Bringing up Turing's 'Child-Machine'

S. G. Sterrett[+]

Department of Philosophy, Carnegie-Mellon University, Pittsburgh PA USA 15213
susangsterrett@gmail.com

Abstract. Turing wrote that the "guiding principle" of his investigation into the possibility of intelligent machinery was "The analogy [of machinery that might be made to show intelligent behavior] with the human brain." (Turing 1948) In his discussion of the investigations that Turing said were guided by this analogy, however, he employs a more far-reaching analogy: he eventually expands the analogy from the human brain out to "the human community as a whole." Along the way, he takes note of an obvious fact in the bigger scheme of things regarding human intelligence: grownups were once children; this leads him to imagine what a machine analogue of childhood might be. In this paper, I'll discuss Turing's child-machine, what he said about different ways of educating it, and what impact the "bringing up" of a child-machine has on its ability to behave in ways that might be taken for intelligent. I'll also discuss how some of the various games he suggested humans might play with machines are related to this approach.

Keywords: Machine Intelligence, Artificial Intelligence, Analogy, Reinforcement learning, intelligence, Turing, education, child-machine, Turing test, game show, imitation game, computer, computing machinery, universal machine

1 A 'Guiding Principle'

In his writings on intelligence and machinery, Turing often employs analogies. One analogy he states explicitly and calls the "guiding principle" of his investigation into "possible ways in which machinery might be made to show intelligent behavior" is "the analogy with the human brain." (Turing 1948)

The analogy that Turing employs in the discussions that follow is not a simple analogy between machine and brain; it's more specific, and less physically-oriented, than the brief description of an analogy between computing machinery and the human brain quoted above might at first suggest. Turing says his investigation is mainly concerned with the analogy between *the ways in which a human* [with a human brain] is *educated* such that *the potentialities for human intelligence are realized*, and "an analogous teaching process applied to machines." (Turing 1948).

That is, his investigation concerns identifying and evaluating proposals for filling in the part of the analogy that answers: if we want a machine to fulfil *its* potentialities for intelligence, *how should it be "educated"*? In his 1950 "Computing Machinery and

[+] I should like to thank an anonymous referee for some helpful remarks, and the audience at the Fourth Regional Wittgenstein Workshop held at Washington and Lee University on March 11th, 2012 for discussion on this paper.

Intelligence", he spoke of dividing the problem of building a machine that can imitate the human mind into two parts: "The child program and the education process." He mentions yet a third component, on this approach: "Other experience, not to be described as education, to which it [the machine] has been subjected." That is, there is a distinction between "the education process" and "other experience." But what is it that distinguishes the education process?

2 Intelligent Behavior versus Completely Disciplined Behavior

This analogy -- between a machine that has undergone an education process and a human student who has been educated by a teacher -- provides Turing with the means to respond to one of the most common objections raised against the possibility that a machine could be regarded as exhibiting intelligent behavior. This objection (to the possibility of intelligent machinery) is, in Turing's words, the view that "[i]nsofar as a machine can show intelligence this is to be regarded as nothing but a reflection of the intelligence of its creator." That view, he says, is much like the view that "the credit for the discoveries of a pupil should be given to his teacher", which can be rebutted as follows:

> " In such a case the teacher would be pleased with the success of his methods of education, but would not claim the results themselves unless he had actually communicated them to his pupil. He would certainly have envisaged in very broad outline the sort of thing his pupil might be expected to do, but would not expect to forsee any sort of detail. " (Turing 1948, p. 2)

Turing contrasts "intelligent behavior" of a machine with "completely disciplined behavior." Both are exceptional sorts of behavior for a machine; he says that "Most of the programmes which we can put into the machine will result in it doing something that we cannot make sense of at all, or which we regard as completely random behavior."

