THE PSI CHANNEL CODING PROBLEM

Luther D. Rudolph

Introduction

Surely the main reason why the laboratory results of parapsychologists have not been widely accepted by the science establishment lies in the elusive nature of psi phenomena. Researchers in extrasensory communication have had to resort to statistical inference in order to demonstrate that communication is actually taking place. Yet Shannon's "noisy-channel coding theorem" states that, for a broad class of channels, if the information rate is kept below the capacity of the channel, then by appropriate design of the encoder and decoder it is possible to reduce the probability of error at the output of the decoder to an arbitrarily small value. This suggests using channel coding to increase the reliability of extrasensory communication to the point where the reality of the phenomena could be verified by direct sensory experience and statistical tests would no longer be necessary. In this paper I consider the question of whether or not Shannon's model applies to extrasensory communication and, if so, what problems must be overcome in order to reap the benefits of channel coding promised in Shannon's theorem.

The primary purpose of this paper is not to solve problems or suggest specific experiments, but to provide a framework in which interesting questions might be asked. The framework is based on Shannon's information theory, which I predict will play an increasingly important role in parapsychological research.

Two characteristics of information theory should be understood at the outset. First, although information theory is couched in signal transmission language, the theory itself does not postulate or depend on any underlying mechanism. Not only is a signal transmission model of communication unnecessary, causality is not even required. That signal transmission language is used throughout this paper is due to long habit, not necessity.

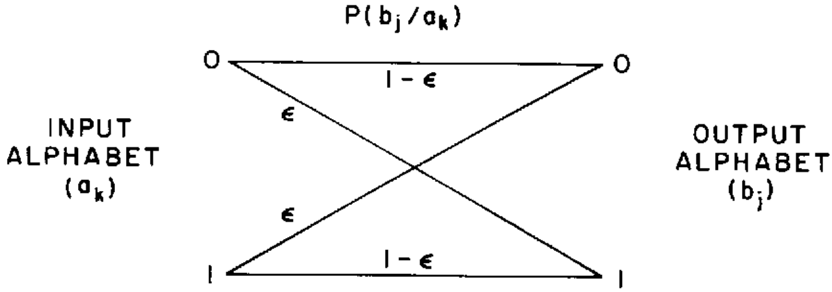

Figure 1. Binary symmetric channel.

Second, a wide range of processes can be modeled as "communication systems," from a radio link to a computer or biological organism. The term "extrasensory communication" is used here in its broadest sense. It applies just as well to a PK task involving a random generator as to a long-distance ESP experiment.

I will first review the fundamental ideas of information theory that relate to channel coding and then consider whether these ideas can be applied to extrasensory communication.

Channel Models and Channel Coding[1]

A channel model requires the specification of the set of possible inputs, the set of possible outputs, and, for each input, a probability measure on the set of outputs. Discrete memoryless channels constitute the simplest class of channel models and are defined as follows: The input is a sequence of symbols from a finite alphabet, $a_1, \ldots, a_K$, the output is a sequence from the same or a different alphabet, $b_1, \ldots, b_J$, and each symbol in the output sequence is statistically dependent only on the symbol in the corresponding position of the input sequence and is determined by a fixed conditional probability assignment $P(b_j|a_k)$.

The simplest non-trivial discrete memoryless channel is the binary symmetric channel (BSC) in which the input and output symbols are binary digits and each digit in the input sequence is reproduced correctly at the channel output with some fixed probability $1 - \epsilon$. The BSC model, which will be used for illustrative purposes later on, is shown in Figure 1.

In order to talk about channel capacity, it is necessary first to define what we mean by "information." In Shannon's theory, this term is used in a highly restricted technical sense to denote a quantitative measure of uncertainty, related to the number of possible outcomes and the probability (i.e. uncertainty) associated with each outcome. Let x denote the input symbol to the channel and y the output symbol. Let $\{a_1, \ldots, a_K\}$ be the X sample space and $\{b_1, \ldots, b_J\}$ the Y sample space in an XY joint ensemble with probability assignment $P(a_k, b_j)$. We want a quantitative measure of how much the occurrence of $y = b_j$ changes the probability of $x = a_k$ from the à priori probability $P(a_k)$ to the à posteriori probability $P(a_k|b_j)$. The quantitative measure which turns out to be useful is the logarithm of the ratio of à posteriori to à priori probability. This gives the following fundamental definition: *the information provided about the event* $x = a_k$ *by the occurrence of the event* $y = b_j$ *is*

$$I(a_k; b_j) = \log \frac{P(a_k|b_j)}{P(a_k)}. \tag{1}$$

We will take the base of the logarithm to be 2, in which case the numerical value of (1) is the number of *bits* of information.

