David Cremins
15 December 2015
Topics in Neurophilosophy
Final Paper

Moving Freedom

Who's In Charge? Free Will and the Science of the Brain (2011) by renowned neuroscientist Michael S. Gazzaniga is, without doubt, a *tour de force*. Topics ranging from evolution to neuroscience to the courtroom weave together over the course of 220 pages to arrive at a single conclusion regarding moral responsibility: "we are people, not brains" (Gazzaniga 218). At face value, this statement is surprising coming from a scientist who never concretely denies the possibility of a deterministic universe in which "mental states do not exist without [cell-to-cell] interactions" (107), but it makes more sense in light of his argument that the mind is an emergent property of the brain which, in turn, can actually "constrain the brain" (4) from which it sprung. As the book continues, it becomes clear that Gazzaniga's interest in this theory of emergence is not even limited to the individual, though, as he outlines how brains can constrain each other, the evolution of our species, and how eventually all of this may impact our legal system. In order to better come to grips with these arguments about freedom and moral responsibility in the latter half of his book, which I will consider against the backdrop of current work in philosophy, I first turn to Gazzaniga's early three very well-developed chapters on "the science of the brain" which are worth considering before turning to his more interdisciplinary claims about issues such as "free will", of which they are the, at times questionable, basis.

The first chapter of *Who's In Charge?* is devoted to describing both the evolution of the human brain and the evolution of our scientific understanding of it. This is clearly material well within Gazzaniga's wheelhouse and would be of good use to any introductory neuroscience student. The informative descriptions of *Homo sapiens'* mental growth and specialization over

millennia are not just useful in explaining how higher cognitive functions and localized circuits came about, though. By taking the time to explain how genetic coding leads to specific neural pathways, the (limited) influence that environmental experience can have on those pathways, and the impact the formation of social groups had on the organization of individual minds and our species, Gazzaniga lays out the foundations of the rest of his book. Even though we are not all wired the same, the evolution of the "specialized…modules" (Gazzaniga 33) of the prefrontal cortex and beyond still allows for all of us to feel like we are the "self" (8), the homunculus, at the controls of life.

Moving on from the evolution of the brain, Gazzaniga spends the next two chapters describing important experiments in the field of neuroscience that will be more directly useful in understanding his later arguments. The race to discover the localization of cerebral functions (led by two lionized pioneers of brain study, Broca and Wernicke) precipitated the detection of very specific, hardwired functional modules in the brain. For example, cortical lesions to an area encompassing the module for identifying food can leave patients unable to "recognize fruit" (Gazzaniga 50). When Gazzaniga started studying split-brain patients – those with a severed corpus callosum and therefore no pathway for communication between the two hemispheres of the brain – in the 1970's, he was able to capitalize on this knowledge and prove with a series of clever experiments that the cortical distribution of functions results in a hemispheric lateralization of capabilities. The right brain largely acts as a "maximizer" (84) that finds and exploits patterns in the world. Meanwhile, the left brain acts as an "interpreter" (83) that processes actions and provides post hoc explanations for them on the fly. For instance, if a split brain patients perform some action that was prompted towards and completed by their right brain, when the left brain is made aware of this action it invents a justification for it seamlessly without realizing that this

conscious explanation is "only as good as the information [the left brain] receives" (89) and is both causally incorrect and temporally lagging. The interpreter is both effective in its confabulation and easily duped when inter-hemispheric functionality is impaired.

What is the upshot of this fact: that our world is processed in parallel by different parts of the brain, yet we experience only one, unified consciousness in our mind because of this interpreter module? In Gazzaniga's view, the "you" of consciousness, the homunculus, is just an interpretation, an "illusion" (Gazzaniga 75) that emerges out of many non-conscious brain processes: "all kinds of local consciousness systems, a constellation…[enable] the feelings of consciousness" (66). The truth behind our personal narrative is that it is an emergent property of a complex, underlying system much like traffic is a phenomenon that is created out of the interactions of thousands of cars and variable road conditions. This fact of complexity "is rooted in the laws of physics" (71). As a final precursor to the rest of his book, Gazzaniga states that, while our sense of agency is largely unfounded in terms of what actually happens in the brain (the interpreter is constantly making "order out of chaos" (102) without our awareness), our conscious mind still can play an important causal role in directing the neurons that allow it to emerge specifically because it is not bound to any specific neural circuitry due to emergence.

