A Review of Todd E. Feinberg’s *From Axons to Identity: Neurological Explorations of the Nature of the Self*

By Alexander Murphy-Nakhnikian


** Doctoral Student, Program in Neuroscience, Cognitive Science Program, Indiana University Bloomington

Overview

Since Phineas Gage’s unfortunate encounter with an iron rod, so-called “experiments of nature” have provided valuable insight into the function of the nervous system. As neuroscience has uncovered more about the workings of the brain, scientists, clinicians and philosophers have found progressively more refined frameworks in which to discuss the effects of neurological dysfunction. In *From Axons to Identity: Neurological Explorations of the Nature of the Self (ATI)*, psychiatrist Todd E. Feinberg explores both clinical and physiological evidence in the hope of clarifying one of the great mysteries of modern science: How do the brain’s diverse and disparate regions produce a unified sense of self?

Feinberg’s thesis is that the self is not a singular entity but a dynamic process arising from the hierarchical structure of the nervous system. In *ATI*, he discusses hierarchical arrangements from both a psychiatric and physiological perspective. In chapters one through four, Feinberg describes a number of psychiatric conditions and hypothesizes that many disorders of the self arise from regression to an earlier stage of psychological development. As a result, we observe in some mentally ill adults psychological coping behaviors typically seen only in children. This suggests, according to Feinberg, that disorders of the self are in fact a loss of higher-level brain function allowing younger, more immature mechanisms to take hold. In the remaining chapters, Feinberg argues that the hierarchical arrangement of the brain is unique in nature and that this uniqueness explains the “hard” problems such as mental unity, the subjective experience of qualia, and intentionality. Though the subject matter of the first and second halves of *ATI* is quite different, the concept of hierarchical organization is the theme that unifies the entire book.

Altogether Feinberg provides an interesting perspective on the concept of self, though some of his conclusions are too strong. He glosses over many of the deeper philosophical issues related to the “hard” problems of consciousness, demonstrating the danger in making strong claims about consciousness at this stage of the game. That being said, Feinberg’s clinical expertise provides an
interesting perspective. The case studies in particular are fascinating. Feinberg quotes extensively from both his and others’ interviews with their patients giving his readers a first-person perspective of mental illness that is touching as well as intellectually stimulating. *ATI* is accessible for lay readers and could also be a valuable resource for university instructors teaching abnormal psychology or cognitive neuroscience at the undergraduate level. I would recommend it to professional scientists and philosophers interested in clinical perspectives on cognition.

**Psychopathology and the Physical Basis of the Self**
Feinberg begins by surveying numerous case studies of patients with disorders of the self arising from trauma, such as stroke, or dementia. Broadly, disorders of the self can disrupt knowledge of one’s own body, produce an abnormal sense of familiarity or non-familiarity with objects and people, corrupt recollection of one’s own past, or break apart the typically cohesive self. Among other things, Feinberg’s descriptions of these patients drive home how the self is inseparable from the brain, and it is disquieting to know that we can “lose ourselves” through trauma or the gradual breakdown of our brain due to age. In two particularly striking examples, one patient insists that paralyzed limbs still work while another develops strong feelings towards a “person” who is in fact her own mirror image.

The powerlessness one faces in disorders of the self is especially disturbing. Terrible though paralysis may be, it is at least something that one can confront and, perhaps, learn to live with. But the sincere belief that one can lift an inoperable limb is something else entirely; under certain circumstances, we can lose our selves and no exertion of will can retrieve what is lost. Outside the enclaves of neuroscience, cognitive science and philosophy of mind, many if not most people instinctively believe in some form of dualism – the assertion that there is distinct “mental stuff” and “physical stuff”. Feinberg’s discussion of these case studies brings into relief the inseparability of the mental and the physical. For many lay readers, this might come as a jarring reminder of just how susceptible our rich inner lives are to the vicissitudes of the physical world. And while physicalism – the belief that there is only physical stuff and the mind emerges out of it – is nothing new or radical in philosophical circles, it is nevertheless worthwhile to read first-hand accounts of people with damaged selves. These cases remind us of the fragility of the physical self – something that is often lost in technical discussions.

**The Hierarchical Development and Dysfunction of the Self**
In chapters three through four, Feinberg develops the hypothesis that a mature ego emerges as progressively more refined psychological mechanisms for navigating the world come into being. In this way the self develops in a hierarchical manner. “Immature” mechanisms such as fantasy are not lost as one makes the transition from childhood to adulthood, rather they become part and parcel of the maturing self. Adults are certainly capable of fantasy, but in healthy adults indulgence in fantasy is modulated by restraints that maintain a
clear distinction between what is real and what is imagined. For this reason, we think little of it if a young child sincerely believes in Santa Claus but this belief is cause for concern in an older child and indicates psychosis in adults.

