ACTA SYSTEMICA

Volume XII
No. 1
(2012)

International Journal
of
The International Institute for
Advanced Studies
in Systems Research
and Cybernetics

George E. Lasker
Editor-in-Chief

Published by
The International Institute for Advanced Studies
in Systems Research and Cybernetics

ISSN 1813-4769

Confusing the "Confusion Matrix": The Misapplication of Shannon Information Theory in Sensory Psychology

Lance Nizami

Independent Research Scholar, Palo Alto, California, USA

nizamii2@att.net

Abstract

Information flow in a system is a core cybernetics concept. It has been used frequently in Sensory Psychology since 1951. There, Shannon Information Theory was used to calculate "information transmitted" in "absolute identification" experiments involving human subjects. Originally, in Shannon's "system", any symbol received ("outcome") is among the symbols sent ("events"). Not all symbols are received as transmitted, hence an indirect noise measure is calculated, "information transmitted", which requires knowing the confusion matrix, its columns labeled by "event" and its rows labeled by "outcome". Each matrix entry is dependent upon the frequency with which a particular outcome corresponds to a particular event. However, for the sensory psychologist, stimulus intensities are "events"; the experimenter partitions the intensity continuum into ranges called "stimulus categories" and "response categories", such that each confusion-matrix entry represents the frequency with which a stimulus from a stimulus category falls within a particular response category. Of course, a stimulus evokes a sensation, and the subject's immediate memory of it is compared to the memories of sensations learned during practice, to make a categorization. Categorizing thus introduces "false noise", which is only removed if categorizations can be converted back to their hypothetical evoking stimuli. But sensations and categorizations are both statistically distributed, and the stimulus that corresponds to a given mean categorization cannot be known from only the latter; the relation of intensity to mean sensation, and of mean sensation to mean categorization, are needed. Neither, however, are presently knowable. This is a quandary, which arose because sensory psychologists ignored an ubiquitous component of Shannon's "system", the uninvolved observer, who calculates "information transmitted". Human sensory systems, however, are within *de facto* observers, making "false noise" inevitable.

Keywords: Shannon Information Theory, sensory psychology, observer

Introduction

A core cybernetics concept, *Information flow in a system*, has been extensively used in Sensory Psychology since 1951, when the "information transmitted" of Shannon Information Theory (Shannon, 1948) was first computed for the results of "absolute identification" experiments using human observers (Garner & Hake, 1951). It was believed that a fundamental aspect of human capability was being measured, the "channel capacity", famously proselytized early on as "The Magical Number Seven, Plus Or Minus Two: Some Limits On Our Capacity For Processing Information" (Miller, 1956). Publicized by Miller, the Garner & Hake computation was enormously influential in the study of behavior; Garner & Hake, with Miller, together have been cited over 4,750 times by workers in psychology, systems theory, management, information engineering, human factors, music, neurology, and general cybernetics. The numbers are from ISI Web of Science; however, according to GoogleScholar, Miller (1956) alone has received nearly 15,000 citations.

Computations of "channel capacity" are ongoing. Nonetheless, the use of information theory in psychology has been declared fruitless, several times (e.g., Gregory, 1980; Luce, 2003; Laming, cited in Luce, 2003; Collins, 2007), yet with some puzzlement, because a compelling explanation of *why* has been lacking. This paper provides such an explanation.

Shannon's "General Communication System", Information Transmitted, and the "Confusion Matrix"

In Shannon's "general communication system", there is "An information source which produces a message or sequence of messages to be communicated to the receiving terminal"; "A transmitter which operates on the message in some way to produce a signal suitable for transmission over the channel"; "The channel is merely the medium used to transmit the signal from transmitter to receiver"; "The receiver ordinarily performs the inverse operation of that done by the transmitter, reconstructing the message from the signal"; and "The destination is the person (or thing) for whom the message is intended". Figure 1 shows the "general communication system".

