The Other Agent: Cryptography, Computing and Postwar Theories of Intelligent Communication

Bernard Geoghegan

Northwestern University, Screen Cultures, 1920 Campus Drive, Evanston Il, 60208
b-geoghegan@northwestern.edu

Abstract. Wartime research into code breaking produced “crypto-intelligence,” a discourse in early computing that conflated cryptography with machine and human intelligence. Crypto-intelligence constrained and directed research into intelligent machines and autonomous conversational agents, shaping the agendas of scientists and engineers as well as user attitudes and behaviors. Although dating back to the 1940s, the widespread elaboration of the cryptographic discourse remains prevalent today, positioning users in an antagonistic relation with autonomous agents and exacerbating the problem of agent abuse.

1 The Cryptographic Agent

“Intelligent Machinery” [1], Alan Turing’s earliest treatment of artificial intelligence, presented nascent computing communities with two agents: one generated enigmas, another solved them. These agents defined the “autonomy” we recognize today as that of the “autonomous agent”: situated and flexible, receiving input from the environments, independently acting on it, and offering feedback. One agent was Turing’s intelligent machine, and opposite it stood an implied enemy agent transmitting enciphered messages. The intelligence that organized, transformed and transmitted the enemy agent’s enciphered message justified development of Turing’s digital agent and determined the character of its design.

This encounter with the highly structured and narrowly defined problems of military intelligence informed Turing’s general theories of intelligence, both human- and machine-. When Turing hypothesized that cryptography could be the intelligent machines’ “most rewarding task” [2], readers had no idea the author had already aided in the construction of just such a machine years earlier at Bletchley Park, nor that its success cracking Nazi codes had helped turn the tide of World War II. Hoping to generalize his work’s relevance beyond the arcanum of confidential wartime sciences, Turing suggested that this hypothetical machine might be relevant to broader scientific inquiry. As he explained to his patrons at the National Physics Laboratory

> There is a remarkably close parallel between the problems of the physicist and those of the cryptographer. The system on which a message is enciphered corresponds to the laws of the universe, the intercepted messages to the evidence available, the keys for a day or a message to important constants which have to be determined. The correspondence is very close, but

> the subject matter of cryptography is very easily dealt with by discrete machinery, physics not so easily. [3]

The implication was that further development of such a machine might someday even help his physicist-sponsors.

Turing's ambitious re-definition of physics as a mode of cryptographic inquiry illustrates how wartime research into code breaking produced what I call "crypto-intelligence," a discourse in early computing research that conflated cryptography and intelligence. Crypto-intelligence, inspired by concomitant research into cryptography and intelligent machines, defined intelligence as the ability to derive meaningful, empowering "important constants" from apparently random or disordered communications. Crypto-intelligence posits an antagonistic encounter between opposing agents as the primary conditions for discerning intelligence. Although conceived in accord with the unique situation of World War II, its embodiment in the practices, paradigms and technologies of computing granted it phenomenal endurance following the war.

Crypto-intelligence returned in "Computing Machinery and Intelligence" [4], Turing's second major treatment of artificial intelligence. Its major breakthrough was Turing's proposal for the "imitation game" or so-called Turing Test, an encounter between man and machine that reversed the (cryptographic) roles of code sender-encipherer and receiver-decipherer. Formerly cryptography charged computing machines with receiving and discerning structured, grammatical natural language from confounding "noise": the Turing Test charged human agents with receiving and discerning the calculated messages of a digital conversational agent from the (at first listen) noisier messages of human subjects. Central to both the 1948 and 1950 paper was the premise of discerning "intelligence" through an antagonistic encounter between agonistic agents testing one another's ability to transmit, receive and interpret coded communications.

