A Conversation with Claude Shannon

One man's approach to problem solving

IN THE COURSE of obtaining information concerning the origins of spread-spectrum communications,[1] Dr. Robert Price visited Dr. Claude E. Shannon. There ensued an extensive (tape-recorded) conversation, slightly edited excerpts of which are presented in this article. These are reorganized for the reader's convenience into four general subject areas: motivation/background for the original Information Theory work; origin of the "entropy" measure in Information Theory; application of noise-like waveforms in multiple-access systems; and Shannon's attraction to problems rather than applications.

Motivation/Background for the Original Information Theory Work

Robert Price: Where and when did you get the idea that information could be modeled as a random process?

Claude Shannon: I was thinking about information theory quite a bit while I was working at Princeton with Hermann Weyl,[2] and while I was at Bell Labs.

When I went to take my National Research Fellowship under Weyl [1940], I told him that I wanted to work on information, the measurement of information, and how much is required. I told him that I had already read Hartley's paper [5], and that it had been an important influence in my life. I think I read it at the University of Michigan; in any event, several years earlier. That paper struck me as important in this area.

Robert Price: Because of the logarithmic measure?

Claude Shannon: Because he was interested in communicating information.

I would say that it was in 1940 that I first started modeling information as a stochastic process or a probabilistic process.

R.P.: Did the idea come to you "out of the blue"? Or because of some Bell Labs work? You were not yet into cryptography in 1940. Do you remember any context in which it occurred to you that if you knew the probability was one or zero ahead of time, there was no information? Hartley's work does not have to have a probabilistic signal, only probabilistic noise. To model the signal as probabilistic, that, to me, is the conceptual breakthrough.

C.S.: I don't regard it as that difficult. What would be the simplest source you might have, the simplest thing you were trying to send? I would think of tossing a coin, heads or tails, and then to try to send that stream of data. That would seem to be the simplest place to start, including the question of whether it is any easier if the coin is loaded. Or if it has six sides like a die, and so forth.

R.P.: In one place in your 1945 cryptography report [6], you actually say "information theory." You just make a passing reference to the phrase, but I think that that report is the first time that those two words appeared together in print.

---

[1]Readers interested in learning more about this work should consult [1-3]; background information concerning the conversation reported herein is presented on pages 86-87 of [2].

[2]This work is discussed extensively in [4]; of particular interest is Shannon's 1939 letter to Vannevar Bush, which is presented on pages 504-505 of [4].

0163-6804/84/0500-0123 $01.00 © 1984 IEEE

"I was a great fan of Edgar Allan Poe's 'The Gold Bug' and stories like that. And I used to solve cryptograms when I was a boy."

C.S.: That could be. That cryptography report is a funny thing because it contains a lot of information theory that I had worked out before, during the five years between 1940 and 1945. Much of that work I did at home.

R.P.: You did that analysis during the war, at home? Wasn't that motivated by cryptography?

C.S.: My first getting at that was information theory, and I used cryptography as a way of legitimatizing the work.

R.P.: Was it an answer looking for a problem? You were delighted to find cryptography coming along during the war as something that was needed and that was a great application of your information theory?

C.S.: In part. I might say that cryptography was there and it seemed to me that this cryptography problem was very closely related to the communications problem. The other thing was that I was not yet ready to write up information theory. For cryptography you could write up anything in any shape, which I did.

R.P.: Do you think, even if there had not been a war effort, you would have been interested in the cryptographic aspects of this?

C.S.: I probably would have been because that's the kind of thing that attracts me. I was a great fan of Edgar Allan Poe's "The Gold Bug" and stories like that. And I used to solve cryptograms when I was a boy.

R.P.: I read that John R. Pierce said that cryptography was an application of information theory. I was pretty sure that that was putting the cart before the horse. I was beginning to think that it was the other way around, and that information theory had come out of cryptography. When I look at this 1945 cryptography report, it has the phrase "information theory" and it says that you are next going to get around to writing up information theory. This makes it sound as if cryptography gave you the mysterious "missing link," but it's now clear that information theory did not come out of cryptography.

