Journal of the History of the Behavioral Sciences, Vol. 38(1), 3–25 Winter 2002

(PHYSIO)LOGICAL CIRCUITS: THE INTELLECTUAL ORIGINS OF THE McCULLOCH–PITTS NEURAL NETWORKS

TARA H. ABRAHAM

This article examines the intellectual and institutional factors that contributed to the collaboration of neuropsychiatrist Warren McCulloch and mathematician Walter Pitts on the logic of neural networks, which culminated in their 1943 publication, "A Logical Calculus of the Ideas Immanent in Nervous Activity." Historians and scientists alike often refer to the McCulloch–Pitts paper as a landmark event in the history of cybernetics, and fundamental to the development of cognitive science and artificial intelligence. This article seeks to bring some historical context to the McCulloch–Pitts collaboration itself, namely, their intellectual and scientific orientations and backgrounds, the key concepts that contributed to their paper, and the institutional context in which their collaboration was made. Although they were almost a generation apart and had dissimilar scientific backgrounds, McCulloch and Pitts had similar intellectual concerns, simultaneously motivated by issues in philosophy, neurology, and mathematics. This article demonstrates how these issues converged and found resonance in their model of neural networks. By examining the intellectual backgrounds of McCulloch and Pitts as individuals, it will be shown that besides being an important event in the history of cybernetics proper, the McCulloch–Pitts collaboration was an important result of early twentieth-century efforts to apply mathematics to neurological phenomena. © 2002 John Wiley & Sons, Inc.

> Logic is concerned with the real world just as truly as zoology, though with its more abstract and general features.
>
> –Bertrand Russell (1920, p. 169)

In 1943, neuropsychiatrist Warren McCulloch (1898–1969) and mathematician Walter Pitts (1923–1969) presented one of the first applications of a logical calculus to the elements of a biological system (McCulloch & Pitts, 1943). Based on the "all-or-none" character of nervous activity, they constructed a Boolean logic to describe neural events and their relations. The all-or-none concept led McCulloch and Pitts to idealize neurons as on-off devices—they either "fired" or they did not. Connecting this to the "true–false" nature of propositions in logic, McCulloch and Pitts constructed hypothetical networks of excitatory and inhibitory neurons, with varying patterns of connection, and demonstrated an isomorphism with these hypothetical arrangements of neurons and the logic of propositions.

The McCulloch–Pitts concept of a logical neural network has been described as a landmark event in the history of cybernetics. One of the goals of the cybernetics movement was to find common elements in the functioning of animals and machines. As their paper had conceptual connections to Alan Turing's (1912–1954) 1937 work on the "Turing machine" (Turing, 1936–1937) and was significant for the later work of John von Neumann (1903–1957) (Von Neumann, 1945/1981, 1951), the McCulloch–Pitts paper clearly represents an important event, and its legacy, at least within the cybernetics movement, has been well examined by historians of science. For example, the work has been described as integral to the design of digital computers and automata, and to the development of theories of infor-

TARA H. ABRAHAM *is a postdoctoral fellow at the Dibner Institute for the History of Science and Technology at MIT. In 2000, she received her Ph.D. from the Institute for the History and Philosophy of Science and Technology at the University of Toronto, with the thesis* "Microscopic Cybernetics: Mathematical Logic, Automata Theory, and the Formalization of Biological Phenomena, 1936–1970."

mation processing (Aspray, 1990, chap. 8; Edwards, 1996, chap. 6–8; Heims, 1991, chap. 3).

The McCulloch–Pitts work has also been viewed as fundamental to modern cognitive science and neuroscience, particularly Artificial intelligence (AI) and connectionism (e.g., Leiber, 1991; von Eckardt, 1993).[1] AI, which emerged during the 1950s, adheres to a computational theory of mind: in principle, intelligent behavior can be imitated by a digital computer. Within the AI paradigm, cognition is characterized by the manipulation of symbols according to "rules," which are, in essence, a logical description of the desired behavior.[2] Connectionism, or the "neural network" approach, has flourished since the 1980s. Models here are networks of elementary units, each having a certain degree of activation, and active units excite or inhibit other units (Bechtel & Abrahamsen, 1991, p. 2).[3] Two key elements of the McCulloch–Pitts work—that of a logical description of activity and a functionally connected network of idealized neurons—are central to cognitive modeling and to modern conceptions of cognition, neural activity, and the logical organization of the brain.

But how did this new view—that of a network of logically defined neurons—emerge? As yet, the intellectual context of the McCulloch–Pitts collaboration has not been examined in detail. The aim of this article is to shed light on the constellation of concepts, disciplines, and institutional backgrounds that led to the publication of the McCulloch–Pitts paper, namely, their intellectual and scientific orientations and backgrounds, the key concepts that contributed to their paper, and the institutional context in which their collaboration was made. Cybernetics has recently been called a "symbiosis" of elements from mathematics and physiology (Marshall & Magoun, 1998, pp. 261–262). This metaphor maps well on to the McCulloch–Pitts collaboration: McCulloch was trained in neurophysiology, and Pitts had a remarkable fluency in mathematics. Providing an intellectual space for this collaboration was a group devoted to mathematical biology at the University of Chicago, pioneered by the mathematical biologist Nicolas Rashevsky (1899–1972), who saw mathematics as a powerful tool for the study of complex biological phenomena.

Although they were almost a generation apart and had dissimilar scientific backgrounds, McCulloch and Pitts had similar intellectual concerns, simultaneously motivated by issues in philosophy, neurology, and mathematics. By examining the intellectual backgrounds of McCulloch and Pitts as individuals and the institutional context of their collaboration, this article will illustrate how these issues converged and found resonance in their model of neural networks.

The Philosophical Psychiatrist: Warren S. McCulloch

In 1961, Warren Sturgis McCulloch (Figure 1) told a story about a formative event in his intellectual development. In 1917, while a student at Haverford College in Pennsylvania, a teacher asked him what he planned to do with his life. McCulloch said that he hoped to answer the following question: "What is a number, that a man may know it, and a man, that he may know a number?" The first part of the question, "What is a number?", McCulloch

1. For an examination of the cybernetic roots of modern cognitive science, see Dupuy (1994/2000).

2. For an introduction to some of the central ideas of AI, see Pratt (1987).

3. In contrast to the AI approach, in neural network models, there is no need for a precise, explicit, logical description of the desired behavior, rather, the neural net itself embodies the description implicitly, in the pattern of connection, or architecture, of the system. For detailed analyses of the contrasts between AI and connectionism, see Cowan and Sharp (1988) and Bechtel and Abrahamsen (1991, chap. 1).

FIGURE 1.
Warren McCulloch in the Navy at Yale University, 1918. From McCulloch (1989, Vol. I).

recalled, was answered by the mathematicians. The second, more difficult part of the question was to direct his life's work (McCulloch, 1965a).

By the end of the First World War, McCulloch transferred to Yale University, to join the United States Naval Reserve's training program for students, which was not offered by Haverford (R. McCulloch, 1989, p. 1). McCulloch received his bachelor's degree in philosophy and psychology from Yale in 1921, and an M.A. in psychology from Columbia in 1923.[4] According to his own recollection, during these early years McCulloch's main interest was the nervous system (McCulloch, 1965a). Eventually he went on to medical school at Columbia, receiving his M.D. in 1927. In 1928, McCulloch interned as a neurologist at Bellevue Hospital in New York City, doing experimental research on epilepsy and head injuries. In 1932, he began psychiatric training at the Rockland State Hospital for the Insane in Orangeburg, New York. Here, he worked with the German-born psychiatrist Eilhard von Domarus (1893–1958), who was to have a profound influence on his intellectual development

4. For biographical information on McCulloch based on the first-hand experience of one of his students, see Arbib (2000).

over the next few years, and whom McCulloch later referred to as "the great philosophic student of psychiatry."[5]

While at Rockland, McCulloch learned to understand the logical difficulties in cases of schizophrenia and psychopathia, not from a clinical perspective, but as von Domarus had understood them through his contact with the likes of Bertrand Russell (1872–1970), Alfred North Whitehead (1861–1947), and F. S. C. Northrop (1893–1992). In McCulloch's words, von Domarus, like McCulloch himself, was "forced into neuropsychiatry by philosophic problems" (McCulloch, 1967a, p. 350). In 1930, von Domarus had written his thesis, "The Logical Structure of Mind: An Inquiry into the Philosophical Foundations of Psychology and Psychiatry," under Northrop, with help, he acknowledged, from Warren McCulloch (Von Domarus, 1930/1967). Von Domarus was the author of many works in psychiatry, including a paper presented at the 1939 meeting of the American Psychiatric Association on "The Specific Laws of Logic in Schizophrenia" (Von Domarus, 1944). Von Domarus's approach was interdisciplinary. Well-versed in both neuroanatomy and philosophy, his goal was to connect scientific treatments of the brain with philosophical conceptions of the mind and the logical structure of reasoning. Through this, Domarus hoped to develop a "logic of intentional relations" (McCulloch, 1967b).