In both intelligent behavior and completely disciplined behavior, we are able to make sense of the machine's behavior. But the kind of sense we make of it differs. When a machine is carrying out computations, the machine's behavior is "completely disciplined" and what we strive for is to have "a clear mental picture of the state of the machine at each moment in the computation." (Turing 1950, p. 459) When a teacher is educating a machine with an intent to produce an intelligent machine, the goal of the education process entails that some of the machine's rules of behavior will be undergoing change. The teacher will be able "to some extent to predict the pupil's behavior", but, in contrast to the case of programming it to carry out computations, won't have a clear picture of what is going on within the machine being educated. Intelligent behavior is not a large departure from completely disciplined behavior, but it does differ qualitatively from completely disciplined behavior: the sense we make of it is distinctively different. Intelligent behavior escapes the predictability of completely disciplined machine behavior without veering off into random behavior.

One qualitative difference between completely disciplined behavior and intelligent behavior is the presence of initiative. When describing a universal machine with no special programming but able to carry out whatever program is put into it, Turing remarks that after carrying out the actions specified by the program, it would sink into inactivity until another action is required. It would lack initiative. This is one reason that the universal computer, even if produced from a child-machine by some machine analogue of an education process to produce completely disciplined behavior, is not a good candidate for a machine analogue of a human --- even though the actions that it does take would be faultless. Turing thinks that an intelligent machine will be fallible, and that the feature of fallibility can be (or might be an indication of something that is) an important advantage. He also thinks that it might be required to have some sort of random element in a machine in order to produce a machine that is amenable to undergoing the kind of process that is analogous to the education of a human.

So, randomness probably has a part to play in producing a machine that might possibly be said to exhibit intelligence. Yet, intelligent machine behavior is not random behavior. The part that randomness plays in intelligent machinery is in the generation of possibilities among which some search process is then employed. (Turing 1950, p. 459) Turing writes of using a random element to generate forms of behavior at one point (Turing 1950, p. 459); at another point he speaks of using a random element to generate different child-machines among which one then selects the best ones. (Turing 1950, p. 456). However, in both places he indicates that random generation alone does not seem very efficient, and that he would expect to supplement the generation of alternatives or the search among alternatives with some more directed, more informed process. Of the process of "finding" an appropriate child-machine, he writes:

> 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 be 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 of the weakness he can probably think of the kind of mutation which will improve it." (Turing 1950, p. 456)

And, of the education process, which aims to find the appropriate behavior:

> "The systematic method [of trying out different possibilities in the search for a solution] has the disadvantage that there may be an enormous block without any solutions in the region which has to be investigated first. Now the learning process may be regarded as a search for a form of behaviour which will satisfy the teacher (or some other criterion). Since there is probably a very large number of satisfactory solutions the random method seems to be better than the systematic." (Turing 1950, p. 459)

The education process is a matter of "intervening" on the machine. Just as the behavior of the early machines could be changed by using a screwdriver to change the machine's physical configuration by physical means, so the behavior of digital computers can be

changed by using communication with it to change its rules of operation in some way. These two kinds of intervention are referred to as "screwdriver intervention" and "paper intervention", respectively. In "Intelligent Machinery", the guiding principle (the analogy mentioned earlier) is employed here, too -- with some qualifications. Turing notes that human life is such that "interference is the rule rather than the exception." He identifies which part of human life he means to compare to a machine that might be regarded as exhibiting intelligence:

> " [A human] is in frequent communication with other [humans], and is continually receiving visual and other stimuli which themselves constitute a form of interference. *It will only be when the [human] is'concentrating' with a view to eliminating these stimuli or 'distractions' that he approximates a machine without interference*." (Turing 1948, p. 8; emphasis added)

The human behavior during a time period when the human approximates a machine without interference, though, "is largely determined by the way he has been conditioned by previous interference."

Since, as he says, humans are constantly undergoing interference, how is the analogy between humans and machines supposed to go here? What is the difference between undergoing an education process and being intervened upon in other ways? Well, he seems to think of education as a special kind of interference: it involves a teacher who intentionally tries to affect the behavior of the machine. It's interference directed towards some goal. So, even though humans undergo interference as a rule as they go about their daily lives (except for the times when they withdraw and concentrate on something), we still want to distinguish the kind of interference that is education from other kinds of interference.

