way, of course, this question is not really fair. Gazzaniga’s concerns are quite different. But I would argue that mine and his are not unrelated. By his opening the door within a neuroscientific perspective to freedom and personal responsibility, how much else has gained entry? What sort of “freedom” are we really talking about here, anyway? Does it include possibilities for self-transformation?

In chapter 4, Gazzaniga asks the rhetorical question, when arguing for the illusory nature of free will (which, remember, he distinguishes from the freedom that has its roots in social mind), “What on earth do humans want to be free from? Indeed, what does free will even mean” (1,679)? We want the world around us to behave reliably, he points out, for people to behave as we expect them to, and so on. All true. But my answer to his question would be this. We want to be free from the necessity of living our lives entirely constrained by the interlocking network of causes and conditions that have defined our lives so far. If we drink too much alcohol, we would like to believe that today, we can stop. If we have been chronically depressed and fearful, we would like to believe that we can overcome those tendencies, and live a happy, confident life. If until now we have struggled financially, we would to believe that today, we can begin to change that. And so on.

I agree with Gazzaniga absolutely that the freedom that we have to make those choices and changes exists only and absolutely in the social mind that arises from the interaction of our brain with others’. But I would extend the argument. While for the sake of scientific analysis it is of course helpful to distinguish between tendencies and capacities inherent in the brain’s structure and those whose primary source is outside of it, in truth, as I have argued above, we can never separate those two worlds, even for a moment. There is no brain independent of our relationship with the outside world, any more than, from a human perspective, there is an outside world independent of our brain. Mind—and I believe, as I hope my glosses above showed, that Gazzaniga is saying this, although not quite as directly—emerges not simply from brain, but from the never-ending interaction of brain with the outside world (and with already existing mind). I have already argued that this inevitably suggests that mind and brain (and social mind) co-emerge, that one of these cannot be said to precede the others, though this does not seem to be a conclusion that Gazzaniga himself has reached.

In these senses, any fundamental distinction among brain, individual mind, and social mind is as much an illusion as the encapsulated self, the “I” running the mental show, that Gazzaniga argues convincingly and, I believe, correctly, against. (Provisional distinctions among the three, as long as we do not grant them a solidity they do not actually possess, can, of course, be useful.) And we could extend this argument even further. While calling on ideas from chaos, complexity, and system theories, most prominently that of emergence, Gazzaniga argues, as we have seen, that mind is itself an emergent phenomenon, with its own rules and timetables—in some sense independent of the brain, but, of course, intimately and ineradically interconnected with it. But we may agree with system theorists
who argue that mind, understood as the ongoing, constant, dynamic interaction with and adjustment to the outside, by any physical system—its interiority, in a word—is ubiquitous, at all levels, and that these minds are as interlocked as the phenomena themselves, with no clear demarcations between them. From this perspective, individual minds, while in some sense “existing,” are, in fact, an inextricable part of a human social mind, as Gazzaniga rightly claims, but also, perhaps startlingly, of a “mind” permeating all phenomena.

We end up, then, following this line of thought, with a model of the brain, operating according to its own internal rules and logic, constantly interacting with individual mind, social mind, and environmental conditions of all sorts, all co-emergent with it and with each other, all interacting with all. The question for me is whether individual mind, that provisionally existing “player” in this game, can be the source of an individual’s inner transformation and, following from that (due to all those ever-present dynamic interactions), social and environmental transformation as well. Does the freedom the mind possesses in a social context extend to freedom to transform itself?

I would argue that the answer is “yes,” although it’s not clear to me that neuroscience can yet explain why, or even has a strong opinion on the matter. The human mind’s capacity for conscious focus can be turned onto itself (Macy 1991, 155). If mind constrains the brain and social mind constrains the individual mind, as Gazzaniga argues, perhaps a human mind can also become better aware of its own functioning, more tuned in to certain signals from among the cacophony of inputs competing to make it into consciousness, using its capacities for symbol, metaphor, intuition, and feeling to create new synergies and harmonies, winners and losers, patterns of interactions, among them. And perhaps, by turning those newfound, still tentative reconstructions toward the world “out there,” brain–mind can move that world, even just a little, in a direction more conducive to individual and social material–spiritual happiness. Then the world, slightly altered, can better support such ongoing internal change, and so on, in a rich example of the Baldwin effect at work.