Filmer Stuart Cuckow Northrop, who was von Domarus's supervisor, began his graduate work at Yale in 1917 and received his Ph.D. from Harvard in 1924, with the thesis "The Problem of Organization in Biology." In 1923, he became an instructor in philosophy at Yale, remaining on the faculty for almost 40 years. An expert in philosophy, science, anthropology, and law, Northrop had an interdisciplinary approach to philosophical and scientific problems. As one biographer wrote, he "used the scientific method as his philosopher's stone" (DiPalma, 1992, p. 89). Especially notable were his studies of scientific events of the early twentieth century, and his efforts to relate these events to cultural and philosophical issues. In an examination of important events in the history of science, Northrop emphasized the importance of theory, arguing that ". . . what made a science out of chemistry was not mere observation and experiment but Lavoisier's attention to theory. . . . Not experiment alone but experiment guided by relevant theory made a science out of chemistry . . ." (Northrop, 1938, p. 213). Northrop also made a strong argument for the importance of the methodology of physics on biological science, and he believed that formal logic and mathematics played a role in the historical development of any science (Northrop, 1940). As a science develops historically, its ultimate state of maturity is achieved through the incorporation of formal, deductive methods:

> The history of science shows that any empirical science in its normal healthy development begins with a more purely inductive emphasis, in which the empirical data of its subject matter are systematically gathered, and then comes to maturity with deductively-formulated theory in which formal logic and mathematics play a most significant part. (Northrop, 1940, p. 128)

To illustrate his argument, Northrop used the example of physics, which in his words had an "inductive" or "natural history" phase from the ancient Greek period to the Middle Ages, gaining a basis in "deductively-formulated theory" with the work of Galileo and Newton. Biology, for Northrop, was at present struggling with this transition. The descriptive, classificatory stage in biology that began with Aristotle had begun to move toward a formal,

5. Von Domarus received his M.D. from the University of Jena in 1922. In 1928, he emigrated to the United States, and received a Ph.D. from Yale in 1932.

deductive stage with the work of Joseph Henry Woodger (1894–1981) (Woodger, 1937) and Nicolas Rashevsky (1899–1972). In these works, formal logic and mathematics played a strong role, and thus biology as a discipline was reaching a more mature stage, as it began to incorporate the scientific method of physics, that is, using theoretical analysis and mathematical formulations. Northrop's arguments for the value of formalization in biology were connected to his larger vision of "dissecting the given scientific theories which . . . scientists have verified, to determine what concepts and principles are taken as primary or undefined" (Northrop, 1931, p. xiii). Through this process, scientific theories across all branches of science, including physics and biology, could be reduced to a set of primary, foundational concepts, or "first principles." This was the goal of Northrop's philosophy of science. A greater emphasis on theory and mathematical formulation in biology would allow physics and biology to be integrated. In Northrop's words, "there can be no adequate biological or medical theory of the concrete individual until there is a verified theory of the inter-relation of the basic concepts of the sciences . . . to possess such a theory is to possess an experimentally verified philosophy of science" (Northrop, 1938, p. 231).

During his period of contact with von Domarus and Northrop, McCulloch was drawn to this "formal," philosophically motivated approach to biological problems.[6] His work was still directed by the question of knowledge and its logical foundations, and when he asked "What is a man, that he may know a number?", he was pondering the logical nature of human thought, and its physiological basis in the brain. In 1961, he recalled:

> In 1923 I gave up the attempt to write a logic of transitive verbs and began to see what I could do with the logic of propositions. My object, as a psychologist, was to invent a kind of least psychic event, or "psychon," that would have the following properties: First, it was to be so simple an event that it either happened or else it did not happen. Second, it was to happen only if its bound cause had happened . . . that is, it was to imply its temporal antecedent. Third, it was to propose this to subsequent psychons. Fourth, these were to be compounded to produce the equivalents of more complicated propositions concerning their antecedents. (McCulloch, 1965a, p. 8)

Now, what does all this mean? McCulloch later explained that the "psychon" was a "simplest psychic act." His conception of a psychon was, in his words, "what an atom was to chemistry, or a gene to genetics. . . . But my psychon differed from an atom and from a gene in that it was to be not an enduring, unsplittable object, but a least psychic event" (McCulloch, 1965b, pp. 392–393). The notion of an event occurring "only if its bound cause had happened" and proposing this to "subsequent psychons" implied the notion of a network of logically connected elements. Around 1929, McCulloch realized that these could be conceived as the all-or-none impulses of neurons. As McCulloch recalled, he "began to try to formulate a proper calculus for these events by subscripting symbols for propositions in some sort of calculus of propositions . . ." (McCulloch, 1965a, p. 9).

The next decade saw McCulloch grounding his ideas in experimental work on the brain. Following his work at Rockland, McCulloch returned to Yale, working in the Laboratory of Neurophysiology under the Dutch physiologist, Johannes Gregorius Dusser de Barenne (1885–1940).

6. Incidentally, McCulloch also came into direct contact with Northrop as early as fall 1923, at a Frank Pike graduate seminar on the theory of the nervous system. They were also in contact at the Wednesday evening seminar conducted by Clark L. Hull at Yale in 1936 (Smith, 1986, pp. 180–181).

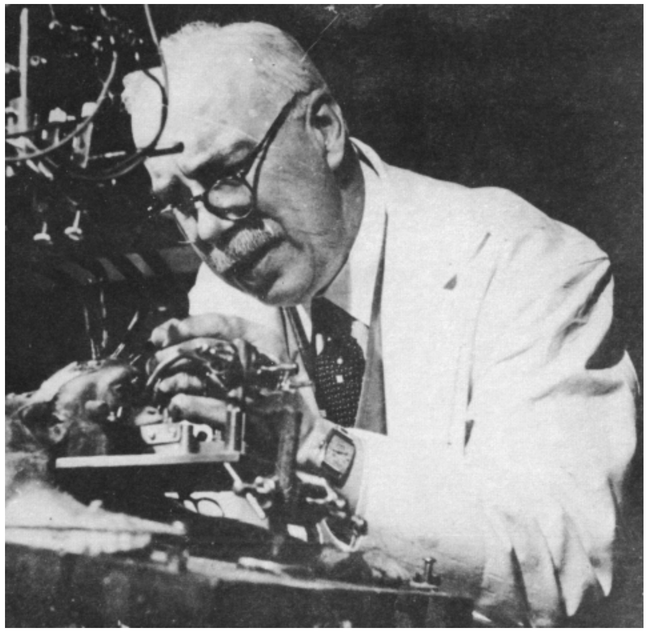


FIGURE 2.
Dusser de Barenne, 1936. Yale Laboratory, New Haven, Connecticut. From McCulloch (1989, Vol. I).

J. G. DUSSER DE BARENNE AND LOCALIZATION OF FUNCTION IN THE CEREBRAL CORTEX

Early in his career, Dusser de Barenne (see Figure 2) had done much work investigating sensory mechanisms and functions in the spinal cord, while in the Laboratory of Physiology at the University of Amsterdam (Fulton & Gerard, 1940). By 1916, his attention turned to the localization of function in the cerebral cortex, the postulated "material source" of psychological functions in the brain. This endeavor was tied to the doctrine of associationism, which was the dominant philosophy within psychology by the turn of the twentieth century (Boring, 1929, p. 67).[7] Originally, the idea was connected to the notion that the mind was composed of many separate "ideas" that were bound together to form complexes of ideas through a number of associations.[8] Within the context of late-nineteenth-century psychology, associationism concerned the functional relationship between the physiology of the brain and the psychological processes of sensation and perception. With the work of the brilliant histologist Santiago Ramón y Cajal (1852–1934) (Ramón y Cajal, 1911/1995), who produced countless detailed drawings of the cells of the nervous system, the concept of associationism could be connected to the brain physiology: connections between thoughts or ideas were believed to have a physical, biological correlate in the structure of the brain. In Boring's

7. For a description of the principle of associationism as formulated by the eminent nineteenth-century philosopher-psychologist William James (1842–1910), see James (1890, pp. 550–604).

8. The classic historical account of the nineteenth-century origins of cerebral localization is Young (1970). For a philosophical analysis of the method of localization as central to mechanistic explanations in science, see Bechtel and Richardson (1993).

words, “it was supposed that [in the brain] the fibers merely formed a complicated network . . . and that the physiological account of mind was somehow to be gained from a further knowledge of this network” (Boring, 1929, p. 66). The more well developed knowledge of the structure of the brain became, experimental efforts were made to seek localization of mental functions in the brain (pp. 67–68).

At the start of the twentieth century, two methods of studying localization were prominent: the “lesion” or “extirpation” method and the method of electrical stimulation. In the lesion method, the loss of tissue in specific areas of the brain was related to loss of function. In the method of electrical stimulation, different parts of the brain were subjected to an electrical current, and sensory and motor functions were mapped onto the brain depending on the location of stimulation. In Dusser de Barenne’s view, the method of stimulation had yielded no significant results concerning the problem of localizing sensory functions in the cortex, and he advocated the use of the strychnine method—the local application of strychnine to the cerebral cortex. He saw several advantages of the strychnine method over the extirpation method. First, the strychnine method caused symptoms of excitation with certainty and allowed an easier interpretation of symptoms. In contrast, the extirpation method only resulted in symptoms of impairment of sensation, which were often vague. The strychnine method was also simpler and induced less stress in the animal under study, while the extirpation method was often accompanied by subsidiary pathological changes that confused results. Unlike the extirpation method, the strychnine method could be used without damaging the brain, and could be used to produce precise results with relatively short experiments (Dusser de Barenne, 1916, pp. 357–360).

Dusser de Barenne’s early work using the strychnine method involved experiments on the cerebral cortex of the cat. The cat was first anaesthetized using chloroform and ether. The region of the cortex to be experimented on was then exposed, and any excess cerebro-spinal fluid was absorbed by dabbing the surface of the cortex with cotton. A 1% strychnine solution, colored with toluidin blue, was applied to the cortex using a tiny wad of cotton wool at the end of forceps, and any excess strychnine solution was removed. The resulting poisoned spot on the cortex was then seen as a small blue area of a few square millimeters. The cat’s skin was then stitched back, to prevent cooling (Dusser de Barenne, 1916, p. 360). Following recovery from narcosis, one could then observe and compare symptoms when the sensory cortex was stychninized within a certain region of the cerebral cortex and outside this same region, observing disturbances in the cat such as paralysis and hypersensitivity. In the spring of 1924, Dusser de Barenne went to the laboratory of Charles Scott Sherrington (1857–1952) at Oxford, to study sensory symptoms through the application of strychnine to the cerebral cortex of rhesus (macaque) monkeys (Dusser de Barenne, 1924). He refined his technique here and produced results that delimited the sensory cortex of the monkey.