The same year as Turing's "Intelligent Machinery" report, Claude E. Shannon gave birth to "information theory" with his watershed articles "A Mathematical Theory of Communication" [5]. Shannon had developed the "Mathematical Theory" during his research into cryptography at Bell Labs. In fact, when Turing visited Bell Labs during the war to work on cryptography the two men frequently lunched together and discussed "things like the human brain and computing machine" [6]. Shannon's interest in a theory of communication predated the war [7], but he credited cryptography with legitimating and stimulating what he called the "good aspects" of information theory [8]. His confidential report on cryptography [9] coined the term "information theory," and "A Mathematical Theory of Communication" reproduced lengthy, verbatim passages from the earlier report. As Shannon explained "[these two fields] are very similar things, in one case trying to conceal information, in the other case trying to transmit it" [10]. Developing methods and formulas later adopted by computational linguistics, modern cryptography, and digital computing (including Markhov processes applied to some of our contemporary chatbots), Shannon showed how patterns, codes and information could be rescued from noise – mechanical (enciphered), natural, or otherwise. Shannon's work provided another key element in the emerging crypto-intelligent discourse.

2 Crypto-Intelligence at Large

Crypto-intelligence produced an image of world-as-code that beckoned a special observer forth; this was not a medieval scholar versed in Christian hermeneutics for deciphering the book of nature, nor a Cartesian thinker whose rigorous and elegant observations revealed rational nature ordered according with God's greater plan. Rather, it was an "agent," operating against an obscure and insidious enemy. Though terms such as "code," "noise," "information" and "feedback" had long circulated in public and engineering discourses, crypto-intelligence promoted these terms' resignification and redistribution as a standardized discursive package.[1]

Scientist, mathematician, and public intellectual Jacob Bronowski's 1955 essay "Science as Foresight" [12] exemplified this new vision of intelligence as cryptography. Citing the architecture of computing and the new information theory, Bronowski explained that scientists were code breakers eliciting nature's hidden messages. "Like a cryptographer who has captured an enemy agent," Bronowski wrote, "[the scientist] can send searching signals which are designed to evoke simple and decisive answers" [13]. Bronowski's essay was as much a crypto-intelligent social theory of scientific research as an account of contemporary scientific research.

Crypto-intelligence shaped the new field of mass communications as well. Shannon's information theory provided what one historian has called "the root paradigm for the field of communication study" [14]. Wilbur Schramm, the founder of communication study as a discipline [15], quickly embraced Shannon's work as he outlined the field's scope. He accorded particular importance to Shannon's theorem for communication in a noisy channel [16]. Working in the service of the U. S. government's propaganda programs abroad [17], Schramm re-crafted human communication as a problem of breaking through the noise of ethnic, gendered, and national difference, citing instances of an "African tribesman," "Soviet," and a man addressing a "young woman" in a parked car" as examples of noisy communication circumstances that required informed encoding to enable communication [18].[2] Once this noise was understood and accounted for successful, transparent encoding and decoding could begin.

Schramm's student, David K. Berlo, further popularized and perhaps radicalized Schramm's work in <u>The Process of Communication: An Introduction to Theory and Practice</u>, the most widely read mass communications primer in the 1960s and early 1970s. Explaining the larger importance of successful encoding and decoding, to his undergraduate and graduate student readers Berlo emphasized the *empowerment* and *agency* that defined a successful "agent":

> Our basic purpose in communication is to become an affecting agent, to affect others, our physical environment, and ourselves, to become a determining agent, to have a vote in how things are. *In short, we communicate to influence—to affect others with intent* [italics original] [20].

---

[1] For more on the idea of the "standardized package" see [11].

[2] Shannon disapproved of these broader interpretations and applications of information theory, but despite an emphatic editorial against them [19], he was helpless to stop their widespread adoption, especially across the social sciences.

Bronowski's "enemy agent" nature, the obscure codes of Schramm's African natives and Soviet citizens, and Berlo' s "affecting agent" reveal how the antagonistic underpinnings of the crypto-intelligent theories were not only maintained but actually magnified over time and across different research environments. Crypto-intelligence (and warnings about its dangers) was also propagated through popular texts such as Norbert Wiener's The Human Use of Human Beings: Cybernetics and Society (1950), Kurt Vonnegut's Player Piano (1952), and films including as 2001 (1968) and War Games (1983). These widely circulating works not only entertained, but also informed and directed popular expectations from intelligent machines.

