representation and rule, or data and program, was not a real distinction until von Neumann designed a physical machine to realize Turing’s machines.

Notably, since Turing’s model did not specify any goals beyond the desired final states of specific sequences of state transitions, Von Neumann was free to design the physical machine to meet engineering goals. A Central Control unit containing stored programs was more efficient for sequencing operations even though it minimized the flexibility (self-modifiability) of the machine’s internal operations (1966: secs. 2.2, 2.3); flexible intelligence could be left to the human programmer. Even the fact that digital computers are “fragile” in that they break down (stop operating) at an error was part of their explicit design (von Neumann 1966: 73). This latter feature became a fatal bug in the eyes of connectionists and their champions (e.g., Churchland) who sought a more neurally realistic design given the common goal of materialism.

 The incorrect inference from observing the behavior of machines built to satisfy these engineering goals is that computing can’t explain human intelligence (Dreyfus 1992). Of course von Neumann also had human intelligence in mind, in that he sought a general theory of information processing in humans and machines (Aspray 1985). But he was also well aware of the differences. Computers were much faster than brains and stopped to fix errors as they arose; brains had far more memory and isolated problems for working on them on the side. Von Neumann estimated that while the human nervous system had a million times as many components as the ENIAC, each of the machine’s components was about 5,000 times faster than a neuron; overall, “the nervous system outperforms the machine by a factor of roughly 200.” [[1]](#footnote-2)

If anything, the assimilation of minds to stored-program computers by psychologists and others aiming directly for materialism (e.g., Newell, Shaw, and Simon 1958, Newell and Simon 1976) may be understood as an expression of hope about just how much mind could be explained with these first incarnations of Turing’s model.[[2]](#footnote-3) For despite this over-optimism about what such computers could do, they did show that complex outputs need not require mysteriously complex internal operations or state transitions. For experimental psychologists, who faced a growing pile of anomalies that motivated looking inside the behaviorist’s black box (Miller 1956; Chomsky 1959; Bisiach and Luzzatti 1978; Rescorla 1988), this sufficed to provide a viable non-intuitionist alternative framework.[[3]](#footnote-4) Turing’s machine was “complicated enough to do everything that cognitive theorists have been talking about” (Miller, Galanter, and Pribram 1960: 43). The model was thus elaborated directly in psychological terms. Symbols were interpreted as natural-language-like concepts, and rules were the rules of deductive logic or heuristics (Fodor 1975, Newell and Simon 1976; Miller, Galanter, and Pribram 1960: 3). The stored-program computer was the machine for which these first psychologically interpretable internal state transitions – high-level software programs – were developed.

These agent-level reinterpretations of Turing’s model also helped clarify the role of experimental psychology in the new science of cognition: to individuate and empirically articulate capacities whose effects – cognitive deficits, reaction time differences, and other tests – can be experimentally isolated or distinguished at the organism level, including behaviors that can only be observed in experimental situations and are inaccessible to intuition or self-reflection. If psychologists took away from Turing the lesson that “if they could describe exactly and unambiguously anything that a living organism did, then a computing machine could be built that could exhibit the same behavior with sufficient exactitude to confuse the observer” (Baddeley 1994: 46), then their role was to satisfy the antecedent, not the consequent. The autonomy of psychology from lower-level theories should be understood from this perspective.[[4]](#footnote-5)

From the same perspective, “strong AI” and “weak AI” (Searle 1980) are different attitudes towards the same point on a continuum of systems of different degrees of freedom of operation. Turing’s mathematical model is theoretically apt for any type of agent – any point on the continuum. Searle’s claim that there is no understanding in a system that realizes an unelaborated Turing machine – a claim that injected a large dose of reality into the AI community – is a way of saying humans do not occupy that point on the continuum. But Turing’s model is not the whole story.