Dusser de Barenne came to Yale in September 1930 as Sterling Professor of Physiology in the School of Medicine, and in 1934 began collaborating with Warren McCulloch. Broadly speaking, their work focused on the influence of one cortical area upon another, and the interaction between different areas of the cerebral cortex; that is, on cortico-cortical connections was well as straightforward localization.[9] Their early collaborations at Yale’s Laboratory

9. Dusser de Barenne was critical of the “classical” localization theory, with its assumption of a “sharp, point to point, geometrical projection of the body on the cortex.” (Dusser de Barenne, 1934, p. 96). Viewing the functional organization of the cerebral cortex as complex and plastic, in 1934, he argued that “with regard to the cortical representation of the somatic functions, there is not one type of functional localization in the cortex, but more, perhaps as many as there are senses” (Dusser de Barenne, 1934, p. 103).

of Neurophysiology involved performing electrical stimulations of the motor cortex of the monkey, to study the phenomenon of cortical "extinction" or inactivation.[10] In 1936, the pair published their first joint work employing the strychnine method (Dusser de Barenne & McCulloch, 1936a). Here, through the coupling of the strychnine method with the recording of action potentials, they established that there were functional boundaries between the main subdivisions of the sensory-motor cortex. They observed that the application of strychnine was accompanied by large, rapid changes in the action potentials (which they termed "strychnine spikes") recorded from specific areas of the cortex; and that the spikes were dissimilar when recorded from different areas of the cortex. Dusser de Barenne and McCulloch also found that strychninization of a small area in one of the subdivisions of the sensory cortex, for example the arm area, could result in changes in action potentials of the whole subdivision. They observed that if local strychninization was performed in some area of the sensorimotor cortex, the "spikes" recorded were not restricted to this area but were also found in other areas. Further, the distribution of these spikes was specific for each area strychninized. And finally, their work revealed that there were *directed* functional relations between areas: for example, they observed that if one strychninized region A, spikes were recorded from region B, but if region B was strychninized in a separate experiment, no spikes were recorded from region A. Their work confirmed Dusser de Barenne's earlier hypothesis that complex functional relationships exist between different areas of the cortex (Dusser de Barenne & McCulloch, 1936b, 1938a, 1938b; Dusser de Barenne, McCulloch, & Ogawa, 1938). Dusser de Barenne and McCulloch concluded that, for certain areas of the cortex, the effects of strychnine superseded architectonic or structural boundaries of the cortex but respected functional boundaries.

Dusser de Barenne and McCulloch (1939) also used the strychnine method for delimiting neurons in the cerebral cortex, a procedure called "chemical neuronography." Although their method was similar to that used in their earlier work, their goal here was to understand communication in the cortex by deducing specific pathways of neural impulses. Based on their previous work mapping functional areas in the cerebral cortex, Dusser de Barenne and McCulloch aimed to correlate these findings with the neuronal structure of the cortex; to determine if, in the strychninized area, neurons originate that *end* in the area where electrical activity is recorded. Drawing on some neuroanatomic evidence regarding the direction of neuronal connections in the cortex, Dusser de Barenne and McCulloch concluded that when one strychninized a particular region A, and recorded "spikes" from region B, the neurons that were strychninized in region A had an ending in region B. They argued that local strychninization coupled with the recording of action potentials was a powerful tool for delimiting the origins and endings of neurons in the central nervous system.

It was under Dusser de Barenne's influence that McCulloch was able to pursue his quest for an "experimental epistemology"—the idea that physiology can explain knowledge. Between 1918 and 1930, Dusser de Barenne had worked as a Lecturer and Privat Dozent in the Departments of Pharmacology and Physiology at Utrecht University, along with Rudolf Mag-

10. This was not cortical *inhibition,* which was cessation of a response to stimulation if a second, "antagonistic" part of the cortex was simultaneously stimulated. Inactivation described the phenomenon of decreased response to stimulation of the cortex. In an early study, if the cortex was stimulated in succession, with a 13-second interval, the second response was much smaller (Dusser de Barenne & McCulloch, 1934). Dusser de Barenne's and McCulloch's early work also examined the phenomenon of facilitation, that is, the "appearance" or increase of response if the second stimulation occurred within only a few seconds after the first (McCulloch & Dusser de Barenne, 1935). From these stimulation experiments, Dusser de Barenne and McCulloch concluded that extinction was a highly localized phenomenon, most probably connected to the pyramidal cell layer of the cortex.

nus (1873–1927). Magnus had become well known for his theory of the "physiological *a priori.*" He argued that Kant's synthetic *a priori* should be interpreted "philosophically-psychologically," that is, as an aspect of the psyche, and that the *a priori* must have a physiological basis. This was related to the notion that one does not come to sensory data as a "blank tablet," but rather brings a sort of relational structure within the nervous system to interpret sense data. The nature of sensory impressions is determined *a priori*, by the physiological sensory apparatus of the brain (Magnus, 1930). Through his work with Dusser de Barenne, McCulloch was informed by a similar mingling of philosophical and physiological concerns. Although McCulloch performed many productive experiments on the primate brain, gaining extensive training in the anatomy and physiology of the brain, his earlier preoccupation with more philosophical questions remained. In McCulloch's view, the question "What is a man that he may know a number?" still formed the basis of his work here. As opposed to classical epistemology, McCulloch aimed to base a theory of knowledge on experiment, to found a physiological theory of knowledge. In his work with Dusser de Barenne on localizing function in the cerebral cortex and mapping neuronal connections within the central nervous system, McCulloch was able to relate the psychological functions of sensation and perception to the neurophysiology of the brain.

McCulloch collaborated with Dusser for six years, and they published over 20 papers together. Their last joint publication, on the sensory cortex in the chimpanzee, appeared in 1940. Here they collaborated with Hugh W. Garol and Percival Bailey, who was then on leave from the Department of Neurology and Neurosurgery at the University of Illinois (Bailey, Dusser de Barenne, Garol, & McCulloch, 1940). In 1941, after Dusser de Barenne's death the previous year, McCulloch, along with others from the lab at Yale, was invited by Bailey to join the Illinois Neuropsychiatric Institute (NPI) in Chicago. The NPI, having nine floors and a basement, has been described as "the largest and most complete neurophysiological unit in the world" (Hughes, 1993). During the 1930s, Bailey had organized an informal "neurology club" in Chicago, which brought together local scientists who were interested in the nervous system (Marshall, 1987). Bailey had hoped to establish a division of neurosciences at the University of Chicago, but when the administration proved unsupportive Bailey moved to the University of Illinois in 1939, and became the director of the newly formed Neuropsychiatric Institute (Blustein, 1992). When McCulloch came to work at the NPI, he was also named Professor of Psychiatry at the University of Illinois. It was here that he met the young Walter Pitts.

The Young Mathematician: Walter F. Pitts

Walter F. Pitts (see photo in Figure 3) has been called "the real driving intelligence" behind Warren McCulloch (Cowan, 1998, p. 104).[11] An autodidact, he taught himself logic and mathematics and was able to read Latin, Greek, and Sanskrit at an early age. Much of what we know about Pitts has come from the recollections of Jerome Y. Lettvin (1998a), who met Pitts, his "life-long friend," at the University of Chicago, where they became inseparable. Lettvin relates how Pitts had come into contact with Bertrand Russell's work at an early age, providing his entry into the world of mathematical logic.

> At the age of twelve [Pitts] was chased into a library by a gang of ruffians, and took refuge there in the back stacks. When the library closed, he didn't leave. He had found

11. For a vivid account of Pitts's life, see Smalheiser (2000).

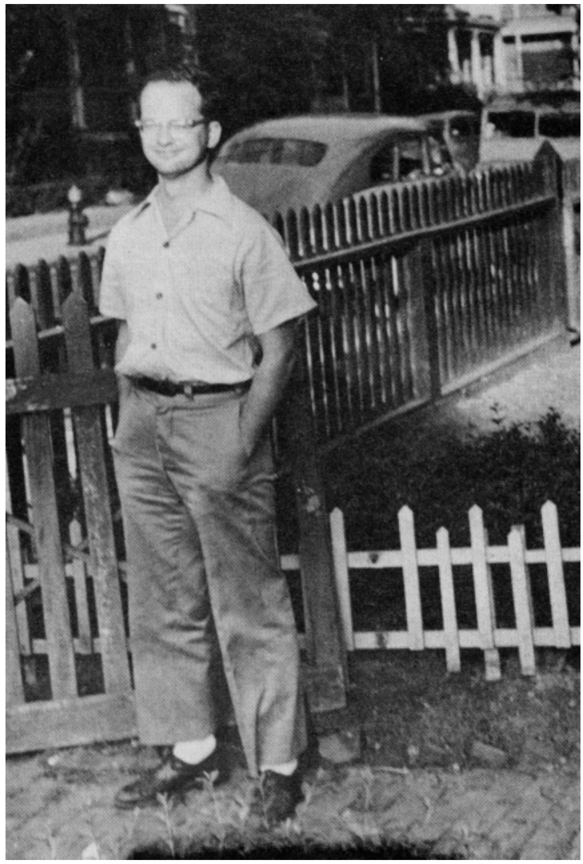


FIGURE 3.
Walter Pitts, 1949. From R. McCulloch (1989, Vol. II).