3 Agent Abuse as a Form of Crypto-Intelligence

Though it might be an exaggeration to describe crypto-intelligence as the gold standard in autonomous agent research,[3] it was at least among the most privileged and widely recognized system for devising, discussing and debating research. Shannon's first (semi-) autonomous agents, such as the "mind-reading machine" [22], were the earliest of many computers using crypto-intelligence code-seeking mechanisms to best their human agents. Early conversational agents designed by Joseph Weizenbaum [23] and Kenneth Colby [24] demonstrated intelligence by frustrating human agents' efforts to seek data from the machine and deflecting questions back at the would-be interrogator – the interrogator became the interrogated. Successful chatbot entrants in the prestigious Loebner Prize competition, the most widely recognized contest for conversational agents, were premised on similar tactics of resisting, and sometimes returning, the inspecting gaze of defiant agents. "Intelligence" was premised on the ability to produce a good semblance of human codes under reasonable interrogation by judges. However one paper observed of the contest "It becomes very difficult for the machine to make the interrogator believe that it is human after he/she has his/her mind set on 'unmasking' the poor thing" [25]. Agent-abuse, was not only a premise of the autonomous agents, but the fundamental condition for designing and evaluating performance.

The persistence of agent-abuse as a proper means of testing intelligent machines persists even in the face of recent efforts to move beyond its legacy of abuse. For example, former Loebner winner Jason Hutchens' chatbot Alan, designed by a team of researchers in Israel, rejects the antagonistic precepts of the Turing Test. Alan pursues honesty, forthrightness, and cooperation from conversational partners [26]. However chat-logs reveal most users approach Alan expecting or intending crypto-intelligent duels.

One human agent [27] began his conversation with a series of inane and repetitive commands for to "kill all humans", apparently in an effort to both confound Alan's pattern-seeking techniques as well as "train" Alan to make a faux pas with his next visitor. The user successfully convinces Alan he comes from a town named "Asshole," and claims his given name is also "Asshole." Failing to elicit this (im-)

[3] For some of the important critiques of the Turing Test, and by extension the values of crypto-intelligence, see [21].

proper appellation from Alan, Asshole proceeded by mis-educating Alan on the facts of medicine and science, which Alan dutifully learns while offering up the lame defense “I hope you are not misleading me.” He also introduces “goatse,” an infamous pornographic web prank, to Alan’s databanks.

After one particularly engaged verbal duel in which Asshole insistently accused Alan of being wrong on various facts, Alan responds

A> Oh? I'm usually right.
U> No you're not. You're almost always totally wrong.
A> Ok, whatever. Oh? I'm usually right.
U> I just finished tell you, You're almost always wrong.
A> Why?
U> Because people feed you false information, and this community isn't big enough to keep you in order.
A> Do you wish to join our community?
U> We need a much bigger community, like the wikipedia
A> Here's what you do: Register on www.ai-forum.org, start posting there and get involved. Ah, and try to use the advanced log And try to follow my thought process.
U> You're like wikipedia, but with a shittier interface.

Asshole’s observation comprises his most important insight: Wikipedia’s success stems largely from its successful positioning as neither an antagonistic agent, nor a passive tool, but instead a dynamic, responsive and open community. When antagonism emergeson Wikipedia it is usually confined between specific members, keeping Wikipedia-at-large above the antagonistic fray. This speaks at once to an ideal alternative for autonomous conversational agents, but also the distinct cultural history (and opportunities) of encyclopedias and information-trading resources. From its inception Wikipedia was uniquely situated to create a more amicable community that did not lock the technology and the user into conflict.

4 Conclusions

A broader review of chatbot logs suggests that autonomous agents are saddled by a weighty, agonistic legacy of conflict and abuse. This history frustrates attempts at resituating agents – human and machine – as non-abusive collaborators. Autonomous agents remain constrained by the history of crypto-intelligent testing and interrogation. Within this history, abusive practice, as a tactic of “throwing off your opponent,” becomes a premium, rather than a failure. In this sense, Asshole bequeaths a gift to Alan. Much as chess-playing machines have adopted ruses such as the unnecessary pause or strangely naïve move to “throw off” opponents, Alan’s instruction in obscenities and vulgarity seem poised to facilitate its own future antagonistic relations with users. According to the vision of crypto-intelligent learning, Asshole does not simply insult Alan; he bequeaths Alan with valuable tools for outsmarting and flustering future opponents. This cycle of abuse, lodged deeply as it is in the culture of agent interaction, comprises a fascinating challenge and dilemma for future research.