The analogy may not be precise, but I think it is pretty clear: humans undergo education processes for a portion of their lives (which Turing estimates at about the first twenty years of their lives), and their behavior after that is very much affected by the education they have received, even though they still receive other interference -- most of the time, in fact. The point is to approximate the human process of education with some analogous process suitable for machines. The major points of his proposal are that, on analogy with a human's life, we plan for these three stages of a machine: first, there is the infant stage of a machine, which is a machine that has not been educated and is at least partly unorganized. It need not be a blank slate, but it is important that large amounts of its behavior are undetermined. This is followed by the child-machine stage, during which the machine is educated. The first stage of education is to get the machine to a point where "it could be relied on to produce definite reactions to certain commands." (Turing 1948, p. 118) Education involves a teacher who is intentionally trying to teach or modify the machine's behavior to effect some specific kinds of behavior. The machine's behavior is in flux during this time. Even if the machine is given the means to educate itself using some kind of program during the child-machine stage, there is still oversight and monitoring by a teacher of sorts who checks up on its progress and intervenes if necessary. At some point the education can be ended, and the machine that results when

education is ended is supposed to behave in a way that can be predicted "in very broad outline" by someone familiar with how it has been educated --- but its behavior might not, in fact probably will not, be fully predictable. Finally, there is the adult-machine, which is still capable of learning, but is also capable of quite complex behavior without additional intervention.

What about a process that would start with an unorganized machine, which we would then 'organize' by suitable interference to be a universal machine (e.g., a digital computer capable of being programmed)? Turing says that researchers should be interested in understanding the process that begins with an unorganized machine and results in a universal machine, but he doesn't regard such a process as the appropriate "analogous process" of human education: a universal machine isn't really the behavioral analogue of an adult. One of the differences between a human adult and a universal machine is the point mentioned above regarding the lack of initiative. There are other reasons, too: such an adult-machine would "obey orders given in an appropriate language, even if they were very complicated; he would have no common sense, and would obey the most ridiculous orders unflinchingly." (Turing 1948, p. 116)

Turing describes an experiment in "educating" machines he carried out. It involved a process meant to be analogous to administering punishments and rewards; of giving the machine something analogous to pain and pleasure. The machine to be educated in his experiment was one whose description was incomplete, as he put it, meaning that its actions were not yet fully specified; thus, the machine's operation would give rise to specific cases where the action called for is not determined. When such a specific case arises, the following is done: an action is selected randomly and applied tentatively, by making the appropriate entry in the machine's description. This is the point where the teacher "educates" the machine.

The general idea of employing pleasure and pain he has in mind is revealed in his discussion of "pleasure-pain systems." We can get the general idea without getting into the details too much. He is considering unorganized machines whose states are described using two expressions, one of which he calls "character": "Pleasure interference tends to fix the character, i.e., towards preventing it changing, whereas pain stimuli tend to disrupt the character, causing features which had become fixed to change, or to become again subject to random variation." When he describes the particular experiment he carried out, though, which he refers to as a "particular type of pain-pleasure system", the analogy seems to employ the brain-machine analogy quite directly: "When a pain stimulus occurs all tentative entries are cancelled, and when a pleasure stimulus occurs they are all made permanent." (Turing 1948, p. 118) At the time, he found it took too much work to pursue this means of educating a machine much farther than the rather simplified version of it he had carried out.

As his friend and colleague Donald Michie put it, they were waiting for hardware. Michie recounts a story about one of Turing's plans to program the "Manchester Baby" to investigate what would happen when two different programs for playing chess were pit against each other: "[Turing] was thwarted (rightly) by . . . the guardian of its scarce

resources, Tom Kilburn." According to Michie, in the years leading up to Turing's 1948 and 1950 papers on intelligent machinery, Michie, Turing, and Jack Good "formed a sort of discussion club focused around Turing's astonishing 'child machine' concept. [1] His proposal was to use our knowledge of how the brain acquires its intelligence as a model for designing a *teachable intelligent machine*." (Michie 2002) The idea that the source of learning might be sought in some random elements of neural physiology was well-known in psychology; decades earlier, in his *Principles of Psychology*, William James had concluded a discussion on the formation of pathways in the brain: " All this is vague to the last degree, and amounts to little more than saying that a new path may be formed by the sort of *chances* that in nervous material are likely to occur." (James, 1890, p. 104) The discussion club worked on developing the analogy for machines; one might say that what they were doing in that discussion club was developing the basic ideas of what has since become known as reinforcement learning. Michie later showed that reinforcement learning could indeed be successfully carried out in machines, proving sceptics wrong. (Michie 1961) In 1948, though, there was still a "wait for hardware", and having to wait for the hardware to be available to test their ideas must have been frustrating.