> Russell and Whitehead's *Principia Mathematica.* He spent the next three days in that library, reading the *Principia,* at the end of which time, he sent a letter to Bertrand Russell, pointing out some problems with the first half of the first volume; he felt they were serious. . . . A letter returned from Russell, inviting him to come as a student to England—a very appreciative letter. That decides him; he's going to be a logician, a mathematician. (Lettvin, 1998b, p. 2)

There are several, not entirely conflicting, stories about Pitts's connections to Rudolf Carnap (1891–1970) and Russell during this time. According to one story, Russell, who was on sabbatical at Chicago, met Pitts in Jackson Park, and took him to meet Carnap, who was in the philosophy department at Chicago. Pamela McCorduck reports this story as follows:

> Walter Pitts was forced to drop out of high school by his father, who wanted him to go to work and earn money. Rather than do this, young Pitts ran away from home and ended up in Chicago, penniless. The fifteen-year-old boy spent a lot of time in the park, where he met and began to have conversations with and older man he knew only as

> Bert. When Bert detected the boy's interests, he suggested that young Pitts read a book that had just been published by a professor at the University of Chicago by the name of Rudolf Carnap. Pitts did, and showed up at Carnap's office. "Sir," he said, "there's something on this page which just isn't clear." Carnap was amused, because when *he* said something wasn't clear, what he meant was that it was nonsense. So he opened up his newly published book to where young Pitts was pointing, and sure enough, it wasn't clear; it was nonsense. Bert turned out to be Bertrand Russell. (McCorduck, 1979, pp. 73–74, n.1)

Whatever the details may have been, through Russell and Carnap, Pitts ended up attending classes and seminars at the University of Chicago, but never registered as a student.[12] In 1938, at the age of 15, he read Carnap's latest book on logic (Carnap, 1938). Again, Pitts found flaws with it, and pointed them out to Carnap, who was at Chicago at the time. This led to Carnap hiring Pitts for "some menial job." It was through Carnap that Pitts met Nicolas Rashevsky, who took Pitts in as part of his group on mathematical biology, and who held weekly seminars on the subject at the University of Chicago (Cowan, 1998, pp. 104–105). This, according to Lettvin, was the only department Pitts ever called home.

Rashevsky's Project in Mathematical Biology

Mathematical descriptions of the behavior of nerves and networks of nerves became prominent in the 1930s and early 1940s, particularly with the work of Nicolas Rashevsky (see photo in Figure 4) and his group at the University of Chicago. Born in Chernigov, Ukraine, in 1899, Rashevsky obtained his doctorate in theoretical physics at the University of Kiev in 1919. One of his most prominent students, Robert Rosen (1934–1998), reported that given Rashevsky's *bourgeois* origins and that he had fought in the White navy during the Revolution, his progress inside the nascent Soviet Union was difficult. Rashevsky eventually emigrated from Russia and in 1920 taught physics at Robert College in Istanbul; in 1921 he became professor of physics at the Russian University in Prague (Rosen, 1972). Immigrating to the United States in 1924, Rashevsky worked as a physicist at the Westinghouse Research Laboratories in Pittsburgh, and as a lecturer in physics at the University of Pittsburgh. It was here that Rashevsky's interest in biology emerged. By 1927, his work at Westinghouse had turned to colloids, polydispersed systems, and spontaneous division in microscopic droplets, and he began developing a mathematical theory to explain the process (Rashevsky, 1928a, 1928b, 1929). He began to see a resemblance between spontaneous division in droplets and the division of living cells. Motivated to learn more about biology and to try and account for the complex process of cell division through the methods of theoretical physics, Rashevsky began studying biological literature and studied laboratory work "informally" with Davenport Hooker (1887–1965), a professor of anatomy at Pittsburgh's School of Medicine (Landahl, 1965).

During the early 1930s, while still at Westinghouse, Rashevsky published several papers on a mathematical theory of nerve conduction, which built directly on his work in cell me-

12. The theoretical biologist Wilfrid Rall reports, from a conversation with Herbert Landahl, an early colleague of Pitts's, that Landahl and others had "tried repeatedly [in the mid-1940s] to help Pitts complete the [Ph.D.] degree requirements; however, Pitts was an oddball who felt compelled to criticize exam questions rather than answer them" (Rall, 1990, p. 3).

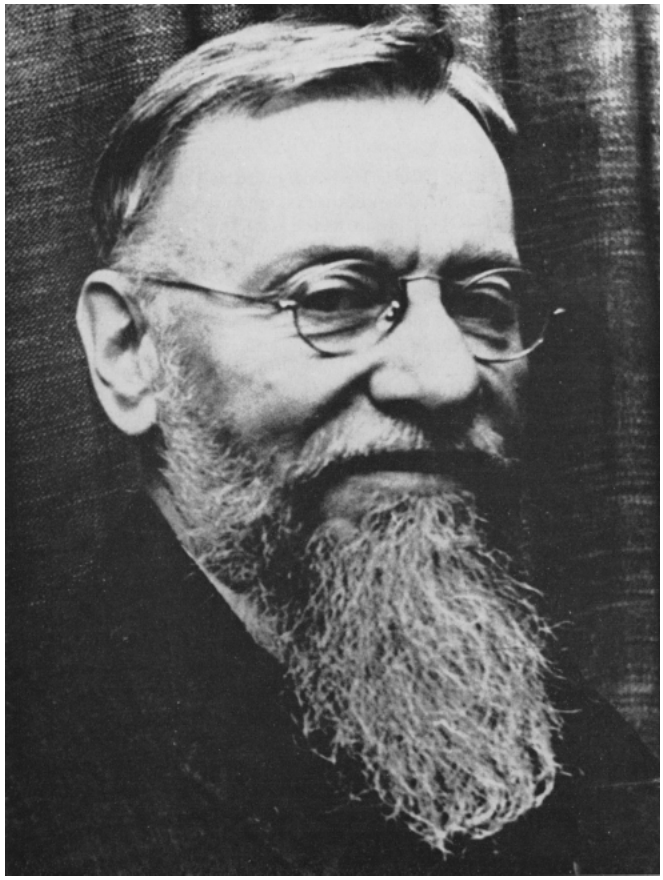


FIGURE 4.
Nicolas Rashevsky, circa 1950. From Bartholomay et al. (1972).

tabolism and division (Rashevsky, 1931a, 1931b, 1931c, 1933). Neurons, after all, were cells, and Rashevsky based his theory of nerve conduction on the notion of diffusing substances and electrochemical gradients. He began with the idea of exciting and inhibiting "substances" involved in nerve centers, and developed equations governing their diffusion and role in nerve conduction. Rashevsky's "two-factor" theory of nerve excitation, first presented as early as 1933, built on the single-factor theory of H. A. Blair (Blair, 1932). Generally, the two-factor linear theory of nerve excitation postulated that the nerve fiber develops two kinds of "substances" or "factors": excitatory substances and inhibitory substances. The rate at which each are produced is a function of the intensity of stimulus, the excess of the exciting substance over its resting value, and the excess of the inhibiting substance over its resting value. Rashevsky developed differential equations governing the relationship between the intensity of excitation, the concentrations of exciting and inhibiting substances, and several constants, which represented "empirical evidence." Altering the relative values of these constants, Rashevsky developed a mathematical theory of excitation and inhibition based on the diffusion kinetics of excitors and inhibitors.[13]

In 1934, Rashevsky was invited to the University of Chicago as a Rockefeller Fellow, for a project on "physico-mathematical methods and biological problems."[14] Rashevsky's association with Chicago is not surprising in certain respects: by this point, he had published several papers on cell division and conduction in nerves, and Chicago was a center of neurophysiological research (Blustein, 1992).

The chairman of psychology at Chicago, Louis Leon Thurstone (1887–1955), was instrumental in bringing Rashevsky to the school.[15] Rashevsky's transfer to Chicago was also facilitated through the efforts of Ralph S. Lillie (1875–1952), Sewall Wright (1889–1988), and Karl Lashley (1890–1958). In 1935 Rashevsky was appointed assistant professor of mathematical biophysics in the Department of Psychology, and eventually joined the Department of Physiology, at the invitation of Anton Julius Carlson (1875–1956), the department's chairman.[16]

According to the recollections of Jack Cowan, a prominent figure in theoretical neuroscience, Rashevsky's relationship with Carlson was somewhat strained:

> [A. J.] Carlson, who was a very famous physiologist, threw him out after a year because he never did any experimental work. The story is that Carlson went into Rashevsky's office, and there was a desk and a chair and Rashevsky, sitting there with a pencil. And Carlson said, "Where is your apparatus?" And Rashevsky said in his Russian accent, "What apparatus? I am a mathematical biologist." (Cowan, 1998, pp. 104–105)

Mathematical biology, for Rashevsky, was analogous to mathematical physics: a field that stood to experimental biology in the same way in which mathematical physics stood to experimental physics. At the time, mathematical biology was a slowly developing field. In the preface to his *Mathematical Biophysics* (1938), Rashevsky cited D'Arcy Thompson's (1860–1948) *On Growth and Form* (1917), as well as the work of Alfred J. Lotka (1880–1949) (Lotka, 1925) and Vito Volterra (1860–1940) (Volterra, 1931) on species interaction in a population of organisms, as seminal (Rashevsky, 1938, p. vii). However, Rashevsky viewed his own project as distinct from these efforts. His mathematical biology, he argued, was not merely the use of mathematics to describe biological *systems*, and the interrelations between organisms. In his own words, Rashevsky aimed to develop a mathematical biology that was a precise analogy to the use of mathematics in "molecular" physics, whereas Lotka's and Volterra's approach, in his view, was analogous to the use of mathematics in thermodynamics. For Rashevsky, "molecular" physicists dealt with *atomic* concepts rather than "gross phenomena." Rather than study the "general relations" between organisms, Rashev-

13. Others, for example, the British physiologist A.V. Hill (1886–1977) took similar mathematical approaches but they presented their work along with experimental data (Hill, 1936a, 1936b). In this paper, Rashevsky admitted that although he was not able to give an interpretation of brain phenomena that would be "altogether compatible" with present neurological and anatomical knowledge, he argued that his work demonstrated that one can *conceive* of physical systems that possess some *fundamental properties* of neurological function.

14. Rockefeller Foundation Archives, Record Group 1.1 (Projects), Series 216D (Illinois-Natural Sciences), Box 11.

15. Thurstone was also the President of the Psychometric Society in Chicago, and editor of its journal, *Psychometrika,* devoted to "the development of psychology as a quantitative, rational science." Thurstone was engaged in the development of theories of intelligence and the application of statistics, specifically factor analysis, to psychological problems. For a contrast of Thurstone's approach to mental testing with that of the French psychologist Alfred Binet (1847–1911), see Martin (1997).