C.S.: Working on cryptography led back to the good aspects of information theory. I started with information theory, inspired by Hartley's paper, which was a good paper, but it did not take account of things like noise and best encoding and probabilistic aspects.[3]

R.P.: You have said to other people that these were closely intertwined, and that cryptography was no mere application of information theory. As you say, you got stimulus. Could I suggest that there is a sort of duality there? The cryptography problem is, in some ways, the "mirror image" of the communications problem, so you naturally got some insights out of it.

C.S.: Yes. I believe that I made some remarks about that in one of my papers. I think that all of these sciences and theories stimulate each other to later developments. In my case, I started with Hartley's paper and worked at least two or three years on the problems of information and communications. That would be around 1943 or 1944; and then I started thinking about cryptography and secrecy systems. There is this close connection; they are very similar things, in one case trying to conceal information, and in the other case trying to transmit it.

R.P.: That is why I see a duality there. Entropy measures can be used in both cases.

C.S.: When I came out with my paper in 1948 [7], part of that was taken verbatim from the cryptography report, which had not been published at that time.

Origin of the Entropy Measure in Information Theory

R.P.: It has been said that [John] Von Neumann gave you the word "entropy," saying to use it because you would win every time because no one would understand it and, furthermore, it fitted $p\log(p)$ perfectly [12,13].

I also heard a different version of this story: that you had independently arrived at the word "entropy" and were thinking of using it but were somewhat dubious, and you got reassurances from people like Von Neumann and people at Bell Labs that "entropy" could be used. You had already made that identification and, furthermore, in your cryptography report of 1945, you use the word "entropy"; you liken it to statistical mechanics. Moreover, I don't believe that you were in contact with Von Neumann in 1945. So, it does not seem to me that Von Neumann suggested the word "entropy" to you.

C.S.: No, I don't think he did. I'm quite sure that it did not happen between Von Neumann and me.

R.P.: I think the fact that it is in your 1945 cryptography report establishes that you did not get the idea from Von Neumann. Rather, you had made the $p\log(p)$ identification with entropy by some other means.

Professor [I. J.] Good told me that [Alan] Turing had brought the entropy measure into cryptography in England as early as 1940. Good talked about this in his book, *Weighting of Evidence*, or some title like that, in 1948. But Good alluded to it only very obliquely because it was still under super-secrecy, and it was not until 1974 that this could be talked about openly. However, the entropy measure was

". . . they are very similar things, in one case trying to conceal information, and in the other case trying to transmit it."

[3]Ed. Note: In later discussion, Dr. Shannon also emphasized the importance of Nyquist's work in the development of his thinking in this area. Still later, he introduced the editor to [10], and provided the note accompanying it in the References.

"Turing and I never talked about cryptography. We talked about things like the human brain and computing machines."

in there as a measure of goodness-of-fit and testing on cryptographic cases.

I had the notion, since Turing had already come up with the concept of entropy in 1940, that might have been a lead for you.

C.S.: Not at all! Turing and I never talked about cryptography. We talked about things like the human brain and computing machines.[4]

R.P.: Turing never mentioned that entropy might be an interesting sort of quantity to consider?

C.S.: Not to me.

Application of Noise-Like Waveforms in Multiple Access Systems

R.P.: There is one particular area that I asked your colleagues about. You proposed using a noise carrier for what is now called CDMA [Code Division Multiple Access]. You seemed to start into this by observing that multiplexing is generally achieved by the use of orthogonal functions, and then you went on and said, "Why don't we use quasi-orthogonal functions?" Then it suggested itself to you to use *noise* waveforms.

C.S.: Brilliant idea!

R.P.: Well, your colleagues thought so.

C.S.: I appreciated it, too. I thought it was clever.

R.P.: This is a very applications-oriented idea, very different from information theory. You're actually proposing to mechanize a noise carrier rather than sinewaves. This idea really stands out to me. Both [John R.] Pierce and [Brockway] McMillan said that that idea occurred to you at about the same time that your *Proceedings* paper [8] was published.[5]

C.S.: I think that's right. It would fit logically into my thinking at that time. That's the kind of idea I had at about that time.

