Jordan Bailey

Classic Paper 1: Review/Analysis

Title and Author

Title: *Computing Machinery and Intelligence*

Author: *A. M. Turing*

Summary/Hook

Within Alan Turing's paper, "Computing Machinery and Intelligence," he proposes the question, "Can machines think?", and challenges it by creating a test called the Imitation Game – eventually known as the Turing Test. Within the paper, he describes the Turing Test in detail, defines the machines taking part in the test and how they work, and further elaborates on Discrete State Machines and Universal Turing Machines. The rest of the paper is dedicated to analyzing nine objections/arguments to his question, "Can machines think?": The Theological Objection, "Heads in the Sand" Objection, Mathematical Objection, the Argument from Consciousness, Arguments from Various Disabilities, Lady Lovelace's Objection, the Argument from Continuity in the Nervous System, the Argument from Informality of Behavior, and the Argument from Extrasensory Perception.

Knowledge Related to the Cognitive Science Program Learning Outcomes

1. **Foundational Assumptions**
   It was suggested tentatively that the question, "Can machines think?" should be replaced by "Are there imaginable digital computers which would do well in the imitation game?" If we wish we can make this superficially more general and ask "Are there discrete-state machines which would do well?" But in view of the universality property we see that either of these questions is equivalent to this, "Let us fix our attention on one particular digital computer C. Is it true that by modifying this computer to have an adequate storage, suitably

increasing its speed of action, and providing it with an appropriate programme, C can be made to play satisfactorily the part of A in the imitation game, the part of B being taken by a man?"

2. **Formal Systems and Theories of Computation**

There are a number of results of mathematical logic which can be used to show that there are limitations to the powers of discrete-state machines. The best known of these results is known as Godel's theorem (1931) and shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent. There are other, in some respects similar, results due to Church (1936), Kleene (1935), Rosser, and Turing (1937). The latter result is the most convenient to consider, since it refers directly to machines, whereas the others can only be used in a comparatively indirect argument: for instance if Godel's theorem is to be used we need in addition to have some means of describing logical systems in terms of machines, and machines in terms of logical systems.

3. **Consciousness and Controversies**

This argument is very, well expressed in Professor Jefferson's Lister Oration for 1949, from which I quote. "Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain-that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants."

4. **Psychological Investigations (psychological processes pertaining to learning)**

The use of punishments and rewards can at best be a part of the teaching process. Roughly speaking, if the teacher has no other means of communicating to the pupil, the amount of information which can reach him does not exceed the total number of rewards and punishments applied. By the time a child has learnt to repeat "Casabianca" he would probably feel very sore indeed, if the text could only be discovered by a "Twenty Questions" technique, every "NO" taking the form of a blow. It is necessary therefore to have some other "unemotional" channels of communication. **If these are available it is possible to teach a machine by punishments and rewards to obey orders given in some language, e.g., a symbolic language.** These orders are to be transmitted through the "unemotional" channels. The use of this language will diminish greatly the number of punishments and rewards required.

5. **Darwinian Processes and Phenomena**

We have thus divided our problem into two parts. The child programme and the education process. These two remain very closely connected. We cannot expect to find a good child machine at the first attempt. One must experiment with teaching one such machine and see how well it learns. One can then try another and see if it is better or worse. There is an obvious connection between this process and evolution, by the identifications: *Structure of the child machine = hereditary material*, *Changes of the child machine = mutation*, *Natural selection = judgment of the experimenter*. One may hope, however, that this process will be more expeditious than evolution. The survival of the fittest is a slow method for measuring advantages. The experimenter, by the exercise of intelligence, should he able to speed it up. Equally important is the fact that he is not restricted to random mutations. If he can trace a cause for some weakness he can probably think of the kind of mutation which will improve it.