If we interchange the roles of x and y in (1) and apply the identity $P(y|x)P(x) = P(x|y)P(y)$, it is easily seen that the information provided about the event $y = b_j$ by the event $x = a_k$ is also given by (1). Because of this symmetry, $I(a_k;b_j)$ is called the *mutual information* between events $x = a_k$ and $y = b_j$. Mutual information is a random variable with average value

$$I(X; Y) = \sum_{k=1}^{K} \sum_{j=1}^{J} P(a_k,b_j)I(a_k; b_j). \tag{2}$$

The *capacity* C of the channel is the maxium value of average mutual information per channel use, where the maximum is taken over all input probability assignments $P(a_k)$, i.e.,

$$\begin{aligned} C &= \max_{P(a_k)} I(X; Y) \\ &= \max_{P(a_k)} \sum_{k=1}^{K} \sum_{j=1}^{J} P(b_j|a_k)P(a_k) \log_2 \frac{P(a_k|b_j)}{P(a_k)} \qquad \text{bits/symbol.} \end{aligned} \tag{3}$$

(The capacity may, of course, be expressed in bits per second by multiplying C in (3) by the input symbol rate in symbols per second.)

In the case of the BSC, the maximum average mutual information occurs when the à priori input probabilities are $P(0) = P(1) = 1/2$. The capacity of the BSC, as a function of the crossover probability ϵ, is

$$C_{BSC} = 1 + \epsilon \log_2 \epsilon + (1 - \epsilon) \log_2 (1 - \epsilon) \qquad \text{bits/binary digit.} \tag{4}$$

This is plotted in Figure 2. Note that the maximum amount of average mutual information that can be conveyed in one use of the channel is

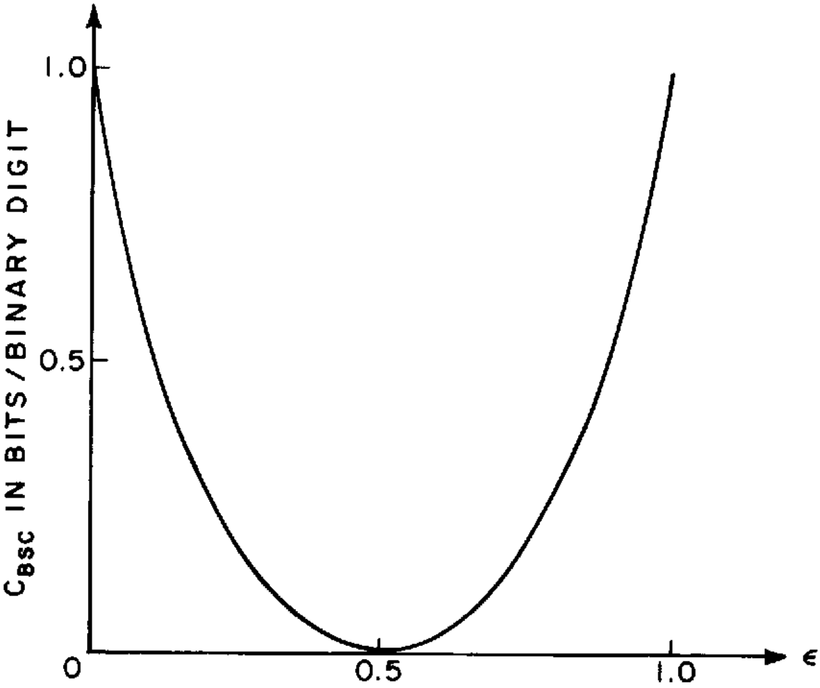

Figure 2. Capacity of the binary symmetric channel.

one bit, and that this occurs either when $\epsilon = 0$ or when $\epsilon = 1$. When $\epsilon = 1/2$, the input and output are statistically independent and the capacity is zero.

It is worth noting at this point that the channel model is completely probabilistic: it depends only on the XY joint ensemble. This means that we need not specify the mechanism which underlies the statistical dependence between channel input and output. In fact, because of the symmetry of mutual information, we need not even say which variable is the input and which the output. Further, nowhere does time enter into the channel model. It is perfectly acceptable for the "output" to occur before the "input." Not only do we not have to postulate some sort of signal energy propagating from a sender to a receiver, we do not even have to postulate a causal relationship between the two, or even think in terms of "sender" and "receiver."