Probabilities lead to information as follows (Shannon, 1948). n events are possible. The event that occurs is the outcome. When n > 1, the outcome is uncertain. What is certain is each event's probability of occurrence, p_i, i = 1, ... , n. Shannon proved that the requisite [amount of] "uncertainty", "choice", or "information", called I_S , is

$$\textit{source (signal) uncertainty (source information)} = -K \sum_{i=1}^{n} p_i \log p_i, \quad K > 0. \quad (1)$$

Shannon set K =1. When events are symbols "k",

$$I_S = -\sum_k p(k) \log p(k). \quad (2)$$

Shannon assumed, for simplicity's sake, that any symbol received is one of the set of symbols sent. Not all symbols will be received as transmitted; unintended errors occur. Let the probability of transmission of symbol k given reception of symbol j be denoted $p_j(k)$. Then

$$\textit{information transmitted } I_t = I_S - E_S = \sum_k p(k)\log p(k) + \sum_j\sum_k p_j(k)\ \log p_j(k), \quad (3)$$

$$E_S = -\sum_j\sum_k p_j(k)\ \log p_j(k) \textit{ is the stimulus equivocation/uncertainty/entropy, called } H. \quad (4)$$

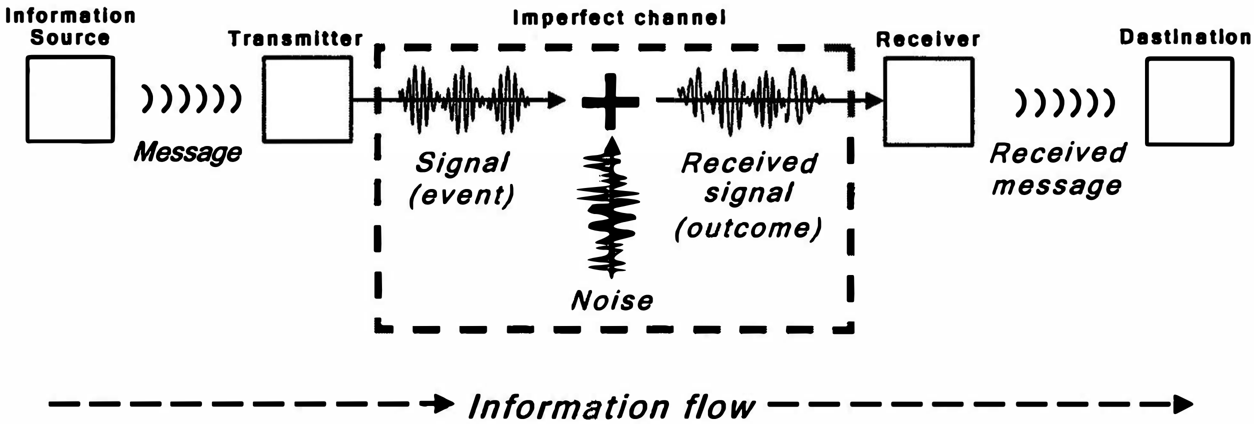

Figure 1. A "general communication system" (after Shannon, 1948), to which Shannon applied his Information Theory.

I_t can be computed by knowing (a) what symbols were transmitted, (b) what symbols were received, and (c) for each symbol sent and each symbol received, the number of times that the latter corresponded to the former. Those numbers form the basis of the confusion matrix, whose columns are labeled by the symbol transmitted (the event) and whose rows are labeled by the symbol received (the outcome), such that each matrix entry depends upon the number of times that the particular outcome corresponded to the particular event. Figure 2 shows the Shannon confusion matrix. The confusion matrix can be assembled as follows for a set of transmitted symbols "k". N_{jk} = the number of times a symbol transmitted as k is received as j; $p(j) = N_{j.}/N$ = the probability that j was received; $p(k) = N_{.k}/N$ = the probability that k was transmitted; $p_k(j) = N_{jk}/N_{.k}$ = the probability that j was received if k was transmitted; and $p_j(k) = N_{jk}/N_{j.}$ = the probability that k was transmitted, if j was received. Non-zero off-diagonal elements in the confusion matrix represent lack of transmission fidelity ("noisiness"). When what is transmitted is identically received, the non-zero entries lie on the diagonal, and $I_t = I_S$.