16. For an analysis of Rashevsky's early work as part of a "culture of the artificial" during the 1930s, see Cordeschi (1991).

sky's mathematical biology considered the detailed structure of individual organisms (Rashevsky, 1938, p. vii).

The main object of study for mathematical biophysics was the "fundamental" living unit: the cell. And in justifying this approach, Rashevsky again referred to the mathematical methods of physics:

> Following the fundamental method of physicomathematical sciences, we do not attempt a mathematical description of a concrete cell, in all its complexity. We start with a study of highly idealized systems, which at first may not even have any counterpart in real nature. . . . The objection may be raised against such an approach, because systems have no connection to reality; and therefore any conclusions drawn about such idealized systems cannot be applied to real ones. Yet this is exactly what has been, and always is, done in physics. The physicist goes on studying mathematically, in detail, such nonreal things as "material points," . . . "ideal fluids," and so on. *There are no such things as those in nature.* Yet the physicist not only studies them but applies his conclusions to *real things.* And behold! Such an application leads to practical results—at least within certain limits. This is because within these limits the real things have common properties with the fictitious idealized ones! Only a superman could grasp mathematically at once the complexity of a real thing. We ordinary mortals must be more modest and approach reality asymptotically, by gradual approximation. (Rashevsky, 1938, p. 1, emphasis in original)

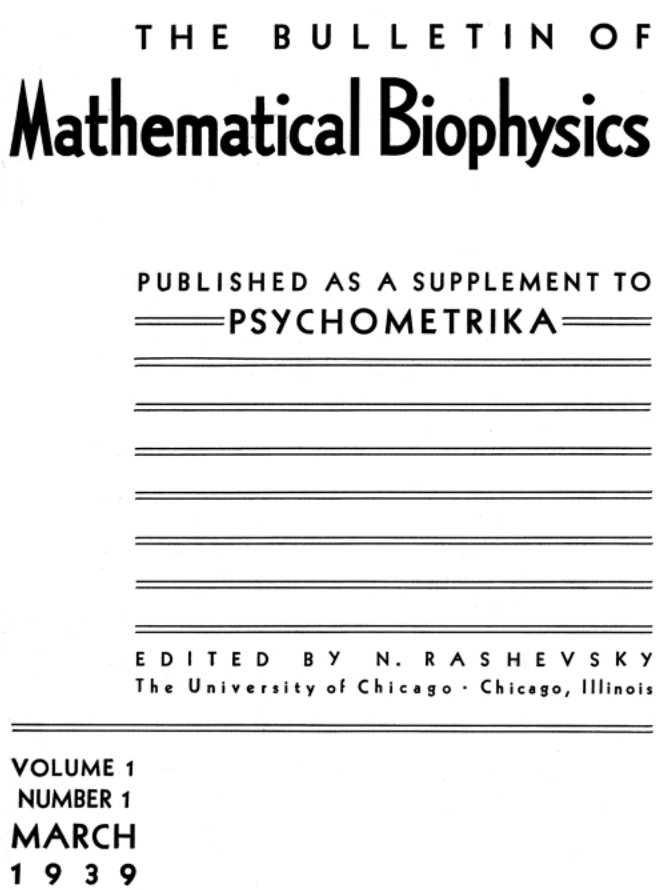
Here, Rashevsky had outlined a fundamental aspect of his project, and a clear justification of a theoretical approach to biology. Complex phenomena in biology are ubiquitous, he argued, and it is through a simplification or "idealization" that one may begin to understand such phenomena. Such an approximation may be achieved through the use of mathematics.

By the end of the 1930s, Rashevsky began an independent group for "mathematical biophysics" at Chicago. His first students were Herbert D. Landahl (b. 1913), Alston S. Householder (1904–1993), and Alvin Weinberg (b. 1915). Eventually, through the aid of the University of Chicago and the Rockefeller Foundation, he created a doctoral program in mathematical biology, and his group became known as the Committee on Mathematical Biology. Rashevsky's students began producing papers, and they needed a place to publish them. Although Rashevsky managed to find a forum for some of his earlier work in mathematical biophysics, he clearly saw a need for a journal totally devoted to this new field. In January 1939, Rashevsky approached Thurstone about the creation of a new journal devoted to mathematical biophysics. By March of that year, Rashevsky formed an agreement with the corporation that a new journal, the *Bulletin of Mathematical Biophysics* (see Figure 5), would be published by the Psychometric Corporation in Chicago, as a supplement to the quarterly issues of *Psychometrika*.[17] The *Bulletin* was a forum for mathematical treatments of psychological and neurological phenomena, and became the "classical journal in general mathematical biology," serving as a principal publication outlet for most mathematical biologists (Bartholomay, Karreman, & Landahl, 1972).[18]

Drawing on some of Rashevsky's early work on nerve conduction, a number of his Committee members developed models of excitation and inhibition in simple neural networks. However, these analyses were mostly in terms of differential equations, looking at thresholds

17. Warren Weaver Interview, January 19, 1939, Rockefeller Foundation Archives, Record Group 1.1, Series 216D, Box 11, Folder 148. Beginning in the summer of 1940, the *Bulletin* was published by the University of Chicago Press.

18. For a comprehensive list of papers published in the field of "mathematical biophysics" through to 1945, see Rashevsky (1946).


THE BULLETIN OF
Mathematical Biophysics

PUBLISHED AS A SUPPLEMENT TO
PSYCHOMETRIKA

EDITED BY N. RASHEVSKY
The University of Chicago · Chicago, Illinois

VOLUME 1
NUMBER 1
MARCH
1939

FIGURE 5.
The first cover of the *Bulletin of Mathematical Biophysics.*

and electrical currents in the excitation and inhibition of neurons. In 1936, Rashevsky had published work on simple neural networks using differential equations, linking the intensity of excitation, threshold, and frequency of impulse (Rashevsky, 1936, p. 9). Rashevsky's 1938 monograph also included analyses of excitatory and inhibitory nerve fibers connected in simple networks, with differential equations describing their activity (Rashevsky, 1938, chaps. XXII–XXVII). In a series of papers published in the *Bulletin*, Alston Householder developed a linear model of excitation based on Rashevsky's 1938 work (Householder, 1941a, 1941b, 1941c, 1942). As in Rashevsky's model, Householder aimed to connect the behavior of "complexes" of nerve fibers to the dynamic properties of the individual fibers and the structural relations among fibers in the network (Householder, 1941a, p. 63).

Joining Rashevsky's group as early as 1940, Walter Pitts took up some of the problems tackled by Rashevsky and Householder (Pitts, 1942a, 1942b, 1943). Rather than analyzing the steady-state activity of networks, Pitts was more concerned with initial nonequilibrium cases, and how a steady state could be achieved. Adopting Householder's model of neural excitation, Pitts developed a simpler procedure for the mathematical analysis of excitatory and inhibitory activity in a simple neuron circuit, and aimed to develop a model applicable to the most general neural network possible.

Clearly, Rashevsky's early monographs and papers would not have been published without the help of his students and colleagues in Chicago, who were largely members of his Committee on Mathematical Biology. In the preface to the first edition of *Mathematical Biophysics* (1938), he acknowledged the help of his students, such as Householder and Landahl. By the publication of the second edition of *Mathematical Biophysics*, in 1948, he mentioned several colleagues who had produced papers "which were published too late to be included in the book" (Rashevsky, 1948, p. xix). Included in this list were Warren McCulloch and Walter Pitts.

The Collaboration

According to his own recollection, during the 1930s McCulloch had begun attempts to formulate a logical calculus to describe the all-or-none activity of neurons, but he had no strong foundation in mathematical logic. By the time McCulloch had arrived in Chicago, Pitts was working with Rashevsky's group on mathematical biology. Early in 1942, Jerome Lettvin introduced Pitts to McCulloch, Lettvin having come into contact with McCulloch as a medical student.[19] Pitts was about 17 at the time, and McCulloch was in his mid-forties. Lettvin reports that McCulloch was "enchanted" with Pitts, and McCulloch invited both him and Lettvin to stay at his house with his family.

McCulloch and Pitts soon learned that, although from very different intellectual backgrounds, they had similar intellectual concerns. According to Lettvin, Pitts had been reading Leibniz (1646–1716), who had related the notions of computation, logic, and algorithms in the seventeenth century (Schrecker, 1947), and who had demonstrated that "any task which can be described completely and unambiguously in a finite number of words can be done by a logical machine" (Lettvin, 1998b, p. 3). In retrospect, McCulloch said that it was another "engine of logic" that had inspired him and Pitts in their neural network paper: Alan Turing's idea of a "logical machine." In 1936, Turing had developed a theoretical machine for the process of mathematical computation (Turing, 1936–1937). Simply put, Turing was able to define the complicated process of computation in "mechanical" terms, with the notion of a simple algorithm so exhaustive, rigorous, and unambiguous that the executor would need no "mathematical knowledge" to carry out its task. Turing had linked the behavior of humans and machines: in both cases, "computing numbers" involved a finite number of "states of mind" or "configurations." These "states of mind," according to Turing, were irreducible: the "operations" performed by a logic machine or a human computer can be split up into "simple operations" so elementary they cannot be further divided. This concept was central to the work of McCulloch and Pitts on neural networks; it was clear to them that the Turing machine was a "logical machine" in the sense of Leibniz. The task of computation could be described "completely and unambiguously" in a finite number of steps, and, as such, Turing's machine could be seen as an "engine of logic."

19. McCulloch himself reported that in 1941 he had presented some of his ideas on neural activity to Rashevsky's seminar, and had met Pitts here (McCulloch, 1989).

This notion fascinated Pitts. Through their discussions, Pitts and McCulloch wondered if the nervous system itself could be conceived as such a device. By the end of the year, McCulloch and Pitts had completed the essay that was to become their famous 1943 paper.