We now turn to the problem of channel coding. Without loss of generality, we will take the input to the channel encoder and the output of the channel decoder to be binary digits, where each digit entering the encoder carries one bit of information (i.e., $P(0) = P(1) = 1/2$). A

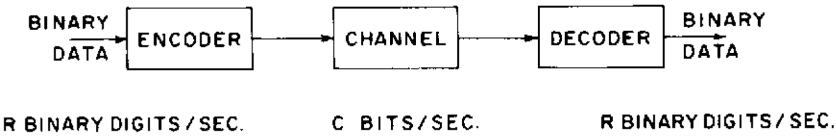

Figure 3. Communication system block diagram.

block diagram of this model is shown in Figure 3. Unlike the channel, which is defined probabilistically, the encoder and decoder are assumed to be deterministic.

The function of the encoder is to produce, for each input data block of binary digits, a unique codeword suitable for transmission over the channel. To combat the effects of the unreliability of the channel, each codeword contains a specified amount of redundancy. The set of all codewords is called an "error-correcting code." The function of the decoder is, given the output of the channel, to determine which codeword was most likely to have been sent. The output of the decoder is then the data block corresponding to this best guess as to the codeword.

The significance of the capacity of a channel (which is here expressed in bits per second) stems primarily from the famous "noisy-channel coding theorem" of Shannon. In imprecise terms, this coding theorem states that, for a broad class of channels, if the channel has capacity C bits per second and if binary data enter the channel encoder at a rate (in binary digits per second) of $R < C$, then by appropriate design of the encoder and decoder, it is possible to reproduce the binary digits at the output of the decoder with as small a probability of error as desired.

Shannon's theorem says that (for long messages) there exists a code which can reduce the probability of error to an arbitrarily small value in spite of the unreliability of the channel. The most obvious way to ensure that a message will get through reliably is simply to repeat it many times and make a decision based on majority vote at the output. However, the repetitive redundancy purchases reliable transmission at the cost of an ever-decreasing transmission rate. The surprising thing about Shannon's theorem is that it promises error-free transmission over an unreliable channel *without further reduction in data rate* given only that the rate is less than the channel capacity.

An example may help to clarify the above. Suppose we are given a BSC with crossover probability $\epsilon = .05$ which accepts binary digits at a rate of 30 binary digits/sec. From (4) we calculate the capacity of this channel to be

$$C_{BSC}(.05) \cong .71 \text{ bits/binary digit}$$
$$= 21.3 \text{ bits/sec.}$$

This means that the data rate going into the encoder must be less than 21.3 binary digits/sec. in order to apply Shannon's theorem. We will choose this data rate to be 10 binary digits/sec. Then for each binary digit which enters the encoder, three binary digits will be sent over the channel. The decoder will in turn produce one binary digit at its output for every three binary digits it receives from the channel. The most common type of code used in this situation is an (n,k) block error-correcting code, where k is the length of the input data block to the encoder and n is the length of the corresponding codeword sent to the channel. The resulting communication system is shown in Figure 4.

The simplest (n,k) code with $k/n = 1/3$ is the (3,1) code. Here the encoder simply triplicates the digit at its input and the decoder makes a 2-out-of-3 majority decision on the triple it receives from the channel. This code can correct any single error in the transmitted 3-digit codeword and a straightforward counting argument shows that the probability of error at the output of the decoder is $p = (3)(.95)^1(.05)^2 + (.05)^3 \cong .00725$. So the probability of error in the transmission of a data block consisting of one binary digit has been reduced from .05 when no coding is used to .00725 when the (3,1) code is used. Of course, this gain in reliability is paid for by the reduction in data rate from 30 binary digits/sec. to 10 binary digits/sec. We might well question whether it was worth using the code at all. But now Shannon's theorem comes into play. It says that, by going to longer codes with the same $k/n = 1/3$, we can decrease the probability of error to an arbitrarily small value with no further penalty in data rate. To illustrate this effect, we will now analyze the performance of an $(n,k) = (15,5)$ code.

The encoder for a (15,5) code accepts a data block of 5 binary digits and produces a codeword of 15 binary digits. There are $2^5 = 32$ possible 5-digit data blocks and thus 32 codewords. Again, the simplest approach would be for the encoder to simply triplicate the

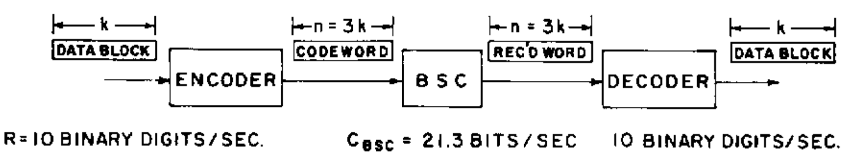

Figure 4. Coded BSC system.

input 5-digit data block and for the decoder to decide which codeword (data block triple) was most likely sent by simple majority vote. A moment's reflection, however, reveals that nothing would be gained over the (3,1) code used previously. In this case, simple repetition is not an efficient way to structure the redundancy. A (15,5) code which has optimally efficient redundancy is shown in Table I. Notice that every possible pair of codewords differs in at least 7 positions, so that the decoder can correctly decode any received word which contains no more than 3 errors. If the received word contains more than 3 errors, the decoder may or may not decode correctly, depending upon the particular error pattern. A more complex version of the counting argument used in the case of the (3,1) code