At this point it is worth focusing a bit more on this topic of emergence that Gazzaniga repeatedly touches on throughout his work. As reviewer Richard Wilson points out, Gazzaniga's arguments for "strong emergence" (Gazzaniga 124), in which our physical brains can be constrained from a mind that operates on a higher plane than the cells and particles which reisde in the domain of neuroscience, leans implicitly on work in "systems philosophy" that broadly argues that "the brain…is a *structure*, in which the *process* of mind manifests" (Wilson 112, italics in original). However, this less empirical, philosophical work, which has gone into formulating how

certain properties in dynamic systems can emerge spontaneously out of lesser structures, is not without critics. Can physical events possibly not be explained fully in terms of a closed causal chain? The philosopher Jaegwon Kim, for example, might reasonably take issue with this violation of his explanatory exclusion principle. Kim and others hold that there cannot be "multiple explanations of a single explanandum", that two distinct theories (or sets of laws) cannot hold for one event, there can only be one "causal explanation" of the same "data" or phenomena (Kim 77, 79). In application here, there is still plenty of debate over whether there is any proof that certain systems can give rise to behavior that conforms to level of explanation different from the underlying system. Kim would find it theoretically inconceivable that "laws are not universal…it depends on which level of organization you are describing" (Gazzaniga 130). These reservations are useful to keep in mind moving forward because the acceptance of emergence as a valid explanation of what creates mind is a major part of Gazzaniga's argument when he turns to moral responsibility, shortly after delving a little bit more into the free will debate.

In chapter four, "Abandoning the Concept of Free Will," Gazzaniga turns distinctly from reveling in the successes of neuroscience, prompted by himself and others in the last several decades, towards the implications of this work for some traditional philosophical topics. One of the main issues with this chapter is that there is little discussion of exactly what concept of free will we are supposed to be abandoning. Neuroscientific skepticism of free will in the popular press is no new thing (David Eagleman's *Incognito,* Sam Harris' *Free Will*, etc.), but usually these texts – and the many that have come before – at least define the sense of free agency in question. For his part, though, Gazzaniga does not go beyond contrasting some sense of free will with the notion of determinism in which everything is "causally necessitated by preceding events combined with the laws of nature" (111). Gazzaniga goes on then to provide some scientific evidence for a "bleak

view" (Gazzaniga 112) of free will that eradicates agency and allows for "the universe made my brain make me do it" defense. One novel example of this argument he references is a transcranial magnetic stimulation study showing that conscious interpretations of when an act occurred can be directly manipulated by applying pulses to areas involved in perceiving intention and acting on it (113). However, the relative dearth of directly grappling with free will and all its nuances still looms large, and this may be because Gazzaniga's interests are slightly different than the typical neuroscientific skeptic: in his own words, "the issue isn't whether or not we are 'free.' The issue is that there is no scientific reason not to hold people accountable and responsible" (106). For all the time Gazzaniga spends supporting determinism in which "the brain functions automatically" (127), he repeatedly argues that this automatic functioning is tangential to the issue of what we should do regarding areas such as moral responsibility on a societal level. His discussion of a deterministic understanding of the universe and the brain is underscored by much of the book up until this point, and it will remain relevant to some lesser degree until the end. But, this is tempered throughout by the idea of emergence, laid out in previous chapters, that will eventually save some concept of freedom (if not necessarily free will, an important distinction that will move us from the perspective of just neuronal firings to that of responsibility) for Gazzaniga.

Despite the subtle intricacy of Gazzaniga's claims, they mostly do little more than bump up against hefty philosophical topics, and there are some points he makes about free will that are well worth contemplating more before moving onto his conclusions about freedom more generally. An exploration of some alternative viewpoints will show that this book largely ignores important additional arguments in favor of directly advancing Gazzaniga's views.

Gazzaniga outlines the dominant view of the neuroscientific community as "upward causality" in which the brain deterministically constrains the mind, yet he never wholly endorses