	Symbol sent (event)						
	1	2	…	k	…	n	Row totals
Symbol received (outcome)							
1	N_{11}	N_{12}	…	N_{1k}	…	N_{1n}	$N_{1.}$
2	N_{21}	N_{22}	…	N_{2k}	…	N_{2n}	$N_{2.}$
…	…	…	…	…	…	…	
j	N_{j1}	N_{j2}	…	N_{jk}	…	N_{jn}	$N_{j.}$
…	…	…	…	…	…	…	
n	N_{n1}	N_{n2}	…	N_{nk}	…	N_{nn}	$N_{n.}$
Column totals	$N_{.1}$	$N_{.2}$	…	$N_{.k}$	…	$N_{.n}$	Σ=N

Figure 2. Shannon's Information Theory "confusion matrix" (Shannon, 1948).

Here,

$$\textit{total number of symbols received} = \sum_{j=1}^{n} N_{j.} = \sum_{k=1}^{n} N_{.k} = \textit{total number of symbols sent} = N.$$

The Confusion Matrix in Sensory Psychology

Shannon's Information Theory has been employed in psychology to quantify a subject's performance in the "category" or "absolute identification" experiment. The computational method has been attributed to Garner and Hake (1951). A set of sensory stimuli are made to vary in only one sensory attribute. Here, as in most actual experiments, let that attribute be the intensity. The experimenter partitions the intensity continuum into ranges called "stimulus categories" and "response categories". The response categories can be named or numbered, usually rising with rise in intensity, for example, "very weak" to "very strong" or 1-10. Each stimulus from a stimulus category evokes a sensation; as intensity increases, sensation increases. Corresponding to the response categories, the subject has a set of non-overlapping ranges of sensations. In the experiment, the subject is exposed to randomly chosen intensities, presented one after the other in a block of intensities spanning the intensity range to be explored. Based on sensation, the subject states the response category that they believe corresponds to the given stimulus.

Figure 3 shows the confusion matrix of sensory psychology. The rows of the confusion matrix are labeled by response category and the columns of the matrix are labeled by stimulus category. Each matrix entry is the number of times that a stimulus taken from the respective stimulus category was classified by the subject as falling within the respective response category. According to Garner and Hake (1951, p. 452), I_t expresses "the amount of information about the event continuum which a particular range of stimulus values can transmit".

	Stimulus category						
Response category	1	2	⋯	k	⋯	n	**Row totals**
1	N_{11}	N_{12}	⋯	N_{1k}	⋯	N_{1n}	$N_{1.}$
2	N_{21}	N_{22}	⋯	N_{2k}	⋯	N_{2n}	$N_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
j	N_{j1}	N_{j2}	⋯	N_{jk}	⋯	N_{jn}	$N_{j.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	N_{n1}	N_{n2}	⋯	N_{nk}	⋯	N_{nn}	$N_{n.}$
Column totals	$N_{.1}$	$N_{.2}$	⋯	$N_{.k}$	⋯	$N_{.n}$	$\Sigma=N$

Figure 3. The "confusion matrix" of sensory psychology (after Garner & Hake, 1951).

Problems with the Confusion Matrix in Sensory Psychology: (1) Categories, and their Manipulation to Maximize Information Transmitted

In Shannon Information Theory, the "outcomes" are elements of the set of known possible "events". Thus, "events" and "outcomes" are similar things. Not so, in point of fact, for the sensory psychologist. In absolute identification, stimulus intensities are the "events", and responses to the stimuli are verbal, written, or electronically signalled "outcomes". But responses ("outcomes") and stimuli of given intensities ("events") are self-evidently different things. They do not have the same units, or even the same *dimensions* of units. Stimulus intensity has physical units, whose particular dimensions are some combination of mass, length and time; in contrast, the response to a stimulus is simply a number, having no physical units at all.