Of course children can transition between the world of fantasy and that of reality; however, the intensity of childhood fantasies differentiates children from adults. When a child enters into a fantasy world, that world becomes the child’s reality – at least for a time – whereas a healthy adult does not conflate his or her daydreams with what actually is. The strength of childhood fantasies could explain, for instance, how children develop intense relationships with imaginary friends or enter a world in which they have special powers. It is also noteworthy that intense fantasy in children can serve as both a type of play or as a psychological defense mechanism. An imaginary companion may, for instance, provide a means by which a child develops creativity and social skills, but can also serve to ameliorate loneliness or as a scapegoat whom the child blames for his or her own actions. Feinberg emphasizes that if an adult becomes completely immersed in fantasy, the indulgence in imagination reflects pathology and not normal psychological development.

Feinberg presents a formal ordering of psychological coping mechanisms adapted from Vaillant [1]. At the first level are “psychotic mechanisms”, those that involve complete denial of reality and typically disappear around age five in normal development. The occurrence of level one mechanisms in adults almost always indicates pathology. At level two we see “immature mechanisms” such as projection hypochondriasis and simple “acting out”. Such behaviors are often observed in adolescents, as well as adults who suffer from depression or personality disorders. At the third level, there are “neurotic mechanisms” which, while not necessarily healthy, are seen in most adults to some degree. These include repression and reaction formation. The highest level, so-called “mature mechanisms”, are those employed by mentally balanced adults to deal with stress, complicated emotions and psychological strain. Altruism and use of humor are examples of mature mechanisms.

Feinberg argues that the emergence of mature mechanisms in late adolescence and early adulthood reflects the progressive development of the brain. As “higher” level processing centers such as the prefrontal cortex develop, the brain is more able to exert inhibitory control and employ feedback mechanisms to modulate its own behavior. In a sense, the part of our mind that cannot distinguish between reality and fantasy is never really gone. Immature aspects of the self simply come under the control of more mature mechanisms that dampen their influence on a person’s behavior. Loss of higher-level mechanisms does not, in Feinberg’s view, produce new behaviors. Rather, such damage allows previously suppressed mechanisms to reemerge in adults.

It follows from the framework described above, and Feinberg argues convincingly based on clinical data, that there exist many similarities between normal behavior
in children and pathological behavior in adults. For instance, patients who insist that paralyzed arms still function are comparable to children who create fantasy worlds when their wishes conflict with reality. To cite another example, Feinberg describes adults who become extremely attached to imaginary friends. In two tragic cases one elderly patient prepares a place setting for a teddy bear while another attempts to feed a photograph of her dead husband. The first patient describes her stuffed animal as a lively companion while the second adamantly avers that her husband is not dead. From these and other examples, Feinberg argues that adults with certain mental illnesses, much like children, use denial of reality as a compensatory mechanism because they have lost the ability to utilize higher-level coping strategies.

Two Types of Hierarchies
Having argued that the self evolves hierarchically from childhood to adulthood and that disorders of the self come about when higher levels of the hierarchy break down, Feinberg devotes the remainder of ATI to a discussion of the organization of the brain and how this organization dovetails with a hierarchical view of the self.

Feinberg differentiates two types of hierarchies: non-nested and nested ones. A non-nested hierarchy is very much like a vertical flow chart, with higher-level nodes exerting influence on the nodes directly beneath them. Such hierarchies are often compared to a military structure in which the highest-ranking officer’s orders influence behavior at lower levels of the system while at the same time lower ranking members of the hierarchy provide information to members at the level above their own. The key feature of a non-nested hierarchy is the distinct separation between different levels. In contrast, nested hierarchies are those in which higher levels are composed entirely of lower level constituents. For instance, a cube of ice is composed of water molecules, which are themselves amalgamations of hydrogen and oxygen. Of course a water molecule isn’t an ice cube anymore than an oxygen atom is a water molecule. Though the levels are non-identical, they are compositionally dependent upon one another in ways that the members of non-nested hierarchies are not. Feinberg argues that the brain displays the traits of both a non-nested and nested hierarchy and that this unique feature explains the problems of unity, qualia, and intentionality.

Hierarchies and the Self
In chapter five of ATI, Feinberg describes the hierarchical evolution of the nervous system. It is largely a review of basic neural development that sets the stage for chapter six. While much of the material is review for professionals and advanced students, it will provide a challenging yet comprehensible introduction for non-specialists.