**II. 2. 1943: McCulloch and Pitts: Brainware (neural logic-gates)**

*We do resent the hiatus between our mental terminology and our physical terminology. It is being attacked in a very realistic fashion today.*

 McCulloch 1943 (from the Warren S. McCulloch Papers, cited in Piccinini 2004)

 Explaining agency materialistically, given the known biological agents, requires a theory of how the brain (or central nervous system) might be the sort of machine that could produce mental phenomena. McCulloch and Pitts provided this theory. They interpreted neural activity as logic gates and theorized that networks of neuron-like units could be equivalent in computing power to a simple Turing machine. In short, brains were Turing machines.[[5]](#footnote-6)

A McCulloch-Pitts neuron is an abstract biological analogue of an electrical relay, a basic component of a von Neumann computer (von Neumann 1945: 4.2, 4.3; Wiener 1948: Ch. 5; Arbib 2000: 212). [[6]](#footnote-7) These neuron units were binary in operation, so could be associated with propositions;

1. Rosenblueth et al. provide a simple analogy: a television receiver can be thought of as a single-cone retina that scans swiftly (obtains orderly successive detection of a signal at a rate of 20 million per second), whereas a human eye, with about 6.5 million cones and 115 million rods, uses a spatial rather than temporal multiplier of the signal. [↑](#footnote-ref-2)
2. Newell and Simon’s physical symbol system hypothesis – that a physical symbol system has the necessary and sufficient means for intelligent action – covered humans and computers alike. They allowed that only systems of sufficient complexity and power could exhibit general intelligence: in effect, “intelligent action” did not mean “action that ensues from human-like general intelligence.” Turing (1950) also linked this internal processing story to human linguistic behavior by proposing the Turing Test, in which an interrogator tries to determine if her hidden interlocutor is a human or a computer. He was over-optimistic about when a computer would pass the test (50 years; no computer has yet); similarly, Newell, Shaw, and Simon predicted a chess-playing computer would beat a human chess champion by 1957 (Big Blue beat Gary Kasparov in 1997). [↑](#footnote-ref-3)
3. Radical behaviorism never took hold in developmental, comparative, social, perceptual, or clinical psychology, and was not dominant outside the U.S.. (Greenwood 1999; Miller 2003); even Skinner was conflicted (Baars 2003). However, within its pockets of influence, its grip was profound: even Neisser’s 1967 *Cognitive Psychology* text, which gave a label to post-behavioristic experimental psychology, had six chapters on vision, four on audition, and just one slim final chapter on higher cognitive functions. But radical behaviorism left two important legacies. First, the demand for observable behavioral evidence of psychological claims (“methodological” behaviorism) is now entrenched. Second, by focusing on behavior rather than consciousness, behaviorism “helped to break down the distinction between the mental behavior of humans and the information processing of lower animals and machines” (Aspray (1985: 128). Bringing humans down a peg has been accompanied by raising other animals a notch, as evidenced by a corresponding transition, albeit with continuing controversy, from behavioristic ethology to cognitive ethology (e.g., de Waal, Emery and Clayton; Tomasello and Call). [↑](#footnote-ref-4)
4. As Newell, Shaw, and Simon (op.cit.: 163) put it: “Discovering what neural mechanisms realize these information-processing functions in the human brain is a task for another level of theory construction.” Off-loading problems that are not of direct interest, particularly if the technology for investigating them is not yet available, is not an assertion of autonomy in some strong sense; psychology may well proceed without that information up to a point, and then further progress requires neural information too. [↑](#footnote-ref-5)
5. As Piccinini (2004: 205) puts it, “After 1943, computing could be thought of as all that humans did.” [↑](#footnote-ref-6)
6. Thus, McCulloch and Pitts provide a common origin for both classical and connectionist versions of computationalism (Boden 1991). In addition to the electrical relay, the Central Control and Memory of von Neumann’s machine were intended to “correspond to the associative neurons in the human nervous system” (von Neumann: 3, sec. 2.6; see also sec. 4.0, 4.2) – that is, the hidden layers of a connectionist network. [↑](#footnote-ref-7)