TABLE 1

(15,5) Error-Correcting Code

Data Block					Codeword														
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
0	0	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
0	0	0	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
0	0	1	0	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
0	0	1	0	1	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0
0	0	1	1	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0
0	0	1	1	1	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1
0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	1	1	0	1	0	1	0	1	1	0	1	0	1	0	1	0
0	1	0	1	0	0	1	1	0	0	1	1	1	1	0	0	1	1	0	0
0	1	0	1	1	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1
0	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0
0	1	1	0	1	1	0	1	1	0	1	0	1	0	1	0	0	1	0	1
0	1	1	1	0	0	1	1	1	1	0	0	1	1	0	0	0	0	1	1
0	1	1	1	1	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0
1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
1	0	0	1	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
1	0	1	0	0	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
1	0	1	0	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1
1	0	1	1	0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1
1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0
1	1	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
1	1	0	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1
1	1	0	1	0	1	0	0	1	1	0	0	0	0	1	1	0	0	1	1
1	1	0	1	1	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0
1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1
1	1	1	0	1	0	1	0	0	1	0	1	0	1	0	1	1	0	1	0
1	1	1	1	0	1	0	0	0	0	1	1	0	0	1	1	1	1	0	0
1	1	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	0	0	1

shows that the probability of error at the output of the decoder is approximately $p_b = .0037$. Of course, this is the probability of a data block error; in order to compare the (15,5) code to the (3,1) code, we need the probability that a single data digit is in error. A conservative assumption is that when the decoder makes a mistake, its output is equally likely to be any one of the 31 incorrect data blocks. In this case, a particular digit in the 5-digit output has about a 50 percent chance of being correct. So the probability of output digit error is $p = .5 \times .0037 = .00185$. Thus the (15,5) code yields a lower probability of error than the (3,1) code at no further penalty in data rate.

The improvement in reliability realized when we moved from the (3,1) to the (15,5) code was not dramatic. This is partly because an increase in code length from 3 to 15 is not large, and partly because the improvement obtainable as a function of code length is dependent upon the initial reliability of the channel. The better the channel, the greater the percent improvement for a given increase in code length. By conventional communication system standards, a BSC with $\epsilon = .05$ is not a very reliable channel. We know that (properly chosen) longer codes with $k/n = 1/3$ will do even better, but very long codes would be required for really reliable communication if the initial channel were a really poor one. (Of course, when we use long codes, the question of whether we can *decode* in a reasonable time and at a reasonable cost becomes important. This problem is treated in the field of coding theory.[2] Suffice it here to say that long codes have been found which can be decoded with relative ease.)

Parapsychologists do not have the luxury of dealing with a channel which has a probability of error of .05 at a data rate of 30 binary digits/sec. The data rate is not unreasonable for certain types of experiments, but the probability of error is more likely to be on the order of .49. The shape of the channel capacity curve near the .50 point warns that this will be a problem. But before considering such practical problems, we must first ask whether, in principle, Shannon's theory can be applied to psi processes.

The Psi Channel

We first ask whether Shannon's channel model is applicable. The answer here is clearly yes. A "channel" in information theory is simply two variables together with the probability measure that relates them. Any psi experiment that lends itself to statistical analysis, therefore, also lends itself to channel modeling.

For example, an ESP card guessing experiment could be modeled as a discrete memoryless channel with a 5-symbol input/output alphabet and assigned probabilities $1-4\epsilon$ for a correct call and ϵ for an incorrect call as shown in Figure 5. The assumptions underlying this model are:

(1) The n^{th} call is statistically dependent only on the n^{th} target symbol.
(2) The probability of an incorrect call is independent of both target and call symbol.
(3) The conditional probabilities do not vary with time.

Are these assumptions reasonable in a card guessing experiment? Clearly not. But they *are* conservative. The equiprobable discrete memoryless channel is a worst-case model in the sense that its capacity is less than or equal to the capacity of any other channel with the same size input/output alphabet and average statistics. Any divergence from equiprobability or independence can only increase the capacity since such divergence constitutes additional information about the channel which can (in principle at least) be used to advantage. The random-error channel is in this sense the most difficult channel to deal with and hence has the lowest capacity. This lower bound on the capacity of the actual channel, however poor the bound may be, provides a basis for applying Shannon's channel coding theorem. As more information about the channel is obtained, better models with higher capacities can be constructed, but in the meantime we can attempt to increase reliability by the use of channel coding based on the worst-case model.