Turing also outlined other approaches he would have liked to try: one might program one's "teaching policies" into the machine, and let it run for awhile, modifying its own programs, and periodically check to see how much progress it has made in its education. He regarded the problem of building a machine that would display "initiative" as well as "discipline" (as he put it) as crucial. Achieving discipline in a machine: that we can see how to do. What initiative adds to discipline in a machine: this is a matter of comparing humans and machines with complete discipline, and asking what humans that are able to communicate had in addition to discipline. Then, one could address what it was that should be copied in the machine. A question remained as to what process to use that achieves ending up with a machine that had both. In particular, in what order should these two be instilled in the machine: first, discipline, then initiative, or somehow both together?

3 Teachers, singular and plural

In most of Turing's examples of a teacher educating a machine, it seems he is thinking of one or at most a few individual teachers. The kind of machine under consideration is a universal (i.e., programmable) machine, specifically, a digital computer, equipped with a means of "at most, organs of sight, speech, and hearing." His investigations are, as a result, biased towards activities that "require little contact with the outside world." (Turing 1948, p. 117) There are other obstacles, too: even if a machine were equipped with the ability to navigate physically, there are limitations on its abilities to be

[1] Turing actively sought out discussion with colleagues. Another such colleague was the philosopher Ludwig Wittgenstein, whose seminar he attended. For Turing's 'contructive uses' of their discussions, see Floyd (to appear).

socialized. More than once, he mentions the advantage that human learners have in that they are able to benefit from interactions with other humans.[2]

It is interesting that a major piece of research in cultural anthropology on cross-cultural features of child-rearing appeals to neural processes very much in line with Turing's "pain-pleasure systems", except that it is evaluations of goodness and badness, rather than physical pleasure and physical pain, that are administered. What is interesting is that this work (Quinn 2003) provides a model of what the education process of a human child would be on the "it takes a village to raise a child" view: "Cultural models of child rearing, thus, exploit the neural capacities of the children so reared, to achieve a result, adulthood, that could not be accomplished by the human brain alone." One can see the kind of issues this might raise: what if different members of the community contradict each other? This is exactly the issue ("constancy of [the child's] experience") that the cultural anthropologists in (Quinn 2003) discovered was universally deemed important.

Turing does not talk about this kind of education -- education by a community, or education constrained and informed by cultural norms -- of a child-machine, but he does talk about the role of the human community in the intellectual activity of humans. It occurs in his discussion of initiative. In that discussion, by the time he got to talking about human community, he had already treated the issues of discipline and initiative separately, and the education of the child-machine had been limited to the instillation of discipline into the machine. Yet, given that he did, albeit very briefly, indicate that the analogy between brain and machine might involve how the human community is involved in the education process of a human, the question of an analogue for a community as teacher during the education process of a machine arises quite naturally.

We might ask, what kind of community? Turing considered digital computers that have "organs" for sight, speech, and hearing. The newest cellphones (e.g., equipped with Siri) have such "organs", and they incorporate some machine learning capabilities, including learning the preferences and habits of their owners. These kinds of machines (state of the art cellphones) generally interact with and "learn from" a single human owner. Yet, they interact with other virtual agents who communicate with them. Now that we have these possibilities not available to Turing, we might consider following through in more detail on remarks that Turing made about "intellectual search" by humans immersed in a human community, and consider what the analogous processes for a machine might be.

4 Imitation and Intelligence

Turing concludes his 1948 paper "Intelligent Machinery" with a section on the concept of intelligence, which ends in the description of an experiment. The experiment involves three people and the game of chess.