this view (Wilson 108). However, this does not stop an author such as Alfred Mele in his book *Free* from claiming that Gazzaniga does not go far enough to outright reject the claims of his neuroscientific brethren such as Benjamin Libet[1]. Mele accuses Gazzaniga of selling "our mental lives short" (Mele 53) in his attempt to connect the neuroscience discussed previously with everyday social realities involving conscious decision-making. In a nutshell, Mele and others caution against naively overgeneralizing from limited neuroscientific results. Showing that some conscious processes lag behind some detectable neuronal firings does little to satisfactorily explain away the influence of conscious intentionality and rationality, it may sometimes simply "take a little time for our decisions to show up in our consciousness" (16). Further, there is as of yet "no good reason to believe that early brain activity" (39) indicates an early decision. While Gazzaniga believes that his previously described conception of "emergence is not a mystical ghost" (Gazzaniga 136), he attracts Mele's criticism because he does argue that the phenomenal agency we experience is still some sort of a ghost in the machine (i.e. "we [naively] believe that we [act of]…free will" (105)). Even if a modular account of consciousness, as advocated for by Gazzaniga in light of his split-brain research and other experiments, holds up empirically, the bearing of these neural circuitry findings on the existence of free will is fuzzy at best, especially given the limits and statistical uncertainty of current techniques, according to critics of neuroscientific determinism. As Mele points out, "free will might work very differently" (Mele 15) than a small batch of modern neuroscientists envision it. This point is never seriously taken on in Gazzaniga's book. Just showing empirically that the functionality of "you" is automatically created is not

[1] Gazzaniga does actually also criticize Libet in a section on complementarity that even the academic reviewer Richard Wilson agrees is denser and less well argued than other parts of the book. Gazzaniga holds that it does not matter if "brain activity goes on before we are consciously aware" because consciousness is an abstraction that happens "on its own time scale" (141), a point in fact echoed more articulately in Mele's own book (see below).

sufficient for dismantling and reformulating any and all concepts of free will and moral responsibility.

There is pushback from the other direction as well – theoretical criticism that Gazzaniga sells our neural lives short, too, in attempting to unify them with mental experience. In the same chapter in which Gazzaniga states reasons for free will skepticism he also lays the foundation for his closing arguments that center around the claim that there is also "downward causality" in which the mind constrains the brain (Wilson 108). Because of this, "reductionism…has been challenged by the principle of emergence" (Gazzaniga 134). In support of the notion that neural reductionists will never be able to explain the macro (functionality, consciousness, etc.) in terms of the micro (brain states), Gazzaniga brings forth the work of Eve Marder. She showed, by modeling the relatively simple nervous system of lobsters, that an "enormous diversity of network configurations…could lead to an identical behavior" (131). To some, this indicates that emergence must be correct and that the study of neurons will never be enough to fully explain behavior and mental experience. Patricia Churchland may step in here, however, and argue that even if "the functional set" of behaviors is dwarfed by "the whole set" of neuronal configurations (132), identity relationships that entirely relate the mental with the physical (in the case of humans and lobsters) are still possible. Not every higher-level state needs to have "identifications" on a one-to-one basis with a singular lower-level state, and rules which may be "observationally obvious" are "not guaranteed correctness" just because neurobiological theory is as of yet inchoate (Churchland 284, 293). Churchland goes on to argue that neuroscience should not at this stage be hindered in its explanatory ambition because so far its evidence is continually marching towards better explanations of psychological findings in terms of biology. This recommended approach is nearly antithetical to Gazzaniga's who is already urging neuroscientists to not buy into one explanandum

of all physical systems and instead asks "hard determinists [to accept] that there is more than one level" (Gazzaniga 127). Suffice to say, the debate over whether analysis has or can show that the whole (mind) constrains the parts (brain) – and therefore has properties in addition to said parts – is kept alive and well by yet another philosophical exemplar in Patricia Churchland.

If the bar Gazzaniga holds for dismantling free will and reconstructing it in his own fashion is possibly too low in a variety of respects (as seen in the preceding two paragraphs), he in turn is able to end his book on a relatively comforting note ("we are people, not brains") because the bar he holds for dismantling moral responsibility is still quite high. Before reaching this conclusion, though, Gazzaniga directs his penultimate chapter towards the neuroscience of interactions, in showing that "we are wired from birth for social interaction" (Gazzaniga 144). This chapter, "The Social Mind", includes fascinating discussions of social norms, cooperation in primates, the coevolution of brain and environment, theory of mind, mirror neurons, localized belief attribution in the right hemisphere, ingrained notions of acceptability, and how moral intuitions vary across cultures. The chapter in itself is a powerful argument for the benefit of thinking and communicating across disciplines. In addition, this last strong showing of Gazzaniga's scientific acumen results in some of the most theoretically unobvious and convincing claims of the book: social processes constrain individual minds, the environment created by ourselves and our peers has a real impact on our biology and vice versa, and "the supreme achievement of the cortex" (158) is socialized behavior.