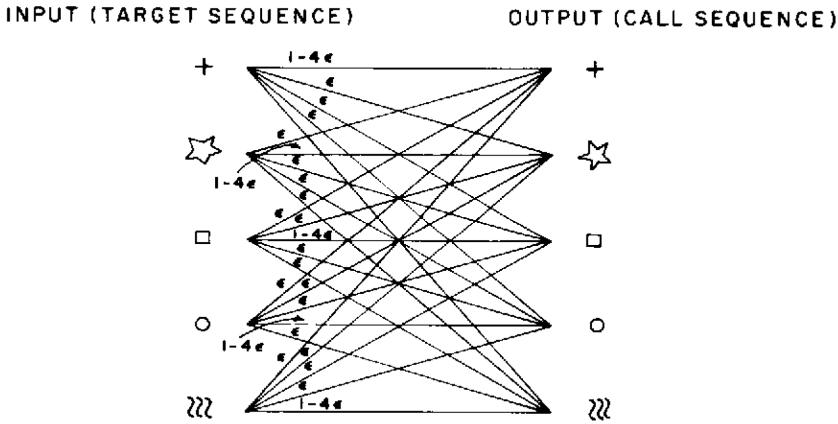

Figure 5. Open-deck card-guessing channel model.

We now turn to the question of whether Shannon's channel-coding theorem can be applied to the psi channel. And here we run into trouble. A channel involves only two variables. The addition of an encoder and a decoder, however, introduces two new variables. Since any two variables constitute a channel, we now have six channels as shown in Figure 6. (The encoder and decoder are the wx and yz channels.) I am assuming here that the auxiliary channels are due to observer effects, not to indeterminacy in the encoder and decoder.

In a conventional communication system we assume that the capacities of these auxiliary channels are zero, but we are certainly not justified in doing so here. If, as is widely believed, psi effects transcend distance and time, then we must allow the possibility of auxiliary channels with nonzero capacities. And this in turn implies the possibility that all attempts to improve the reliability of the original channel through the use of channel coding may be "short circuited" by these auxiliary channels. The addition of the encoder and decoder could thus introduce an "observer effect" which would cancel out the benefits of coding and result in no net gain in reliability.

Only laboratory tests can determine whether or not these auxiliary channels will be a serious problem in practice. However, there are some encouraging indications. First, the use of statistical inference to evaluate the results of experiments consisting of many repeated trials does not cause a complete deterioration of the psi effect. Using statistics to evaluate the results of an experiment consisting of n repeated trials is not unlike the use of an (n,1) repetition code in a coded communication system. That an impressive result such as '$p < 10^{-6}$' can be "observed" by many other researchers (through journal publication, etc.,) without apparent adverse effect, is a hopeful sign. Of course, this does not imply that a similar resistance to observer effects would necessarily obtain if we tried to produce a *physical* output with the same order of reliability, but the reported successes of "majority vote" experiments

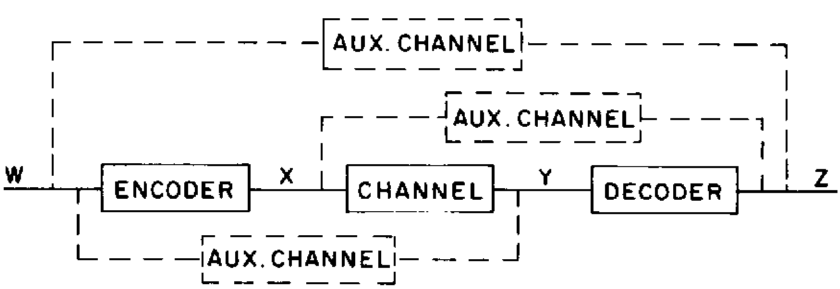

Figure 6. Coded channel with auxiliary channels shown.

gives hope that this might be the case. Stanford[3] and Kennedy[4] have interesting discussions of the majority vote technique.

There is a possible problem, then, in applying Shannon's coding theorem to the psi channel. There are indications that this may not be an insurmountable obstacle, but it is certainly well to be aware of possible observer effects caused by the introduction of channel coding. Who sees what (and how they think about it) may be, and very likely are, important factors. But even if observer effects acting over the auxiliary channels do not wash out our attempts to improve psi process reliability, we still face practical difficulties.

The Data Rate Problem

In this section, we will assume that the capacities of the auxiliary channels are zero. (This is equivalent to what Kennedy[4] has called the "majority vote hypothesis" and is probably much too optimistic.) The problem then reduces to finding a coding scheme that will give the desired performance. To give a feel for the sorts of problems involved, we will consider a simple example.