[2] Harry Collins' research on imitation games extends this observation about the value of interactions with others. He also ties the kind of knowledge gained in this way with the kind of knowledge that can be exhibited using imitation games. (http://www.cardiff.ac.uk/socsi/contactsandpeople/harrycollins/expertise-project/imitationgameresearch.html)

> "It is not difficult to devise a paper machine which will play a not very bad game of chess. Now get three [humans] as subjects for the experiment A, B, C. A and C are to be rather poor chess players, B is the operator who works the paper machine. . . . Two rooms are used with some arrangement for communicating moves, and a game is played between C and either A or the paper machine. C may find it difficult to tell which he is playing." (Turing 1948, p. 127)

By 'paper machine', Turing means creating "the effect of a computing machine by writing down a set of rules of procedure and asking a man to carry them out. " (Turing 1948, p. 113) So, the human who operates the paper machine, in conjunction with the written rules of procedure, is imitating a machine. Now, although it is meant to be straightforward to imitate a machine by this method, Turing does not consider it trivially easy, for he advises using someone who is both a mathematician and a chess player to work the paper machine.

Now, this experimental setup is most assuredly not intended to be an *objective* measure of intelligence of the paper machine. To prevent any charge of interpretive license, let me quote Turing from this last section of the paper, which bears the heading "Intelligence as an emotional concept": "The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration."

Upon what, then, does regarding something as intelligent depend? His answer is given in terms of what it is that would rule out regarding something as intelligent: "If we are able to explain and predict its behaviour or if there seems to be little underlying plan, we have little temptation to imagine intelligence." Different people bring different skills with respect to explaining and predicting the behavior of something: "With the same object therefore it is possible that one man would consider it as intelligent and another would not; the second man would have found out the rules of its behavior."

The experiment is set up as a comparison: between A, a "rather poor" chess player, and B, a paper machine. The experiment is not in terms of whether B can beat A at chess -- the way the experiment is set up, B, which is a man imitating the behavior of a machine by implementing rules that could be carried out by a machine, but which are written by a human and intended to be read by a human, will likely win some rounds. The comparison is not between chess-playing abilities, but between how transparent it is to C that B's behavior is being produced by following a set of written rules capable of being carried out by a machine.

While "Intelligent Machinery" closed with the description of a three person game about telling the difference between a performance generated by 'rules of behavior' and one by a human, "Computing Machinery and Intelligence" opened with such a three person game. The three persons in the game (called an imitation game) were named A, B, and C, too, and C was to distinguish between A and B. There was a difference, though: the moves being communicated were not positions in a game of chess, but taking one's turn

in a conversation. The distinction was not a matter of distinguishing between which player was a person and which a machine, but between which conversationalist was a man and which was a woman.

The experimental setup in Turing's 1950 paper that I dubbed "The Original Imitation Game Test" is very like a TV game show that premiered six years later, in 1956, called "To Tell The Truth." It was played as follows:

> " Three challengers are introduced, all claiming to be the central character.
>
> [. . .] the host reads aloud a signed affadavit about the central character.
>
> The panelists are each given a period of time to question the challengers. Questions are directed to the challengers by number (Number One, Number Two and Number Three), with the central character sworn to give truthful answers, and the impostors permitted to lie and pretend to be the central character.
>
> After questioning is complete, each member of the panel votes on which of the challengers they believe to be the central character, [. . .] Once the votes are cast, the host asks, "Will the real [person's name] please stand up?" The central character then stands, [. . .] Prize money is awarded to the challengers based on the number of incorrect votes the impostors draw." ("To Tell the Truth", Wikipedia, downloaded Jan. 27, 2012. http://en.wikipedia.org/wiki/To_Tell_the_Truth#1956.E2.80.931968.2C_CBS)

Being a convincing imposter can be difficult. In previously published work, I have argued that the OIG Test is a better game than the one currently referred to as "the Turing Test." (Sterrett 2000, Sterrett 2002a, Sterrett 2002b). One reason I gave for my view was that the task given the machine in the OIG Test is the same as the task set for the human in the OIG Test: to imitate something that it is not.[3] The concept of a machine being set the task of imitation should not seem at all foreign here -- in fact, the term "imitation" is used by Turing in describing a universal machine; he speaks of the ability of a universal machine to imitate other machines. Isn't imitation a straightforward task for a computer, then, you may ask?