That difference may have been what motivated Garner and Hake (1951) to replace Shannon's "events" and "outcomes" by "categories". Thereby, any given stimulus was contained within a stimulus category, and any given response was contained within a response category. Categories had the advantage that they masked any arbitrariness by seeming to be similar, perhaps identical, things. For example, there could be a Category 3 of "events", and a Category 3 of "outcomes". When a stimulus within Stimulus Category 3 was "matched" by an experimental research subject to a response within Response Category 3, both stimulus and response fell within something *called* Category 3. Further, the common practice in sensory science was to make the number of stimulus and response categories the same, in which case "information transmission" for/by/of the stimulus could possibly be "perfect".

The partitioning of stimulus intensity, a continuum, into stimulus categories was always arbitrary. As Garner and Hake (1951, p. 452) put it, "Which events (or how many) are represented by which particular discrete stimulus is an arbitrary matter". That is, stimulus *ranges*, consider to be "event" (i.e., stimulus intensity) *categories*, were represented by individual stimulus intensities. Indeed, individual stimuli were often referred to as stimulus categories (e.g., Garner, 1953; Engen & Pfaffmann, 1959). Conversely, the concept of "stimulus category" was sometimes not applied; stimuli were sometimes simply called "stimuli" (e.g., Beebe-Center et al., 1955; Eriksen & Hake, 1955; Blamey & Clark, 1987; Fulgosi et al., 1987; Mori & Ward, 1995; Petrov & Anderson, 2005). From this point on, what are called "events" will be the particular stimulus intensities used in absolute identification experiments. Possible responses (e.g., "that was stimulus number 6") were not typically called response *categories.* Regardless, all papers referred to the computed "information transmitted" I_t as a source of the theoretically identifiable number of categories of stimuli; when the logarithms in Eqs. (1)-(4) are taken to base X, where X is a positive integer, the number of identifiable stimulus categories is, in principle, X^{I_t}.

Note well, however, that the magnitude of the information transmitted can be increased by changing the numbers and widths of the stimulus categories and response categories of the confusion matrix (Garner & Hake, 1951), in a manner that eliminates or reduces the magnitude of off-diagonal entries. In particular, Garner and Hake (1951) realized that the information transmitted could be maximized by making the stimuli equally discriminable, that is, just as easy to discriminate from their nearest neighbors. Equal discriminability could allegedly be established through an elaborate and time-consuming experiment that was proposed by Garner and Hake (1951), who carried it out for some pure (single-frequency) tones. Unfortunately, such preliminary experiments constitute lengthy procedures in themselves and are therefore beyond the scope of most sensory psychology "information transmission" experiments. (Some, nonetheless, took that track.) Alternatively, such experiments could be done for a broad variety of stimuli, with the derived sets of equally-discriminable stimuli being made publicly known for future use. That, however, has not happened. Regardless, equal discriminability would seem to be a capability of a human being, not of a "communication system".

Indeed, if the magnitude of "information transmitted" can easily be altered by simply manipulating the spacing between the stimuli, then only its maximum seems to have any meaning. Recall from Eq. (3) that actual "information transmitted" depends partly upon the *a priori* probabilities of stimulus presentation, $p(k)$, set by the experimenter. But those alone are not the only stimulus properties that affect "information transmitted" in absolute identification; as noted, the latter is empirically dependent upon stimulus *configuration*, which perforce must alter the *a posteriori* probabilities $p_j(k)$ in Eq. (3). But alterations to $p_j(k)$ imply alterations to the "noise" of the Shannon "communication system". In Shannon's system, noise was imagined as a random, independent feature of the channel, not as a property of the "events"; that is, the "events" are used to test the channel for its information capacity, rather than the channel being used to test the "events" for discriminability through the Garner-Hake equal-discriminability protocol. Stimulus intensity discriminability is presumably a memory-dependent cognitive feature, not a "channel" property *per se*, and as such, less-than-optimal discriminability constitutes "false noise".