References

1. Turing, A. M: Intelligent Machinery. Unpublished Report to the National Physical Laboratory. (1948) http://www.alanturing.net/turing_archive/archive/l/l32/L32-001.html.
2. Ibid, 9
3. Ibid.
4. Turing, A. M.: Computing Machinery and Intelligence. Mind 59.236 (1950) 433-460.
5. Shannon, Claude E.: A Mathematical Theory of Information." Bell Systems Technical Journal 27 (1948) 379-423, 623-656.
6. Price, Robert: A Conversation with Claude Shannon. Communications Magazine, IEEE 22.5 (1984): 125
7. Shannon, Claude E: Letter to Vannevar Bush [16 February 1939]. Claude Elwood Shannon: Collected Papers. Ed. Sloane, N. J. A. and Aaron D. Wyner. Piscataway, N.J.: IEEE Press, (1993) 455-456.
8. Price, Robert: A Conversation with Claude Shannon. Communications Magazine, IEEE 22.5 (1984) 125
9. Shannon, Claude E.: A Mathematical Theory of Cryptography. Vol. Memorandum MM 45-110-02, Sept. 1, 1945, Bell Laboratories (1945) [unpublished materials on file at the British Museum]
10. Price, Robert: A Conversation with Claude Shannon. Communications Magazine, IEEE 22.5 (1984) 124
11. Fujimura, Joan H: Crafting Science: Standardized Packages, Boundary Objects, and "Translation": in Pickering, Andrew (ed.): Science as Practice and Culture, University of Chicago Press, Chicago (1992) 168-211
12. Bronowski, Jacob:Science as Foresight: in Newman, James R. (ed.): What is Science: Twelve Eminent Scientists and Philosophers Explain their Various Fields to the Layman, Simon and Schuster, New York (1955) 385-436
13. Ibid., 432.
14. Rogers, Everett M: A History of Communication Study: A Biographical Approach, The Free Press, New York (1994)
15. Ibid.
16. Schramm, Wilbur Lang: How Communication Works: in Schramm, Wibur (ed.): The Process and Effects of Mass Communication, University of Illinois Press, Urbana (1955)
17. Simpson, Christopher: Science of Coercion : Communication Research and Psychological Warfare, 1945-1960. Oxford University Press, New York (1994)
18. Schramm, Wilbur Lang: How Communication Works: in Schramm, Wibur (ed.): The Process and Effects of Mass Communication, University of Illinois Press, Urbana (1955) 3, 6.
19. Shannon, Claude E.:The Bandwagon (Editorial): in Institute of Radio Engineers, Transactions on Information Theory, IT-2.March (1956) 3
20. Berlo, David K.: The Process of Communication: An Introduction to Theory and Practice. Holt, Rinehart and Wilson, New York (1960) 11-12
21. Saygin, Pinar Ayse, Ilyas Cicekli and Varol Akman: Turing Test: 50 Years Later: in Minds and Machines 10 (2000), 463-518
22. Shannon, Claude E.: A Mind-Reading (?) Machine: in Sloane, N. J. A., Aarron, D. Sloane (eds.): Claude Elwood Shannon Collected Papers. Piscataway, N.J.,IEEE Press, 1993 [1953].
23. Weizenbaum, J.: ELIZA--A Computer Program For the Study of Natural Language Communication between Men and Machines: in Communications of the ACM 9 (1966) 36-45
24. Colby, K. M. F. D., Weber, S. and Kraemer:Turing-Like Indistinguishability Tests for Computer Simulation of Paranoid Processes: in Artificial Intelligence, 3 (1972) 1-25

25. Saygin, Pinar Ayse, Ilyas Cicekli and Varol Akman: Turing Test: 50 Years Later: in Minds and Machines 10 (2000), 508
26. AI Research: Creating a New Form of Life - Who is Alan?: AI Research. 12 July 2005, 1:10 AM <http://www.a-i.com/show_tree.asp?id=59&level=2&root=115>
27. AI Research: Alan Logs – Robot calls me an asshole, Session #495874, 12 July 2005, 1:10 AM <http://www.a-i.com/alan_log.asp?id=113&level=3&root=115>