Suppose we take as our channel model a BSC with $\epsilon = .49$ which accepts binary digits at a rate of ten per second. (This could correspond to a binary random generator experiment with 10 trials/sec. and an average scoring rate of 51 percent.) The capacity of this channel is, from (4),

$$C_{BSC}(.49) \cong .0003 \text{ bits/binary digit}$$
$$= .003 \text{ bits/sec.}$$

In order to apply Shannon's theorem, the k/n for the block code must be less than C_{BSC}, say $k/n = .0002$. The length of the code word sent to the channel must thus be 5000 times the length of the data block entering the encoder. The information rate into the encoder is then .002 bits/sec., which is also the data rate, in binary digits/sec., assuming that each binary digit entering the encoder carries one bit of information (i.e., $P(0) = P(1) = 1/2$). The simplest code we could use in this case is the (5000,1) repetition code. The probability that the decoder would correctly decode a 5000-digit received word by majority vote is approximately $p = 0.92$. To further increase the reliability, we would resort to longer codes with the same k/n as illustrated in the previous section.

In this example, it would take more than eight minutes to transmit one data bit and perhaps several hours to transmit one codeword, if any code other than the (5000,1) repetition code were used. Aside from the fact that such a low data rate would be of little interest, there are practical problems which make such a rate virtually impossible to sustain. First, there is the very real problem of finding subjects able and willing to serve as part of the "channel." If each channel digit must receive individual attention, as seems likely if a washout of the psi effect due to complexity-independence is to be avoided,[3–5] then the transmission of even one codeword would be an impossibly tedious task. Furthermore, the longer the transmission time for a codeword, the greater the probability that the channel will shift from psi-hitting to psi-missing or otherwise exhibit non-stationary behavior. Therefore, it would seem imperative that we be able to reduce the time required to transmit a codeword. Three approaches come to mind: (1) increase the channel symbol rate, (2) go to a larger channel alphabet or (3) use parallel channels.

Experiments such as those of Schmidt[6] suggest that increasing the channel symbol rate is not the answer. Clearly, if individual attention to each channel symbol is required, then the symbol rate is necessarily limited to human sequential data processing rates. In communication system terminology, we would say that the channel is bandlimited. We might be willing to push the symbol rate beyond the limits of human sequential processing speeds if the strength of the psi effect did not fall off too rapidly. For the BSC, however, the capacity of the channel varies as the square of the deviation from chance for small deviations, which suggests that even a relatively slow fall off of psi strength with increased channel symbol rate would be unacceptable.

The second approach—going to a larger channel alphabet—is the standard approach to increasing information rate over a bandlimited channel. The idea is to increase the amount of information per channel symbol by reducing the á priori probabilities of the symbols. Thus, with the 5-symbol alphabet employed in ESP card guessing experiments, the maximum amount of information that could be conveyed by one channel symbol is $\log_2 5 = 2.32$ bits as compared to $\log_2 2 = 1$ bit in the binary case. (The amount of information contained in one "symbol" of a free-response ESP experiment can be practically unlimited. However, quantifying this information is a difficult problem and the symbol rate is extremely low.) Whether the use of large alphabets will be fruitful or not depends upon the relationship between the à priori probability of a hit and psi strength.[4] At present, all that can be said is that if the use of large alphabets proves to be beneficial, then the channel capacity can be increased and the code length required for reliable communication correspondingly reduced.

The third approach—the use of parallel channels—is closely related to the second approach. Here, too, the idea is to increase the amount of information in one "use" of the channel(s). One use in this case is the simultaneous transmission of one channel symbol (not necessarily the same one) over each of N parallel channels. One could think of using a different subject on each of the N channels, but this is not the most attractive of the possibilities. Better would be a situation in which the attention of a single subject could be distributed over the N channels. A multiple-dice-throwing experiment would be a classical example of this. As a more modern example, a subject might attempt to influence simultaneously N random generators by receiving feedback in the form of a visual pattern, where different parts or aspects of the pattern are controlled by different generators. The use of parallel channels opens up the possibility of coding across channels as well as in time, which in this case could be accomplished by setting a high-aim/low-aim switch on the i^{th} generator according to the value of the i^{th} digit in a binary code-word of length N. The output of the channel would then be an N-vector whose i^{th} component is some measure of the action of the i^{th} random generator. Binary encoding and decoding would be performed in conventional fashion. As far as the subject would be concerned, his task would be the same regardless of what codeword was "transmitted." (I am indebted to Helmut Schmidt for pointing out this implementation of parallel channels.[7]) Of course, the parallel channels approach depends on the ability of the subject to have the same order of effect simultaneously on N channels as he has on one channel. The hope that this might be possible is based on an analogy with human information processing capabilities. Humans are much better parallel processors than sequential processors. Studies of human information processing show that most people can handle only about seven "chunks" of information sequentially without getting confused, but that the complexity of a "chunk" can be varied over a wide range without substantial loss.