Problems with the Confusion Matrix in Sensory Psychology: (2) the Complicated Relation of Outcome to Event

Even if the "false noise" could be eliminated by choosing equally-discriminable stimuli and their appropriate response categories, thus maximizing "information transmitted", there is a further source of "false noise", and one that arises from categorization by the research subject. That is, by glossing over the differences in units between "events" and "outcomes", categorization serves a deeper purpose, namely, to downplay the ultimate *source* of that difference in units: that the experimental research subject's categorizations (the "outcomes") relate to the stimulus intensity "events" in a decidedly nonlinear manner. That is, the transformation of "event" to "outcome" is more complicated than a simple straight-line function $y = ax + b$. Consider that the relation of stimulus intensity to category judgment may comprise not one but two successive nonlinearities. The first nonlinear transformation occurs when a stimulus evokes a sensation. The mechanism is still poorly understood, but any contemporary textbook or review paper assumes it to be nonlinear. After all, sensation must depend upon the behavior of the first neurons to respond to a stimulus, the primary sensory afferents, and their stimulus-driven spike firing-rate is a nonlinear function of stimulus intensity.

But there is an even further step beyond sensation-generation. Recall that all of the possible stimuli in an absolute identification experiment are *not* presented for the subject's review every time that the subject is obliged to make a categorization. Such comparisons are possible only during training sessions. Subsequent to the training sessions (which can be brief), the subjects categorize stimuli by comparing their immediate *memory* of the stimulus-evoked sensation to the *memories* of the sensations that were learned during practice.

At this point it proves helpful to adopt some mathematical notation, with the understanding that equations only represent mean values. Let us assume that the stimulus I evokes a mean sensation S according to a relation that is meaningful, in that it is smooth, continuous, and monotonic, call it $S(I)$. Despite the potentially chaotic effects of memory upon the transformation of sensation to judged category, let us assume for simplicity's sake that, on average, the subject assigns $S(I)$ a category judgment (category number) f according to a relation that is meaningful, in that it is smooth, continuous, and monotonic, $f(S)$. Although categories are discrete, it is convenient mathematically to represent them (for the present) as being infinitely thin and hence continuous. By assigning $f(S)$ to S, the subject indirectly relates the assigned category to the stimulus intensity according to a meaningful relation $f(S(I))$, which is smooth, continuous, and monotonic, call it $F(I)$. Figure 4 shows the two-stage process in absolute identification.