Many spiritual traditions have argued that it is possible to attain an inner state of life that finds joy in simply being alive. The inevitable obstacles and struggles of life in this context can be understood to be a cause for growth, not despair. If we return to the idea of a “cosmic mind,” this idea may become more explicable. Our minds are, in this view, fundamentally rooted in cosmic mind, but as I suggested above, physical and mental can never be separated: they are simply complementary perspectives on phenomena. (This, by the way, suggests a nondual perspective that is neither materialist nor spiritualist. Gazzaniga seems to agree with the mainstream neuroscientific position that if the basic model for mind–brain functioning is not materialistic, it must be dualistic. This is simply not correct.) The reality of our individual lives, then, as for all phenomena, is one of a constant, mutually causal relationship, both physical and mental, with all phenomena “outside” of us. When one’s inner perception matches this reality (in Buddhist terms, when subjective wisdom and objective reality are fused), we become free—that is, released from the shackles of choices from the past made based on the illusion of an absolutely existent, unchanging self, exactly the illusion
that Gazzaniga and his fellow neuroscientists argue against as well. And finally, because this state of freedom is utterly rooted in the reality of things-as-they-are, it would seem that we might be able to frame Gazzaniga’s argument for the reality of freedom and personal responsibility in a deterministic universe in, perhaps, a new way. As humans we are born with a capacity for self-reflection that, as far as we know, is unique on this planet. It is a function of our human mind, emergent from the constant interaction of brain—what might be termed “internal cause”—not only with mind, but with external phenomena, including, prominently, other human minds—what might be termed “conditions.” (Students of Buddhism will recognize those terms.) This reflective capacity enables us, at least in theory, to see through the illusion of the homunculus in the machine to the way things really are: not just the reality of the modularity of the brain, but of our own lives, mind, and body/brain, as modules in a societal mind/body, a planetary mind/body, a cosmic mind/body, all functioning according to universal, inviolable rules, all neatly combining a certain inevitability of functioning with emergent, unpredictable phenomena, at least one of which, the human social mind, gives rise to concepts of freedom and responsibility.

The freest human being, in this model, is the one most rooted in reality. Without determinism—the rules that govern the functioning of all things—there can be no freedom, because there’s no reality to be rooted in. One needs those rules to base one’s life on, to discern and reliably use them, as much as humans can, to create something of value. I believe Gazzaniga would agree with this.

(Perhaps, as an aside, in the context of a systems view of the universe, the term “determinism” ought to be jettisoned, suggesting, as it generally does, a kind of linear causality that is nowhere to be seen in reality. I would offer “mutual causality,” Macy’s (1991) fine term, as a successor candidate, but will not insist on it here.)

But the dependency of freedom on rules works in the other direction as well, it seems to me. Inherent in the rules are (a) the potential for the emergence/evolution of sentient, symbolically reasoning entities, such as humans, into being and (b) the potential for humans’ (the term hereafter used as representative of such entities) awakening to, and completing the functioning of, those very rules. Without human freedom, I am claiming, the rules’ full potential, their full flowering, cannot be realized. And only with their realization in this way, through their splendid, intricate, mutually penetrating dance with human consciousness and freedom—with which they are, ultimately, inextricably bound up—does the possibility for human happiness and transformation at the deepest level, both individual and social, emerge.

(The dependency of rules on freedom may be considered from a rather different angle as well. Without human freedom, would they even exist as they are? We know from quantum mechanics that mind, perception, awareness, are embedded with physical reality at the very deepest level, a fact whose full implications at more macro levels we are still struggling to understand.)

So that perceived opposition between determinism/rules and freedom—particularly freedom understood as an enabler of human fulfillment and happiness? Precisely because there are unchanging rules, we can be free to become
happy in the deepest sense. Precisely because we are free, the rules can functional optimally (and perhaps, at all); they can yield their deepest, most beautiful and valuable treasures. Are determinism and freedom not, in fact, inseparable? Can one be conceived without the other? It’s a different line of reasoning from Gazzaniga’s, but it reaches a similar, if stronger conclusion: the dependency is not simply determinism → freedom, but actually works in both directions.

**IN CONCLUSION**

Neuroscience has made some real headway in understanding the built-in capacities and functions of the brain, with discoveries that have helped erase the illusion of an internal agent running the show who is “I,” and therefore of free will understood as a function of that agent. In a significant contribution, Gazzaniga has clearly articulated in his book new conceptions of freedom and personal responsibility, emergent not from a brain in isolation—as if such a thing even exists!—but from “the space between brains.” Our American legal system has long been guided by presumptions of individual agency and free will. The implications of the new model for freedom offered by Gazzaniga, while not from a wholly different world, nonetheless are likely to have eventual, far-reaching effects on understandings of guilt, responsibility, and punishment.

The science he loves and practices so skillfully is clearly beginning to probe those layers of interaction where, Gazzaniga tells us, “the action is.” Where “us” is.

There’s a lot of work to be done! It’s an exciting time.

**NOTE**

1. All references are to locations in the Kindle electronic edition of this book. (There are 4,988, total.) Locations are not equivalent to print-edition pages.

**REFERENCES**

Capra, Fritjof. 1996. *The Web of Life: A New Scientific Understanding of Living Systems*. New York: Doubleday.

Laszlo, Ervin. 1972. *Introduction to System Philosophy*. New York: Gordon and Breach.

Macy, Joanna. 1991. *Mutual Causality in Buddhism and General Systems Theory: The Dharma of Natural Systems*. Albany: SUNY Press.