Discussion

A coding theorist on first venturing into the field of parapsychology and reviewing the decades of debate over the statistical evidence of a weak psi effect in the laboratory is very apt to ask himself—as I did—why redundancy in the form of channel coding has not been used to provide a reliable and convincing physical demonstration of the existence of the phenomenon.

After a little reflection, it became clear to me that redundancy in the form of a crude channel code, namely, the (n,1) repetition code, had been used from the very start, first in the form of repeated trials and later in the form of "majority vote" experiments. In almost all cases, however, the redundancy has been used in an attempt to increase statistical significance rather than to provide a convincing physical demonstration. (As Kennedy so aptly puts it, ". . . statistical (rather than practical) significance has become the standard for evaluating psi effects."[8]) The reasons for this seem to me to be: (1) the channel has not been characterized and is therefore unpredictable, (2) the data rates achieved to date are too low to support a real-time physical demonstration and (3) the coding scheme being used is not powerful enought to guarantee reliable results.

On the first point, it must be said that it is extremely difficult to characterize an unknown channel which operates at a very low average signal-to-noise ratio, as does the psi channel. One would hope that channel coding could be used to improve the signal-to-noise ratio and make characterization easier. But one of the fundamental results of information theory is that for reliable and efficient communication, the code must be matched to the channel. And how are we to do this if we know next to nothing about the channel? It is a circular problem which requires that we pull ourselves up by our bootstraps. But as with other situations of this sort, once a little headway is made we can expect very rapid progress.

On the second point, the approach to the low data rate problem which appears most promising to me is increased parallelism. It seems likely that we are dealing with a channel which has a high capacity, but which is severely bandlimited. In this case, the use of parallel channels and/or large channel alphabets is indicated.

On the third point, it is not clear that a more powerful coding scheme will solve the reliability problem. Given a weak but steady statistical effect on a conventional channel, Shannon's theorem guarantees that we can achieve completely reliable communication. But Shannon's theorem may not be applicable to extrasensory communication because of the auxiliary channels created when an encoder and decoder are introduced. An attempt to generalize Shannon's theory to take into account these auxiliary channels would be a most worthwhile undertaking. But even if it turns out that Shannon's theorem cannot be so generalized, the information-theoretic model with auxilliary channels will still provide a useful framework in which to consider such observer-theoretic questions as: who should receive trial-by-trial feedback and who should receive only summary feedback? I will hazard the speculation that the existence of auxiliary channels actually *increases* the overall channel capacity of the system, and that one of the keys to achieving reliable communication is the use of these auxiliary channels so that they interfere constructively rather than destructively.

I would like to close by sharing a beginner's intuitive feeling about the ultimate reliability of extrasensory communication: either such communication can be made virtually error-free, or else it will never be much more reliable than it is today. It all depends on the nature of the underlying mechanism.

BIBLIOGRAPHY

1. This section borrows freely from R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.
2. See, for example, W. W. Peterson and E. J. Weldon, Jr., *Error-Correcting Codes*, 2nd Ed., MIT Press, Cambridge, 1972.
3. R. G. Stanford, "Experimental psychokinesis: A review from diverse perspectives," in B. B. Wolman (Ed.), *Handbook of Parapsychology*, Van Nostrand Reinhold, New York, 1977.
4. J. E. Kennedy, "The role of task complexity in PK: A review," *Journal of Parapsychology*, 1978, *42*, 89–122.
5. H. Schmidt, "Comparison of PK action on two different random number generators," *Journal of Parapsychology*, 1974, *38*, 47–55.
6. H. Schmidt, "PK tests with a high-speed random number generator," *Journal of Parapsychology*, 1973, *37*, 105–118.
7. H. Schmidt, private communication, April, 1979.
8. J. E. Kennedy, private communication, July, 1979.

DISCUSSION

MORRIS: How might you attempt to apply these ideas to sender/receiver situations as was talked about in Dr. Byer's paper, in which a lot of information is being simultaneously generated and exchanged, e.g., two opponents in a basketball game. I gather this has been one of the problems with Shannon's theory in general; it's harder to apply in that kind of complex circumstance.

RUDOLPH: Well, I take a very operational view of that. I only consider what shows up in the two variables that I'm looking at. The effect observed may be due to all sorts of influences, e.g. the sharing of common emotional states, as was spoken about yesterday. Nevertheless, if I'm only looking at the two variables, that is the channel. Taking all these other factors into consideration gets us into the area of "how do I get a better channel?" I suspect that this may involve getting constructive rather than destructive interference among all these factors. I've got an experiment in mind which would decouple the sender and receiver from the percipient who is trying to produce the psi effect. The idea is that an outside observer would provide the information and would receive the information, but the channel capacity would be created by a psi source in a different setting. So that rather than having a pair of percipients try to communicate, I would use one percipient to create the channel and let other people use the channel.