Gazzaniga (or his publisher) is smart enough to sell some extra books by ending with a pragmatic application of the science and philosophy discussed thus far by placing it all in the context of the courtroom. His extended discussion of the different options for justice – "incapacitation, retribution, or rehabilitation" (Gazzaniga 181) – within chapter six, "We Are the

Law," may be best suited for consideration elsewhere; he outlines them well but does not make any grand assertions about their overall validity. Instead, it is most interesting to consider how our legal system is a mere manifestation of our base intuitions about fairness, which have been honed by evolution (for example, we come "hardwired for reciprocity" (209) within social groups) and how these intuitions should and will be changed in light of neuroscientific discoveries. Gazzaniga is clearly correct that brain scans should be still under serious scrutiny as pieces of evidence in trial. This is supported by Adina Roskies who writes eloquently about how the probabilistic statistical measures used to turn the data of brain scans into colorful pictures involves too great of an "inferential distance" in order to share any of the "evidential characteristics of photographs", despite their "epistemically compelling" nature (Roskies 19). However, Gazzaniga's prediction that the "mind reading techniques" (209) (a slightly facetious term) of future neuroscientists will be deemed valid and admissible by judges seems on the right track.

Why does the contentious influence of neuroscience in legal proceedings matter with respect to this book? All of the work done so far in *Who's In Charge?*, which, as we have seen rests on some shakier foundations in certain parts, comes into play here. It does not strictly matter if free will is real, because "responsibility is a contract between two people rather than a property of a brain": the emergent property of social influences constrain individual minds, individual minds are able to "inhibit their intentions when the cop walks by," and after that "determinism has no meaning in this context" (215). We cannot excuse criminals solely on the basis of their potentially skewed neurobiology because, even if free will is largely illusory, the mind of an individual can still in some capacity reign in the influence of its constituent neurons at times and, furthermore, social contracts and constructs (in this case, laws) clearly modulate the actions of criminals just as they do with every member of a species. While Gazzaniga avoids making any

specific proclamations about culpability (for example, schizophrenics should be half as excusable as dementia patients and two thirds as blameworthy as psychopaths), he does emphatically believe that personal responsibility is real and freedom at the "level of multiple minds interacting" arises "from [the] social mind [and escapes] the determinist trap, centered as it is on single minds alone" (Wilson 108, 110). The question of freedom as it relates in practical ways to moral responsibility has been entirely moved out of the laboratories of neuroscientists and has emerged into the hands of our social, evolving species.

There are dangers and advantages whenever an expert in one field waxes philosophically into another, and Gazzaniga's *tour de force* is both hindered and helped by his particular expertise. He masterfully lays out the neuroscience behind each topic of interest, gives a fair pass at the other scientific disciplines that may affect his claims, but leaves a lot of the philosophical rigor in the background or completely out of his work. Admittedly, doing otherwise would be counter to the purpose and efficacy of this book as a popular science piece, but repeatedly it seems that the fascinating subjects he does take on would benefit from more pushback from other voices, some of which are mentioned in this review. I believe even the title misleads the audience from what the author is truly interested in; Gazzaniga's book is not really about "free will and the science of the brain" as much as it is about moral responsibility and an interdisciplinary approach to the formation of mind. This second tagline probably would not send quite as many copies flying off the shelves, but at the end of the day the controversial and illuminating intersections of philosophy and neuroscience that have been not as well-trod are where most of the interesting material is (or could be), and they should not be shied away from in future works by this gifted author.

Works Cited

Churchland, Patricia Smith. *Neurophilosophy: Toward a Unified Science of the Mind-brain*. Cambridge, MA: MIT, 1986. Print.

Gazzaniga, Michael S. *Who's in Charge?: Free Will and the Science of the Brain*. New York, NY: HarperCollins, 2011. Print.

Kim, Jaegwon. "Mechanism, Purpose, and Explanatory Exclusion." *Philosophical Perspectives* 3 (1989): 77-104. Web.

Mele, Alfred R. *Free: Why Science Hasn't Disproved Free Will*. New York: Oxford UP, 2014. Print.

Roskies, Adina L. "Neuroimaging and Inferential Distance." *Neuroethics* 1.1 (2008): 19-30. Web.

Wilson, Richard H. (2013) Gazzaniga, Michael S., Who's in Charge? Free Will and the Science of the Brain, World Futures, 69:2, 102-118.