R.P.: You want to be quasi-orthogonal to your other friends in the channel. You're sharing the frequency spectrum.

C.S.: Yes. Even more, it seemed like a very democratic way to use up the coordinates that you have, and to distribute the "cost of living," the noise, evenly among everyone. The whole thing seemed to have a great deal of elegance in my mind, mathematically speaking, and even from the point of view of democratic living in the world of communications.

R.P.: And, furthermore, it could be actually *applied*, not like channel capacity; it actually could be instrumented. Therefore, a real-world system could have been configured around it if you had gotten the right encouragement, more than just everyone saying, "Yes, that's a good idea." But nothing further happened to it. Now what's happening is that the FCC has set up a special docket for this very idea. But, in those days, I guess nobody was interested that much in "democracy." Now that the spectrum has gotten more crowded, I can see what you mean by "democracy." A great many people can potentially have access to the same channel, as compared with just one organization having a particular channel on television.

C.S.: I love that part of the idea. More and more people can come, and they would all pay equally, so to speak. If more people were there, gradually the noise level would increase on each channel. But everyone could still talk, even though it might be a pretty noisy "cocktail party" by that time.

R.P.: You conceived that at that time; that's what we *now* call "graceful degradation" in military jargon.

Shannon on Problems vis-a-vis Applications

R.P.: I now want to talk more about military applications.

In your 1949 paper [8], you did this "water pouring" to get the capacity for the colored-noise channel, the nonwhite-noise channel. Given a certain amount of power, it's like water—it always has a positive value, and you distribute this around in the "pockets" where the noise is low. [Robert] Fano may have coined the phrase "water pouring."

So then you ask yourself, *What is the worst possible interference; that is, the interference that will give the lowest capacity?* This is all for a given bandwidth, a given noise, and a given power. You decided that the answer was the flat case. And you said something like, "Whereas this would seem to prove that white gaussian noise is the worst kind of gaussian noise, it's really the worst of *all* noises.

That kind of solution is a game-theory solution; what is the worst kind of noise, and what is the best strategy to transmit against that noise? I looked at that years later and I wondered what your motivations were. Is it conceivable that they could have been military? I asked Brockway McMillan about that, and he said, "No, you were always fascinated by minimax problems." I was reading into that that it is an optimum game-theory jamming/antijamming strategy. I believe that you would agree with McMillan that it is an interesting exercise in minimax thinking.

C.S.: I am very seldom interested in applications. I am more interested in the elegance of a problem. Is it a good problem, an interesting problem?

". . . it seemed like a very democratic way to use up the coordinates that you have, and to distribute the 'cost of living,' the noise, evenly among everyone."

[4]This portion of the conversation is alluding to Turing's secretive visit to Bell Telephone Laboratories during World War II; more information concerning this visit can be found on pages 86, 87, and 90 of [2]. Also, Hodges [10] mentions this trip and other relationships between Turing and information theory.

[5]In a 1949 memo [9], Pierce states that "C. E. Shannon suggested some time ago. . .using as 'code functions' voltages which are approximately orthogonal functions of time. . .for instance, . . . noise signals. . .over a long period of time are approximately orthogonal."

"Like a science-fiction writer, I'm thinking, What if it were like this? *or,* Is there an interesting problem of this type? . . ."

R.P.: After World War II, were you never interested in jamming and antijamming problems? Can you ever recall being interested in jamming and antijamming?

C.S.: Only at the theoretical level indicated in that paper. I was never working on that problem for any government agency. The only application of that work, to my mind, would be to determine the right thing to do if your opponent is trying to do the worst thing against you. Von Neumann's Theory of Games, which came out a little before 1945, was very influential on my thinking here.

R.P.: The Theory of Games for cryptography?

C.S.: Yes. Cryptography, and transmission also.