In chapter six Feinberg argues that as new structures develop and integrate themselves with those that already exist, we observe the formation of both nested and non-nested hierarchies with the emergence of non-nested hierarchies
as the crucial ingredient for consciousness. In Feinberg's scheme, organisms are divided into four categories with respect to their capacity for consciousness.

At the first level, we see simple organisms with no central nervous system. These organisms are, in Feinberg’s view, best thought of as purely nested systems as each level is composed entirely of the elements below it. For instance, proteins make up organelles, which make up cells, which make up organisms. There is no centralized control, only parallel processes.

At level two are organisms with simple nervous systems. These organisms are capable of stimulus response behaviors but not reflection or internal representation of their environment. Though the levels of such organisms’ nervous systems are structurally distinct, they do not display robust non-nested hierarchies in that there is little or no centralized control. Though there is some vagueness regarding the cutoff between the emergence of nested and non-nested hierarchies in the nervous system one can grant Feinberg these distinctions for argument’s sake. Moreover, the evolutionary trends he describes are well established as we see not only an increase in brain volume but changes in the ratio of volumes between different brain structures as we move “up” the phylogenetic tree [2]. Such evidence bolsters Feinberg’s claim that, at some point in evolution, we see the emergence of convergent and divergent information processing in the brain as opposed to the simple parallel processing observed in simpler organisms.

Consciousness emerges at levels three and four. Level three organisms are those with consciousness but not self-awareness, while level four organisms have a well-developed sense of self. In addressing these issues, Feinberg surveys a number of issues in neuroscience and psychology including perceptual binding – the ability of the brain to form unified perceptions out of multifarious inputs – and the brain’s capacity to influence its own behavior. In discussing perceptual binding, Feinberg argues that the “wiring” of the brain alone is not enough to explain the emergence of unified percepts. He presents evidence that other phenomena, such as synchrony in the firing of neurons and coherent oscillations in brain waves, facilitate binding. Both non-nested and nested hierarchies might contribute to such coherence across the brain as high-level nodes in non-nested hierarchies exert simultaneous control over multiple nodes while the elements of nested hierarchies tend to operate in tandem due to their compositional dependence on one another. Feinberg also argues that the brain’s ability to control its own behavior depends critically on properties of non-nested hierarchies. Because they can exert top-down control, non-nested hierarchies allow animals with advanced brains to reflect and choose how to react to stimuli, instead of producing the rigid behavioral patterns seen in organisms with simpler nervous systems. Such coordination allows organisms with highly evolved brains to plan actions based on an overarching goal rather than simple stimulus response.
The Hard Problems

In the concluding chapters of *ATI*, Feinberg grapples with the so-called hard problems of consciousness: mental unity (the emergence of a unified self out of a heterogeneous brain), qualia (subjective experiences such as color and taste) and intentionality (the ability to have mental states such as desire and fear that are directed at external things). He attempts to resolve these problems by applying the hierarchical view of the self presented in the previous chapters, arguing that the self is not as a static entity but rather a dynamic process which emerges from sufficiently complex systems.

Feinberg deals with mental unity both with intuitive and scientific evidence. Though our subjective sense of the unity of our own self is strong, arguments from intuition alone are slippery. Non-Euclidean geometry, for instance, requires considerable intellectual gymnastics to grasp because its postulates are so counter-intuitive. Yet it is only within a non-Euclidean framework that Einstein’s theories of space-time hold. Perhaps with the limits of intuition in mind, Feinberg devotes more space to a discussion of scientific evidence and presents an interesting assessment of disorders in which the unified self breaks down. For instance, Feinberg discusses cases of alien hand syndrome in which a limb becomes divorced from its owner’s will (much like the right hand of the title character in the film *Dr. Strangelove: Or How I Learned to Stop Worrying and Love the Bomb*). These cases, Feinberg argues, demonstrate that the unified self is a property of a physical system given that damage to the system disrupts this unity. This reasoning, however, only provides evidence that physicalism is likely correct. To know that damage to the system causes changes in its properties tells us *what* a system does. But, without a proper theoretical framework in which to interpret such evidence, it tells us little about *how* a system accomplishes its task. To deal with the “how”, Feinberg argues that the qualities of nested hierarchies allow us to account for mental unity as each higher level of the brain is composed of the lower levels, thus providing a natural means by which the brain creates a unified self. While interesting, this theory is not entirely novel, nor is it without problems of its own.