No, I don't think it is. For an uneducated machine (such as an uneducated universal machine) to imitate is one thing -- it amounts to implementing a program. For an educated machine to imitate is quite another. In fact, I argued, what is called for is not really imitation, but figuring out and carrying out what it takes to be a convincing imposter. While the central character in "To Tell the Truth" may give an answer to a question without any fear of being led to another question that he or she cannot answer, the imposter has to think how to keep the conversation from turning to topics that might present problems for an imposter. An imposter has to constantly be on guard to override

---

[3] In (Sterrett 2000) I also show that the two tests in (Turing 1950) give different *quantitative*, as well as qualitative, results. I consider it a major contribution of (Sterrett 2000) to give what amounts to a proof that the two tests are unequivocably different on significant points, and that the OIG Test need not be set up around gender differences. Secondary literature citing (Sterrett 2000) has not always recognized these two major points.

tendencies to respond in ways that have by now become habitual, but which are inappropriate while posing as an imposter. Talk of overriding habits is, I believe, no longer fanciful talk, as reading about work done by robotics researchers on the difficulties faced in applying imitation learning in robotics will reveal. If IBM is looking for suggestions for its next Grand Challenge, let me suggest "To Tell the Truth." [4]

I shall not repeat all the points I made in those earlier works on Turing and tests for intelligence. Rather, my point in this talk about the OIG Test concerns a question germane to the education of Turing's Child-Machines. I suggest that reflecting on the question of how machines produced using different methods for educating machines fare on the OIG Test leads to useful ways of thinking about machine intelligence.

References

BCS Computer Conservation Society (2002) "Recollections of early AI in Britain: 1942 - 1965". (video for the BCS Computer Conservation Society's October 2002 Conference on the history of AI in Britain) transcript downloaded from http://www.aiai.ed.ac.uk/events/ccs2002/CCS-early-british-ai-dmichie.pdf on March 25, 2012.

Floyd, Juliet (to appear) "Turing, Wittgenstein, and Types: Philosophical Aspects of Turing's 'The Reform of Mathematical Notation and Phraseology' (1944-5)", in *Alan Turing - His Work and Impact*, eds. S. Barry Cooper and Jan van Leeuwen (*The Collected Works of A. M. Turing*, revised edn of North-Holland 2001, Elsevier).

James, William (1890) *The Principles of Psychology, Volume I.* New York: Henry Holt and company.

Michie, Donald (1961) "Trial and Error", *Science Survey*, part 2, Harmondsworth: Penguin, pp. 129 - 145.

Quinn, Naomi (2003) "Cultural Selves" *Annals of the New York Academy of Sciences*, 1001: 145-176.

Sterrett, S. G. (2002b) "Too Many Instincts: Contrasting Philosophical Views on Intelligence in Humans and Non-Humans", *JETAI* (Journal of Experimental and Theoretical Artificial Intelligence), Vol. 14, No. 1, pp. 39 - 60. Reprinted in *Thinking About Android Epistemology*, Edited by Ken Ford, Clark Glymour and Patrick Hayes, MIT Press (March 2006).

Sterrett, S. G. (2002a) "Nested Algorithms and ´The Original Imitation Game Test´: A Reply to James Moor" *Minds and Machines*, Vol. 12, pp. 131-136.

Sterrett, Susan G. (2000) "Turing´s Two Tests for Intelligence" *Minds and Machines*, Vol. 10, pp. 541-559. Reprinted in *The Turing Test: The Elusive Standard of Artificial Intelligence*. Edited by James H. Moor. Kluwer Academic, 2003.

Turing, A.M. (1950). Computing machinery and intelligence. *Mind,* 59, 433-460.

Turing, A. M. (1948) "Intelligent Machinery" in *Mechanical Intelligence*, *Collected Works of A. M. Turing.* D. C. Ince, ed. North Holland, 1992, p. 107 -127.

Wikipedia contributors. "To Tell the Truth." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 27 Jan. 2012. Web. 27 Jan. 2012.

[4] IBM's Deep Blue took on the challenge of a machine playing chess at the Grandmaster level. IBM's Watson (with DeepQA technology) took on the challenge of competing in the game show *Jeopardy!*