"A Logical Calculus of the Ideas Immanent in Nervous Activity"

Logic, in its most general sense, is the theoretical study of the structure of reasoning, the analysis of the language of propositions, and their logical relations (Kneebone, 1963, p. 5). The late nineteenth century saw the mathematization of logic: the rules of language, reasoning, deduction, and inference—operations of the mind—were mathematized with the development of mathematical logic, an "algebra of logic." The logic of propositions can be symbolized, with variables or symbols representing propositions or sentences, resulting in what we now call Boolean algebra, after George Boole (1815–1864), the mathematician who contributed to its development (Boole, 1854). Boolean logic is isomorphic with set theory, or Boole's logic of classes. Propositions in logic are simply sentences, or statements, that are free of any ambiguity, and, in certain contexts, are seen as either true or false but not both. As in set theory, which involves the combination of sets, in propositional logic, any two propositions p and q can be combined to form new propositions. Three typical functions (sometimes called Boolean functions) in propositional logic are conjunction (AND, symbolized by "$\cdot$"), disjunction (OR, symbolized by "$\vee$"), and negation (NOT, symbolized by "$\sim$").

By the end of the nineteenth century, the existence of the nerve cell was widely accepted. The idea that nerve cells were connected in a functional manner, and that the nervous system consisted of a "network" of neurons, was developed by cell physiologists and neurologists during the last half of the nineteenth century. Nerve fibers, bundles of nerve cells, were known to conduct electrical impulses, and, by the late 1930s, excitation and inhibition between *individual* nerve cells was directly demonstrated. The all-or-none law had emerged after the accumulation of much experimental evidence during the first quarter of the twentieth century, particularly by the British physiologists Keith Lucas (1879–1916), Edgar D. Adrian (1889–1977), and Charles Scott Sherrington.[20] Simply put, the all-or-none law states that any nerve has a finite threshold and the intensity of excitation must exceed this for production of excitation. Once produced, the excitation proceeds independently of the intensity of the stimulus. McCulloch saw a connection between the all-or-none law and his notion of a "psychon." This, along with other considerations, led McCulloch and Pitts to the observation that as propositions in propositional logic can be "true" or "false," neurons can be "on" or "off"—they either fire or they do not. This formal equivalence allowed McCulloch and Pitts to argue that the relations among propositions can correspond to the relations among neurons, and that neuronal activity can be represented as a proposition.[21]

What distinguished the McCulloch–Pitts neural network from previous concepts of net-

20. For an excellent account of the role of electrophysiological instrumentation in the development of the all-or-none law, see Frank (1994).

21. Independently of McCulloch and Pitts, in 1938, Claude Shannon (1916–2001) first showed that symbolic logic could be applied to the automatic electrical switching circuits used by communications engineers. That is, for each logical function (e.g., AND, OR, NOT), there is a *circuit* that is a *physical embodiment* of the corresponding processes of logical addition, multiplication, negation, implication, and equivalence. Shannon's 1938 paper was based on his master's thesis at MIT. In it, he showed how electronic switching circuits could be expressed in the logical symbolism of the propositional calculus, and it had important ramifications for electrical circuit design in computers (Shannon, 1938).

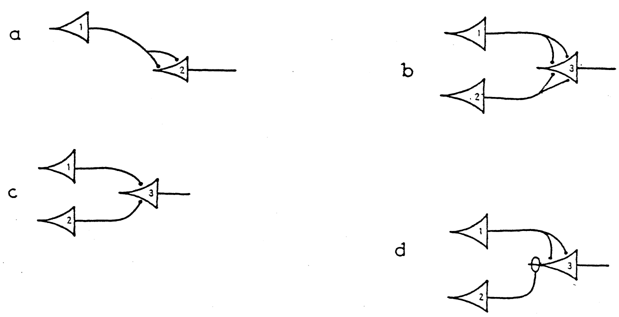
works was its basis in axioms and definitions, and their use of such axioms to construct a logical calculus of the relations between neurons. In the 1943 paper, McCulloch and Pitts began with what was, at the time, generally accepted knowledge about the nervous system. Their first assumption was that the nervous system is a network of neurons, each having a soma and an axon, and that synapses are always between the axon of one neuron and the soma of another. Their second assumption was that at any instant the neuron has some threshold, which excitation must exceed to initiate an impulse. A third important assumption was that excitation occurs mainly from axonal terminations to somata, and inhibition involves the "prevention of the activity of one group of neurons by concurrent or antecedent activity of a second group." Finally, they assumed that excitation across synapses occurs mainly from axonal terminations to somata (McCulloch & Pitts, 1943, p. 115). These were all generally accepted assumptions about neurons gained from decades of empirical investigation.

However, McCulloch's and Pitts's goal was to represent the functional relationships between neurons in terms of Boolean logic: to embody the mental function of reasoning in the actual physiology of the brain. To do this, they needed to make certain theoretical presuppositions. Most notably, they presupposed that the activity of a neuron is an all-or-none process, that is, it either "fires" or it does not. They also presupposed that the structure of the net does not change with time (McCulloch & Pitts, 1943, p. 118). McCulloch and Pitts admitted that these were both abstractions: that the activity of neurons could empirically be shown to be more continuous than discrete, and that phenomena such as "learning" could alter the structure of a net permanently, so that a stimulus which would previously have been inadequate is now adequate. But these issues did not concern them: their goal in the paper was not to present a factual description of neurons, but rather to design "fictitious nets" composed of neurons whose connections and thresholds are unaltered. For nets that can be altered by "learning," they argued that it was acceptable to substitute these "fictitious nets," which have a formal equivalence to "real" ones. They noted emphatically that they were not arguing that any "formal" equivalence between their nets and can be part of a "factual explanation" (McCulloch & Pitts, 1943, p. 117).

Figure 6 shows several simple examples of the McCulloch–Pitts neural networks and their corresponding expressions in propositional logic. In each example shown, a sum of two excitatory synaptic connections (represented in the diagram by dots adjacent to neurons) is required for a neuron to fire. An inhibitory connection is represented by an open circle adjacent to the neuron. In Figure 6(a) , neuron 2 will fire if and only if neuron 1 fires. Logically, this corresponds to the expression $N_2(t) \equiv N_1(t-1)$, which can be read as "neuron 2 will fire at time (t) if and only if neuron 1 fires at time ($t-1$)."[22] Figure 6(b) shows a network that is isomorphic with the Boolean function "OR" in propositional logic. Its expression, $N_3(t) \equiv N_1(t-1) \vee N_2(t-1)$ means that neuron 3 will fire at time (t) if and only if neuron 1 fires *or* neuron 2 fires at time ($t-1$). Figure 6(c) demonstrates the Boolean "AND" function. The expression $N_3(t) \equiv N_1(t-1) \cdot N_2(t-1)$ means that neuron 3 will fire at time (t) if and only if neuron 1 fires at time ($t-1$) *and* neuron 2 fires at time ($t-1$). McCulloch and Pitts also provided an example of the Boolean "NOT" function, with the instance of an inhibitory neuron. In the logical expression corresponding to Figure 6(d), $N_3(t) \equiv N_1(t-1) \sim N_2(t-1)$, means that neuron 3 will fire at time (t) only if neuron 1 fires at time ($t-1$) and neuron 2 *does not* fire at time ($t-1$).

Although McCulloch and Pitts spent the first and last sections of the paper reviewing

22. The logical symbol " $\equiv$ " means "if and only if."



EXPRESSION FOR THE FIGURES

a $N_2(t) \;.\equiv.\; N_1(t-1)$
b $N_3(t) \;.\equiv.\; N_1(t-1) \vee N_2(t-1)$
c $N_3(t) \;.\equiv.\; N_1(t-1) . N_2(t-1)$
d $N_3(t) \;.\equiv.\; N_1(t-1) . \sim N_2(t-1)$

FIGURE 6.
Examples of the McCulloch–Pitts neural networks. From McCulloch and Pitts (1943, p. 130). The dots on either side of the " ≡ " symbol act as separators.

the *physiology* of neurons and the implications of their model for psychology and psychiatry, the bulk of the paper is devoted to *mathematical* considerations. In fact, their only citations are to the work of mathematicians or logicians. Almost anticipating the criticisms of more experimentally oriented neurologists, McCulloch and Pitts made it clear that they do not wish to extend the *formal* equivalence of the behavior of neural networks and the logic of propositions to *factual* equivalence, noting that continuous changes as well as discrete ones are relevant.[23]

CONCLUSION

The 1943 networks were only "possible" and "useful"—in no way did McCulloch and Pitts claim their model was a true description known networks.[24] McCulloch and Pitts acknowledged in their paper that their definition of a neuron was idealized, and that they made physical assumptions that were "most convenient for the calculus" (McCulloch & Pitts 1943, p. 116). Their method here was to begin with theoretical presuppositions and idealizations, and to construct hypothetical networks based on these presuppositions. As such, their diagrams represent hypothetical networks, formally equivalent to Boolean statements, and depict neurons that bear little resemblance to "real" neurons. McCulloch and Pitts saw their theory

23. For an interesting analysis of McCulloch's later views on the schism between "formal" and "natural" accounts of neuron activity, and their connection to work in the artificial intelligence movement, see Moreno-Díaz and Mira (1996).