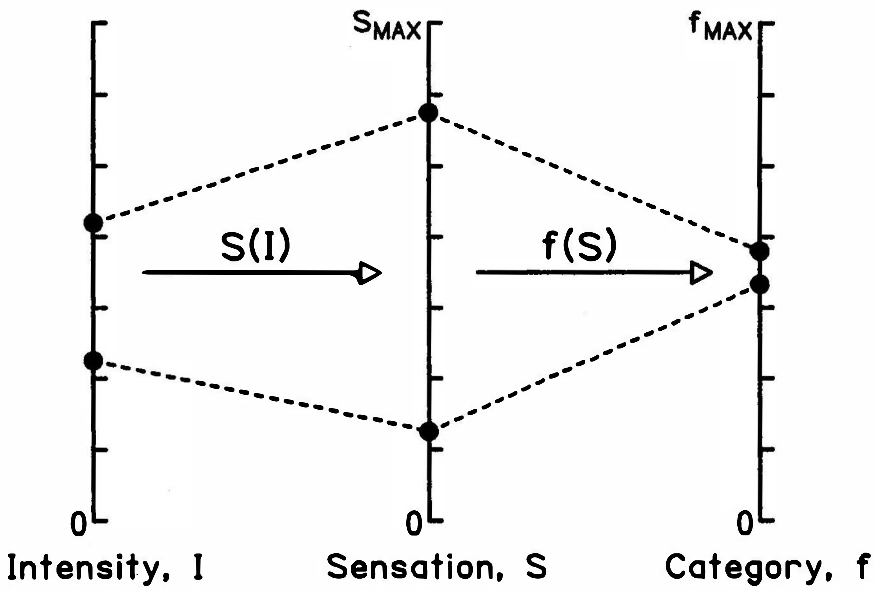

Figure 4. A graphic depiction of two examples of categorization of stimulus intensity. Two intensities I (the black dots on the leftmost vertical number line) are transformed into two respective sensations $S(I)$ (the black dots on the central vertical number line), which in turn are transformed into two respective categorizations $f(S)$ (the black dots on the rightmost vertical number line). The relations $S(I)$ and $f(S)$ are assumed to be smooth, continuous, and monotonic (see text). Therefore, the higher I gives the higher $S(I)$, which gives the higher $f(S)$. However, the transformations are nonlinear, as indicated by the difference in slope between the pair of dashed lines that link I to S, and by the difference in slope between the pair of dashed lines that link S to f. We may assume that $S(I=0)=0$ (i.e., absence of stimulus produces absence of sensation) and that, likewise, $f(S=0)=0$ (absence of sensation results in no assigned category, as represented here by a convention, "zero" category). Sensation is assumed to have an effective maximum, S_{MAX}, for which there is a maximum categorization, f_{MAX}.

Returning to the role of memory: there is considerable accumulated circumstantial evidence that "information transmitted" obtained in absolute identification experiments is a measure of short-term memory for random letters or digits, unless the involvement of memory is lessened or removed, in which case "information transmitted" *in absolute identification* increases perhaps without bound (Nizami, 2010). The latter critical review was later condensed to form the background and justification of what is perhaps the first comprehensive and readable original model of how memory actually influences categorizations (Nizami, 2011).[1] Nizami's conclusion (see Nizami, 2010, 2011) was that memory does not merely *limit* "information transmitted" in absolute identification, as some had suggested; rather, memory capacity is what is actually measured in absolute identification. "Memory noise", a form of "false noise", utterly obscures any role of actual *sensory* noise.

Thus, the results from absolute identification experiments cannot be meaningfully interpreted using Shannon Information Theory, unless the obscuring effect of "false noise" is stripped away. The obvious way to do that is to convert category judgments back to outcomes-as-events. But sensory psychologists do not do that. Regardless, an equation relating categories back to outcomes-as-events could conceivably be devised if we knew the transformations by which a research subject turns stimulus intensities into statements of response category. Whether it is possible to achieve that knowledge is the subject of the following sections.

Problems with the Confusion Matrix in Sensory Psychology: (3) Outcome (Categorization) is Stochastic

There is an important related point, one that should not be further delayed, namely, that categorization is *stochastic*. First of all, the response of a primary afferent neuron is distributed (references too numerous to mention). That is, the spike firing rate can be represented by a probability density function (a "distribution") having a mean value and a variance. As such, the sensations evoked by repetitions of a fixed stimulus, whether dependent upon a single neuron or an ensemble of neurons, will inevitably be distributed. Figure 5 shows this point. This variability of sensation is the *sensory* system's true noisiness, that is, its true tendency to error. However, the subsequent categorization of any given single sensation is memory-dependent and hence imperfect, introducing a new level of variability. Altogether, the judged category to which a fixed sensation is assigned is necessarily distributed. Figure 6 shows this point. Overall, then, a stimulus intensity "event" evokes two steps of randomization by the time the human subject announces the category to which the stimulus intensity belongs. And empirically, category judgments are indeed distributed (e.g., van Krevelen, 1951; Terman, 1965; Miyazaki, 1988; McCormack et al., 2002; Elvevag et al., 2004; Neath et al., 2006; Murphy et al., 2010). Hence, it helps to use mean values of sensations and of categorizations, with the understanding that each mean value reflects an underlying distribution of values.