R.P.: Then, considering the "water pouring" minimax problem, you're now suggesting to me that, in fact, you did have jamming and antijamming in the back of your mind.

C.S.: I think that the only application of the problem that you mentioned (What is the worst noise and the best answer to that noise?) is in the jamming/antijamming battle.

R.P.: It's possible that jamming and antijamming make a sensible application of the minimax approach, but it was not your driving force in presenting that material at that time.

C.S.: Bob, I think you impute a little more practical purpose to my thinking than actually exists. My mind wanders around, and I conceive of different things day and night. Like a science-fiction writer, I'm thinking, *What if it were like this?* or, *Is there an interesting problem of this type?*, and I'm not caring whether someone is working on it or not. It's usually just that I like to solve a problem, and I work on these all the time. This problem, of determining the best strategy against the worst noise, is just the sort of thing that would occur to me. In thinking about, *How would you handle this kind of noise or that kind of noise?*, I would ask, *Well, what would be the* worst *kind of noise?* Suppose someone were trying to do this the worst possible way in constructing noise, then what should be done, and what is the interplay, especially in the background of Von Neumann's Theory of Games, which had just come out about then.

R.P.: So it seems that you worked on problems for the sake of understanding and getting the answers, without worrying about military applications or getting sidetracked by putting too much into applications of any kind.

References

[1] R. A. Scholtz, "The origins of spread-spectrum communications," *IEEE Trans. Commun.*, COM-30, pp. 822–854, May 1982; also reprinted in [11].

[2] R. Price, "Further notes and anecdotes on spread-spectrum origins," *IEEE Trans. Commun.*, COM-31, pp. 85–97, Jan. 1983; also reprinted in [11].

[3] R. A. Scholtz, "Notes on spread-spectrum history," *IEEE Trans. Commun.*, COM-31, pp. 82–84, Jan. 1983; also reprinted in [11].

[4] F.-W. Hagemeyer, "Die Entstehung von Informationskonzepten in der Nachrichtentechnik," ["The origin of information theory concepts in communication technology"], Doctoral dissertation, Free Univ. Berlin, Berlin, Germany, Nov. 8, 1979.

[5] R. V. L. Hartley, "Transmission of information," *Bell Syst. Tech. J.*, vol. 7, pp. 535ff, 1928.

[6] C. E. Shannon, "A mathematical theory of cryptography", Bell Tel. Lab. memo, Sept. 1, 1945; later published in revised form as "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, Oct. 1949. [The appearance of the expression "information theory" in the 1945 Bell memorandum contradicts a statement made in the *Journal of Librarianship*, as cited in *A Supplement to the Oxford English Dictionary* (Oxford: Clarendon Press, vol. II, p. 301, 1970): "It is worth noting that Shannon never referred to his theory as an 'Information theory'."]

[7] C. E. Shannon, "A mathematical theory of communication." *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948; and pp. 623–656, Oct. 1948.

[8] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.

[9] J. R. Pierce, "Time division multiplex system with erratic sampling times—Case 38543," Bell Tel. Lab. technical memo MM-49-150-15, June 15, 1949.

[10] A. Hodges, *Alan Turing: The Enigma*, New York: Simon and Schuster, 1983. [The following information was obtained from C. E. Shannon on March 3, 1984: "On p. 552, Hodges cites a Shannon manuscript date of 1940, which is, in fact, a typographical error. While results for coding statistical sources into noiseless channels using the $p\log(p)$ measure were obtained in 1940–1941 (at the Institute for Advanced Study in Princeton), first submission of this work for formal publication occurred soon after World War II."]

[11] *Spread-Spectrum Communications*, C. E. Cook et al., Eds., New York: IEEE Press, 1983.

[12] R. D. Levine and M. Tribus, Eds., *The Maximum Entropy Formalism*, Cambridge, MA: M.I.T. Press, pp. 2 & 3, 1979.

[13] J. Campbell, *Grammatical Man—Information, Entropy, Language, and Life*, NY: Simon & Schuster, pp. 32 & 277, 1982. ■