Feinberg extends his theory of mental unity when dealing with qualia. As the unified self emerges out of the interactions between different levels of the brain, Feinberg argues that qualia are nothing more and nothing less than the result of interactions among nodes arranged in a sufficiently complex hierarchy. The argument seems very similar to arguments from the concept of supervenience, which posits that mental properties (M) supervene on physical properties (P) when and only when it is impossible for systems with the same M properties to have different P properties. This line of reasoning is not, however, without problems. Kim [3], for instance, argues that mental properties are epiphenomenal in this scheme, as they are ontologically inseparable from physical properties. Therefore, mental properties have no intrinsic qualities or causal power that are not accounted for by their physical correlates. Granted, traditional formulations of supervenience do not take into account the interactions
between nested and non-nested hierarchies that Feinberg describes. Such interactions, however, still do not perform all that we need for a complete physicalist account of the self. The argument still hinges on the supposition that subjective experience either supervenes upon, or is an emergent property of, a particular type of physical system. In this account, as is the case with Feinberg’s explanation of mental unity, we do not have a complete picture of how a system accomplishes its task. Without such an account, the ability of a physical entity of any complexity to experience the subjective aspects of eating an apple, watching a sunset or falling in love remains as shrouded in mystery as ever.

Finally, in discussing intentionality, Feinberg once again posits that it emerges out of the interactions between the various levels of neural hierarchies. He argues that, like qualia, intentionality is non-reducible to the constituent neural events that create it. Ignoring for a moment that strong emergentism is by no means universally accepted by the cognitive science community, this account of intentionality is unsatisfying for the same reason that Feinberg’s discussion of qualia does not close the issue. If we accept that intentional states simply arise out of the interactions among multiple levels of a hierarchy then we still have no account of why a particular thing should “feel” as it does, nor do we have a coherent philosophical explanation of how a physical entity feels anything. Furthermore, if we continue entertaining arguments from emergence, then Kim’s criticisms still apply and must be answered.

In his discussion of all three aspects of the hard problem, Feinberg appeals to the complexity of a system in which nested and non-nested hierarchies operate in tandem. In all three cases, however, there is still no plausible description of how such a system generates consciousness. Such an account of consciousness is tantamount to the understanding of gravity available in Newton’s time; Newton and his contemporaries understood that massive objects attract one another – this is the what – but they could not formulate any coherent story about action at a distance – this is the how. How gravity works hinges on Einstein’s insight that massive objects can bend space-time, a concept that could did not fit into any mathematical framework available to Newton. Likewise, it is probably the case that Feinberg, like anyone currently grappling with these issues, lacks a framework in which to explain how the brain does what it does.

The Jury is Still Out

Though Feinberg has done an admirable job grappling with the hard problems of consciousness his primary conclusions are premature. First and foremost, in all three aspects of the hard problem his argument seems to come down to the following: all the richness and subjective qualities of our inner lives arise from complex interactions among brain regions but are not reducible to these interactions. However, this seems to be a reiteration of strong emergentism, which is not without its detractors. Furthermore, Feinberg consistently states as a matter of fact that the subjective aspects of self simply arise from the hierarchical structure of the brain. This story is incomplete without a coherent
physiological or metaphysical account of how interactions between hierarchies accomplish this task. Without such an account any appeal to hierarchies or any other aspect of the brain simply begs the question.

Finally, many of the ideas present in *ATI* are still topics of on-going discussions in the cognitive science literature. Feinberg’s reliance on strong emergentism would have been strengthened by an attempt to refute its critics. Furthermore, Feinberg’s views of the mind are not entirely novel. The notion of the mind as an ever-evolving dynamic system is seen in work by Van Gelder [4], and further developed in the mechanist account of philosophers such as Bechtel [5] and the dynamical systems theories advocated by Randall Beer and others. Finally, Lucas [6] argues that human minds have a unique ability to perform self-referential reasoning in his critique of strong AI theory, and Hofstadter [7,8] has developed in detail the concept that minds arise out of hierarchical systems that loop back upon themselves.

In *ATI*, Feinberg provides worthwhile food for thought. When sticking to discussions of clinical cases, *ATI* is lively and stimulating reading. Criticisms of the arguments regarding consciousness not withstanding, Feinberg presents interesting concepts in a non-technical and clear style, making *ATI* a valuable resource for lay readers. Feinberg’s insights from clinical experience provide fresh and engaging material for cognitive scientists and philosophers of mind.

**Acknowledgments**
Thanks to Colin Allen, Eric Anderson, Jr. and Robin Murphy for critical readings of this review.

**References**