Problems with the Confusion Matrix in Sensory Psychology: (4) Two Steps Link Event to Outcome, and Both are Unknown

And here we arrive at an apparently intractable problem, as follows. Sensory psychologists ignored the possibility that there was an intervening step between sensation generation and categorization. That is, they imagined that "information transmission" involved only the generation of the sensation from the stimulus, not the subsequent memory-dependent generation of the category judgment from sensation. Given functions $S(I)$ and $f(S)$ which are smooth, continuous, and monotonic, the mean categorization as a function of stimulus intensity, $F(I)$, has an inverse $F^{-1}(I)$, which is such that $I = F^{-1}(F(I)) = F(F^{-1}(I))$. (Imagine, for example, that F is "squared" and F^{-1} is "square root", or vice versa.) Similarly, the equation $f(S)$ has an inverse $f^{-1}(S)$, such that $S = f^{-1}(f(S)) = f(f^{-1}(S))$; the equation $S(I)$ has an inverse $S^{-1}(I)$, such that $I = S^{-1}(S(I)) = S(S^{-1}(I))$. But, also, $I = F^{-1}(F(I)) = F(F^{-1}(I))$. Counterintuitively, the latter stimulus intensity is not the true mean outcome-as-event, as follows.

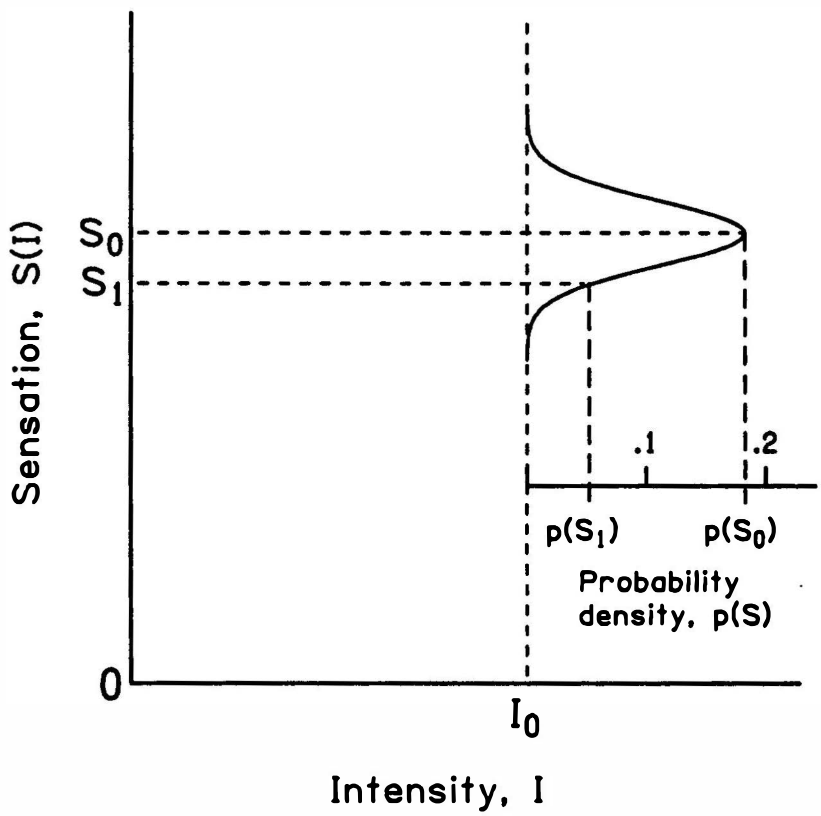

Figure 5. The stochastic nature of sensation. A stimulus of intensity I_0, over repeated presentations, evokes a distribution of sensations (here assumed Gaussian, in the absence of data to the contrary) having a mean value S_0. S_0 occurs with a probability density $p(S_0)$, and, likewise, another sensation S_1 occurs with a probability density of $p(S_1)$ in response to other presentations of I_0. Probability density should be thought of as an axis that extends perpendicularly out of the page.