24. Interestingly, Rashevsky noted two years later that their theory was based more directly on experimental observations than some of his previous work on the mathematical biology of the nerve cell (Rashevsky, 1945, p. 153). In subsequent years, McCulloch and Pitts worked with others in Rashevsky's group to reconcile the two approaches. See, for example, Landahl, McCulloch, and Pitts (1943) and Householder and Landahl (1945), especially Chapters XIV and XV. In retrospect, Rashevsky noted that with an increase of empirical knowledge about neuron interaction, the "better mathematical tool" for representing the activity of neurons was not differential equations, but the logical calculus (Rashevsky, 1960, Vol. II, p. 3).

as contributing a "tool for rigorous symbolic treatment of known nets and an easy method of constructing hypothetical nets of known properties" (McCulloch & Pitts, 1943, p. 132).[25]

Although McCulloch and Pitts came from relatively dissimilar intellectual backgrounds, their intellectual concerns found resonance in their application of logic to neural networks. The product of their collaboration integrated many elements. McCulloch's early contact with von Domarus and Northrop fostered his philosophically motivated approach to psychology and physiology. Through his long, fruitful collaboration with Dusser de Barenne, McCulloch pursued his goal of finding a physiological basis of knowledge—an experimental epistemology. Pitts, a mathematical prodigy, had developed mathematical models of neural activity based on differential equations, but also had a mastery of logic, and was drawn to the concept of an "engine of logic," or a "reasoning machine." Through the McCulloch–Pitts collaboration, these concepts became manifested in their model of logical neural networks, which, while giving the logical nature of human reasoning and knowledge a physiological basis, also had a strong conceptual connection to Turing's "logical machine": embodying the notion of a logical description of behavior in a network of neurons.

Rashevsky's project in mathematical biology had provided an important intellectual space for McCulloch and Pitts. "Mathematical biology," as conceived by Rashevsky, with its emphasis on the formalization of complex phenomena, fit in with McCulloch's quest for a "psychon," a "least psychic event," and Pitts's fascination with mathematical logic. With their pursuit of questions that were at once philosophical and physiological, McCulloch and Pitts were able to collaborate within a community of theoretically-oriented mathematical biologists. Although Rashevsky's concept of "mathematical biophysics" often involved the use of differential equations, rather than logic per se, he made strong arguments about the worth of a mathematical approach to biological systems, particularly, the nervous system. Thus, through his creation of the *Bulletin for Mathematical Biophysics*, Rashevsky created a venue for the McCulloch–Pitts collaboration. Indeed, McCulloch later recalled that they were able to publish their paper "thanks to Rashevsky's defense of logical and mathematical ideas in biology" (McCulloch, 1965a, p. 9).[26] Besides being a formative event in the history of cybernetics and cognitive science, the McCulloch–Pitts collaboration had a history of its own, and was an important result of early-twentieth-century efforts to apply mathematics to neurological phenomena.

An earlier version of this article was presented in November 1999 at the History of Science Society Meeting in Pittsburgh, PA. I would like to thank Paul Thompson, Sungook Hong, Kenton Kroker, Jill Lazenby, Larry Smith, Roberto Cordeschi, David McGee, Raymond Faucher, and two anonymous referees for valuable comments and criticisms. I am indebted to McCulloch's daughter, Taffy Holland, who kindly gave permission to reproduce photographs from his collected works. This article also benefited from helpful discussions with colleagues at the Dibner Institute, especially Jutta Schickore.

References

Aizawa, K. (1996). Some neural network theorizing before McCulloch: Nicolas Rashevsky's mathematical biophysics. In R. Moreno-Díaz & J. Mira (Eds.), Brain processes, theories, and models: An international conference in honor of W. S. McCulloch 25 years after his death (pp. 64–70). Cambridge MA: MIT Press.

25. In their 1947 paper, "How We Know Universals: The Perception of Auditory and Visual Forms," McCulloch and Pitts applied this tool to describe perception, and incorporated more empirical knowledge about the histology and physiology of the brain (Pitts & McCulloch, 1947).

26. As Kenneth Aizawa has noted (1996), however, there were important differences between Rashevsky's approach and that of McCulloch and Pitts, particularly that Rashevsky conceived of neural excitation and inhibition as continuous processes, and thus used differential equations in his models. In contrast, McCulloch and Pitts began with the assumption that neural activity was an all-or-none process, and therefore employed logic in their neural network model.

Arbib, M. A. (2000). Warren McCulloch's search for the logic of the nervous system. Perspectives in Biology and Medicine, 43 (2), 193–216.

Aspray, W. (1990). John von Neumann and the origins of modern computing. Cambridge MA: MIT Press.

Bailey, P., Dusser de Barenne, J. G., Garol, H. W., & McCulloch, W. S. (1940). Sensory cortex of chimpanzee. Journal of Neurophysiology, 3, 469–485.

Bartholomay, A. F., Karreman, G., & Landahl, H. D. (1972). Nicolas Rashevsky. Bulletin of Mathematical Biophysics, 34 (no pagination).

Bechtel, W., & Abrahamsen, A. (1991). Connectionism and the mind: An introduction to parallel processing in networks. Oxford: Basil Blackwell.

Bechtel, W., & Richardson, R. C. (1993). Discovering complexity: Decomposition and localization as strategies in scientific research. Princeton NJ: Princeton University Press.

Blair, H. A. (1932). On the intensity–time relations for stimulation by electric currents. Journal of General Physiology, 15, 709–729.

Blustein, B. E. (1992) Percival Bailey and neurology at the University of Chicago, 1928–1939. Bulletin of the History of Medicine, 66, 90–113.

Boole, G. (1854). An investigation of the laws of thought. London.

Boring, E. G. (1929). A history of experimental psychology. New York, London: Century Co.

Carnap, R. (1938). The logical syntax of language. New York: Harcourt-Brace.

Cordeschi, R. (1991). The discovery of the artificial: Some protocybernetic developments. AI & Society, 5, 218–238.

Cowan, J. D. (1998). [Interview with J.A. Anderson and E. Rosenfeld]. In J. A. Anderson & E. Rosenfeld (Eds.), Talking nets: An oral history of neural networks (pp. 97–124). Cambridge MA: MIT Press.

Cowan, J. D., & Sharp, D. H. (1988). Neural nets and artificial intelligence. Daedalus, 177 (1), 85–121.

DiPalma, A. (1992, 23 July). Prof. F. S. C. Northrop dies at 98; Saw conflict resolution in science. The New York Times, p. 89.

Dupuy, J.-P. (2000). The mechanization of the mind: On the origins of cognitive science (M. B. DeBevoise, Trans.) Princeton NJ: Princeton University Press. (Original work published 1994).

Dusser de Barenne, J. G. (1916). Experimental researches on sensory localisations in the cerebral cortex. Quarterly Journal of Experimental Physiology, 9, 355–390.

Dusser de Barenne, J. G. (1924). Experimental researches on sensory localization in the cerebral cortex of the monkey (Macacus). Proceedings of the Royal Society of London, Series B, 96 (676), 272–291.

Dusser de Barenne, J. G. (1934). Some aspects of the problem of "corticalization" of function and of functional localization in the cerebral cortex. In S. T. Orton, J. F. Fulton, & T. K. Davis (Eds.), Localization of function in the cerebral cortex. In Proceedings of the Association for Research in Nervous and Mental Disease (Vol. XIII; pp. 85–106). Baltimore, MD: Williams and Wilkins.

Dusser de Barenne, J. G., & McCulloch, W. S. (1934). An "extinction" phenomenon on stimulation of the cerebral cortex. Proceedings of the Society for Experimental Biology and Medicine, 32, 524–527.

Dusser de Barenne, J. G., & McCulloch, W. S. (1936a). Functional boundaries in the sensori-motor cortex of the monkey. Proceedings of the Society for Experimental Biology and Medicine, 35, 329–331.

Dusser de Barenne, J. G., & McCulloch, W. S. (1936b). Some effects of local strychninization on action potentials of the cerebral cortex of the monkey. Transactions of the American Neurological Association (62nd Annual Meeting, Atlantic City, NJ, June 1–3, 1936), p. 171.

Dusser de Barenne, J. G., & McCulloch, W. S. (1938a). Functional organization in the sensory cortex of the monkey (Macaca mulatta). Journal of Neurophysiology, 1, 69–85.

Dusser de Barenne, J. G., & McCulloch, W. S. (1938b). The direct functional interrelation of sensory cortex and optic thalamus. Journal of Neurophysiology, 1, 176–186.

Dusser de Barenne, J. G., & McCulloch, W. S. (1939). Physiological delimitation of neurones in the central nervous system. American Journal of Physiology, 127, 620–628.

Dusser de Barenne, J. G., McCulloch, W. S., & Ogawa, T. (1938). Functional organization in the face-subdivision of the sensory cortex of the monkey (Macaca mulatta). Journal of Neurophysiology, 1, 436–441.

Edwards, P. N. (1996). The closed world: Computers and the politics of discourse in cold war America. Cambridge–London: MIT Press.

Frank, R. G. (1994). Instruments, nerve action, and the all-or-none principle. Osiris, 9, 208–235.

Fulton, J. F., & Gerard, R. W. (1940). J. G. Dusser de Barenne 1885–1940. Journal of Neurophysiology, 3, 283–292.

Heims, S. J. (1991). The cybernetics group. Cambridge MA: MIT Press.

Hill, A. V. (1936a). Excitation and accommodation in nerve. Proceedings of the Royal Society of London, Series B, 119, 305–355.

Hill, A. V. (1936b). The strength duration relations for electric excitations of medullated nerve. Proceedings of the Royal Society of London, Series B, 119, 440–453.

Householder, A. S. (1941a). A theory of steady-state activity in nerve-fiber networks, I. Bulletin of Mathematical Biophysics, 3, 63–69.

Householder, A. S. (1941b). A theory of steady-state activity in nerve-fiber networks, II. Bulletin of Mathematical Biophysics, 3, 105–112.

Householder, A. S. (1941c). A theory of steady-state activity in nerve-fiber networks, III. Bulletin of Mathematical Biophysics, 3, 137–140.

Householder, A. S. (1942). A theory of steady-state activity in nerve-fiber networks, IV. Bulletin of Mathematical Biophysics, 4, 7–14.

Householder, A. S., & Landahl, H. D. (1945). Mathematical biophysics of the central nervous system. Bloomington IN: Principia Press.

Hughes, J. R. (1993). Development of clinical psychology in the Chicago area. Journal of Clinical Neurophysiology, 10 (3), 298–303.

James, W. (1890). Principles of psychology. Volume I. New York: Holt.