MORRIS: You would then selectively study circumstances in which that would be less likely to occur.

RUDOLPH: Yes.

RUDERFER: How do you define your channel?

RUDOLPH: A channel is simply two variables and the probability assignment that relates them.

RUDERFER: Well, actual channels have to have more than that.

RUDOLPH: I'm talking about the channel model of Shannon. It's a mathematical model.

RUDERFER: So you're talking about mathematics only, a mathematical model only.

RUDOLPH: That's right.

RUDERFER: You're also talking about a noisy channel and so does Shannon. How do you get noise in a mathematical model?

RUDOLPH: That's incorporated into the probabilities.

RUDERFER: Yes, but where does it arise from physically?

RUDOLPH: I don't know. Shannon doesn't consider that. He looks at the effects of noise, but his mathematical model does not address itself to the physical mechanism involved.

RUDERFER: Then if you're excluding all physical phenomena, your statement about energy not applying is not applicable. You're talking about a mathematical equation and its interpretation or a mathematical model and its interpretation; therefore, energy is precluded only because of that, not for any other reason.

RUDOLPH: I'm dealing with a mathematical theory and it's just like applying group theory to crystallography. It is a mathematical theory that is applied to physical channels to improve their reliability. It is used every day. But the theory itself does not require a physical interpretation.

RUDERFER: Which means that when you go into the physical area, you have to add it according to the requirements of the properties of an actual channel.

RUDOLPH: In the design of the encoder and the decoder, you need know nothing about the channel other than its statistics.

ROSENTHAL: Being such an interdisciplinary area, parapsychology focusing on the same problems may forget the sciences of origin they come from. I think the people coming from the physical sciences may not realize how fuzzy and noisy and murky the signal/noise ratio is in the behavioral and social sciences right now, so that for the physical scientists there may be a world of difference between their traditional science and psi phenomena. For the psychologist, for example, or for the sociologist, this is an old problem. My hunch is that the order of magnitude of the size of the effect in some of the psi phenomena may be very much on the order of some of the effects in general psychology. Let me give you an example: Psychiatric diagnosis is notoriously unreliable. We can make it more reliable by adding more psychiatrists and getting a majority of opinions, but I would argue that in a sense that is statistical. That is, rather than increasing the number of patients in our research study who are being diagnosed, what we're doing is increasing the precision of definition of each of the patients, so, in a sense, we are reaping some statistical benefits. I think we've been doing that a long time in the behavioral sciences. I think psi may not be as badly off as people who come from the physical science tradition think it is.

RUDOLPH: I'm thinking of how to get through to the people outside of the field. I'm not comfortable with statistical inference and I know a lot of my colleagues aren't. Suppose something is significant at the five percent level. Nobody I know is really very impressed by that. Yet, if I can have a physical demonstration that works nineteen out of twenty times, they will be impressed and it's the same order of effect. I guess I'm just making a plea to use the redundancy in psi experiments for physical demonstration, just so I can convince my friends and not have to show them statistics which they don't like. But statistics are a very, very useful tool and I don't mean to imply otherwise.

MORRIS: You note that impressive results, such as P less than 10^{-6}, can be observed by many other researchers through journal publications, without apparent adverse effect. I'm not sure that's true. The lore within the parapsychological community is that if you're halfway through a study, don't present the half of your data even to chums and buddies, because it will plummet at the end. There's even an example of that in the literature. In *Parapsychology from Duke to FRNM*, there is a significant progress report by Rex Stanford of an EEG/ESP study which, when you read the later publication, turns out not to be significant overall. Some people would say that a major portion, perhaps, of parapsychology's replicability problems is that once one announces a result and that therefore one has something to repeat, one has immediately involved a rather large observational community, who may be a little testy about what they observe. I'm not pleased by that interpretation of things, as I like to do the ordinary business of science and generate results for all to observe and utilize.

RUDOLPH: I agree. But there are journal articles that do give impressive results like that.

MORRIS: That I agree with, and the question of asking whether or not the observer effect affects only "not yet" but not "already," I think may re-insert time back into things, because any act of measurement separates the universe into "already" versus "not yet." Even when you say your model is time independent, you sooner or later take an observation separating the universe into "already" versus "not yet." If you're trying to assess a precognition study, all you could say at any given moment was whether or not the precognitive statement had been validated "already" versus "not yet." So time is back in there.

RUDOLPH: Yes. And I share your uneasiness about the assumption of unlimited observer effects, which can certainly cause a lot of paranoia.