Kneebone, G. T. (1963). Mathematical logic and the foundations of mathematics. London: D. van Nostrand.

Landahl, H. D. (1965). A biographical sketch of Nicolas Rashevsky. Bulletin of Mathematical Biophysics, 27, 3–4.

Landahl, H. D., McCulloch, W. S., & Pitts, W. (1943). A statistical consequence of the logical calculus of nervous nets. Bulletin of Mathematical Biophysics, 5, 135–137.

Leiber, J. (1991). An invitation to cognitive science. Oxford: Basil Blackwell.

Lettvin, J. Y. (1998a). Jerome Lettvin. In L. R. Squire (Ed.), The history of neuroscience in autobiography (Vol. II, pp. 222–243). San Diego CA: Academic Press.

Lettvin, J. Y. (1998b). [Interview with J.A. Anderson and E. Rosenfeld]. In J. A. Anderson & E. Rosenfeld (Eds.), Talking nets: An oral history of neural networks (pp. 1–21). Cambridge MA: MIT Press.

Lotka, A. (1925). Elements of physical biology. Baltimore MD: Williams and Wilkins.

Magnus, R. (1930). The physiological a priori. Lane lectures on experimental pharmacology and medicine, Stanford University Publications, University Series, Medical Sciences. Volume II, No. 3. Stanford CA: Stanford University Press, pp. 97–103.

Marshall, L. H. (1987). Instruments, techniques, and social units in American neurophysiology, 1870–1950. In G. L. Geison (Ed.), Physiology in the American context, 1850–1940 (pp. 351–369). Bethesda MD: American Physiological Society.

Marshall, L. H., & Magoun, H. W. (Eds.). (1998). Discoveries in the human brain: Neuroscience prehistory, brain structure, and function. Totowa NJ: Humana Press.

Martin, O. (1997). La mesure en psychologie de Binet à Thurstone, 1900–1930. Revue de Synthèse, 118 (4), 457–493.

McCorduck, P. (1979). Machines who think: A personal inquiry into the history and prospects of artificial intelligence. San Francisco: W. H. Freeman.

McCulloch, R. (Ed.). (1989). The collected works of Warren S. McCulloch (Vols. I–IV). Salinas CA: Intersystems.

McCulloch, W. S. (1965a). What is a number, that a man may know it, and a man, that he may know a number? In W.S. McCulloch, Embodiments of mind (pp. 1–18). Cambridge MA: MIT Press. (Originally published as The Ninth Annual Alfred Korzybski Memorial Lecture, General Semantics Bulletin Nos. 26 and 27. Lakeville CT: Institute of General Semantics, 1961, pp. 7–18)

McCulloch, W. S. (1965b). What's in the brain that ink may character? In W. S. McCulloch (1965). Embodiments of mind (pp. 387–397). Cambridge MA: MIT Press. (Originally presented at the International Congress for Logic, Methodology, and Philosophy of Science, Jerusalem, Israel, 28 August 1964)

McCulloch, W. S. (1967a). Lekton, being a belated introduction to the thesis of Eilhard von Domarus. In L. Thayer (Ed.). Communication: Theory and research. Proceedings of the first symposium (pp. 348–350). Springfield IL: Charles C. Thomas.

McCulloch, W. S. (1967b). Commentary. In L. Thayer (Ed.), Communication: Theory and research. Proceedings of the first symposium (pp. 412–428). Springfield IL: Charles C. Thomas.

McCulloch, W. S. (1989). Recollections of the many sources of cybernetics. In W. S. McCulloch, The collected works of Warren S. McCulloch (Vol. I, pp. 21–49). Salinas CA: Intersystems. (Reprinted from ASC Forum, VI (2), Summer 1974, pp. 5–16)

McCulloch, W. S., & Dusser de Barenne, J. G. (1935) Extinction: Local stimulatory inactivation within the motor cortex. Americal Journal of Physiology, 113, 97–98.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5,115–133.

Moreno-Díaz, R., & Mira, J. (1996). Logic and neural nets: Variations on themes by W. S. McCulloch. In R. Moreno-Díaz & J. Mira (Eds.), Brain processes, theories, and models: An international conference in honor of W.S. McCulloch 25 years after his death (pp. 24–36). Cambridge MA: MIT Press.

Northrop, F. S. C. (1931). Science and first principles. New York: Macmillan.

Northrop, F. S. C. (1938). The history of modern physics in its bearing on biology and medicine. Yale Journal of Biology and Medicine, 10, 209–232.

Northrop, F. S. C. (1940). The method and theories of physical science and their bearing upon biological organization. Growth (Supplement, Second Symposium on Development and Growth), 4, 127–154.

Pitts, W. (1942a). Some observations on the simple neuron circuit. Bulletin of Mathematical Biophysics, 4, 121–129.

Pitts, W. (1942b). The linear theory of neuron networks: The static problem. Bulletin of Mathematical Biophysics, 4, 169–175.

Pitts, W. (1943). The linear theory of neuron networks: The dynamic problem. Bulletin of Mathematical Biophysics, 5, 23–31.

Pitts, W., & McCulloch, W. S. (1947). How we know universals: The perception of auditory and visual forms. Bulletin of Mathematical Biophysics, 9, 127–147.

Pratt, V. (1987). Thinking machines: The evolution of artificial intelligence. Oxford: Basil Blackwell.

Rall, W. (1990). Some historical notes. In E. L. Schwartz (Ed.), Computational neuroscience (pp. 3–8). Cambridge MA: MIT Press.

Ramón y Cajal, S. (1995). Histology of the nervous system of man and vertebrates (Vols. I and II) (N. Swanson & L. W. Swanson, Trans). New York: Oxford University Press. (Original work published 1911)

Rashevsky, N. (1928a). On the size distribution of colloidal papers. Physical Review, 31, 115–118.

Rashevsky, N. (1928b). Zur Theorie der spontanen Teilung von mikroskopischen Tropfen. Zeischrift für Physik, 46, 568–593.

Rashevsky, N. (1929). The problem of form in physics and biology. Nature, 124, 10.

Rashevsky, N. (1931a). Learning as a property of physical systems. Journal of General Psychology, 5, 207–229.

Rashevsky, N. (1931b). Possible brain mechanisms and their physical models. Journal of General Psychology, 5, 368–406.

Rashevsky, N. (1931c). On the theory of nerve conduction. Journal of General Physiology, 14, 517–528.

Rashevsky, N. (1933). Outline of a physico-mathematical theory of excitation and inhibition. Protoplasma, 20, 42–56.

Rashevsky, N. (1936). Mathematical biophysics and psychology. Psychometrika, 1, 1–26.

Rashevsky, N. (1938). Mathematical biophysics: Physicomathematical foundations of biology. Chicago: University of Chicago Press.

Rashevsky, N. (1945). A reinterpretation of the mathematical biophysics of the central nervous system in the light of neurophysiological findings. Bulletin of Mathematical Biophysics, 7, 151–160.

Rashevsky, N. (1946). Development of mathematical biophysics in the U.S.A. from 1939 to 1945 inclusive. Relationes de Avctis Scientiis Tempore Belli, 14, 3–28.

Rashevsky, N. (1948). Mathematical biophysics (Rev. ed.). Chicago: University of Chicago Press.

Rashevsky, N. (1960). Mathematical biophysics (Vols. I–II; 3rd rev. ed.). New York: Dover.

Rosen, R. (1972). Nicolas Rashevsky 1899–1972. In R. Rosen & F. M. Snell (Eds.), Progress in theoretical biology (Vol. II, pp. xi–xiv). New York–London: Academic Press.

Russell, B. (1920). Introduction to mathematical philosophy. London: George Allen and Unwin.

Schrecker, P. (1947). Leibniz and the art of inventing algorisms. Journal of the History of Ideas, 8, 107–116.

Shannon, C. E. (1938). A symbolic analysis of relay and switching circuits. Transactions of the American Institute of Electrical Engineers, 57, 713–723.

Smalheiser, N. R. (2000). Walter Pitts. Perspectives in Biology and Medicine, 43 (2), 217–226.

Smith, L. D. (1986). Behaviorism and logical positivism: A reassessment of the alliance. Stanford CA: Stanford University Press.

Thompson, D. (1917). On growth and form. Cambridge UK: Cambridge University Press.

Turing, A. M. (1936–7). On computable numbers, with an application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, Series *2*, 42, 230–265.

Volterra, V. (1931). Leçons sur la théorie mathématique de la lutte pour la vie. Paris: Gauthier-Villars et cie.

Von Domarus, E. (1944). The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), Language and thought in schizophrenia: Collected papers presented at the meeting of the American Psychiatric Association, May 12, 1939 (pp. 104–114). Berkeley: University of California Press.

Von Domarus, E. (1967). The logical structure of mind: An inquiry into the philosophical foundation of psychology and psychiatry. In L. Thayer (Ed.), Communication: theory and research. Proceedings of the first symposium (pp. 351–411). Springfield IL: Charles C. Thomas. (Original work completed in 1930).

Von Eckardt, B. (1993). What is cognitive science? Cambridge MA: MIT Press/Bradford Books.

Von Neumann, J. (1981). First draft report on the EDVAC. Report prepared for the U.S. Army Ordnance Department under contract W-670-ORD-4926, 1945. In N. Stern (Ed.), From ENIAC to UNIVAC (pp. 177–246). Bedford, MA: Digital Press. (Original work published 1945)

Von Neumann, J. (1951). The general and logical theory of automata. In L. A. Jeffress (Ed.), Cerebral mechanisms in behavior: The Hixon symposium (pp. 1–31). New York: John Wiley.

Werkmeister, W. H. (1936). The Second International Congress for the Unity of Science. The Philosophical Review, 45 (6), 593–600.

Woodger, J. H. (1937). The axiomatic method in biology. Cambridge: Cambridge University Press.

Young, R. M. (1970). Mind, brain, and adaptation in the nineteenth century. Oxford: Clarendon Press.