Suppose that we know $F(I)$. Consider two intensities, I and I'. Then given $F(I)=f(S(I))$, and if f is not the *Identity* transformation $f(S)=S$, then the declaration $F(I')=S(I)$ implies that $I \neq I'$, that is, that the stimulus intensity "event" I' corresponding to $F(I')$ is not the stimulus intensity "event" I corresponding to $S(I)$. To see this, first assume that $F(I')=S(I)$. Consequently, $f(S(I'))=S(I)$, from which $f^{-1}(f(S(I')))=f^{-1}(S(I))$, that is, $S(I')=f^{-1}(S(I))$. Consequently, $S^{-1}(S(I'))=S^{-1}(f^{-1}(S(I)))$. Suppose now that $I=I'$; then $S^{-1}(S(I))=S^{-1}(f^{-1}(S(I)))$, which is true if and only if $f^{-1}(S(I))=S(I)$, that is, if $f^{-1}=Identity$, in which case also $f=Identity$ in order that $f(f^{-1}(S(I)))=f^{-1}(f(S(I)))=S(I)$. Conversely, for $f \neq Identity$, we have $I \neq I'$. Figure 7 shows a graphic version of the argument.

In sum, the mean outcome-as-event that corresponds to a given mean category judgment cannot be known from only $F(I)$, the categorization as a function of intensity. What is required is $S(I)$, the relation of sensation to intensity, and also $f(S)$, the transformation of sensation to category judgment. The proof above is a variation on one by Phillips (1964). Phillips' proof actually concerned $F(I)$ as loudness, a judgment of sensation magnitude, rather than category number; and indeed, Phillips (1964) did not discuss memory or categorization, because he was not concerned with absolute identification. The latter is not the same thing as judged loudness. Nonetheless, Phillips' (1964) proof is relevant, and it proceeds as follows. Phillips noted that $I=S^{-1}(S(I))=S(S^{-1}(I))$, hence $F(I)=F(S^{-1}(S(I)))=f(S(I))$. Phillips emphasized that, if we know $F(I)$, but if we have no reason to know $S^{-1}(I)$ (and indeed we do not), then we cannot know $S=S(I)$ without knowing $f(S)$. In other words, *no relation of intensities to numbers which results from two unknown successive stages can reveal those stages.* To give credit where credit is due, Phillips' proof deals with concepts that were actually pursued earlier (using similar algebra) by Attneave (1962), whom Phillips (1964) does not cite. Regardless, an $f(S)$ for either absolute identifications or magnitude judgments seems unlikely to be discovered any time soon.

Even if $S(I)$ and $f(S)$ were both known, we could never know *exactly* the stimulus intensity I (the outcome-as-event) that corresponded after-the-fact to a known single categorization $F(I)$, due to the stochastic nature of absolute identification. Only mean values could be specified; the closest we might get to obtaining outcomes-as-events might be to simulate the variability inherent in $S(I)$ and in $F(I)$ by using Monte Carlo simulations, as done by physicists. But that would require knowing how $S(I)$ and $f(S)$ are distributed.[2]

Finally, note an interesting property that is illustrated in Fig. 7, namely, that the lowest and highest sensations naturally correspond to the lowest and highest category judgments. But such numbers are all mean values. With mean values of sensation/categorization which are at the extremes of sensation/categorization, the distributions of sensations and of categorizations must both, within this model, become increasingly skewed towards lower sensations/categorizations as sensation/categorization approaches zero, and toward higher sensations/categorizations as sensation/categorization approaches maximum. Such skewing was predicted by Nizami (2011) for categorization, and is found experimentally for categorizations of the duration (McCormack et al., 2002; Murphy et al., 2010) and for the pitch (McCormack et al., 2002) of pure (i.e. single-frequency) tones, and for the lengths of rods (Neath et al., 2006) and for tone duration and for line length (Elvevag et al., 2004).[3]

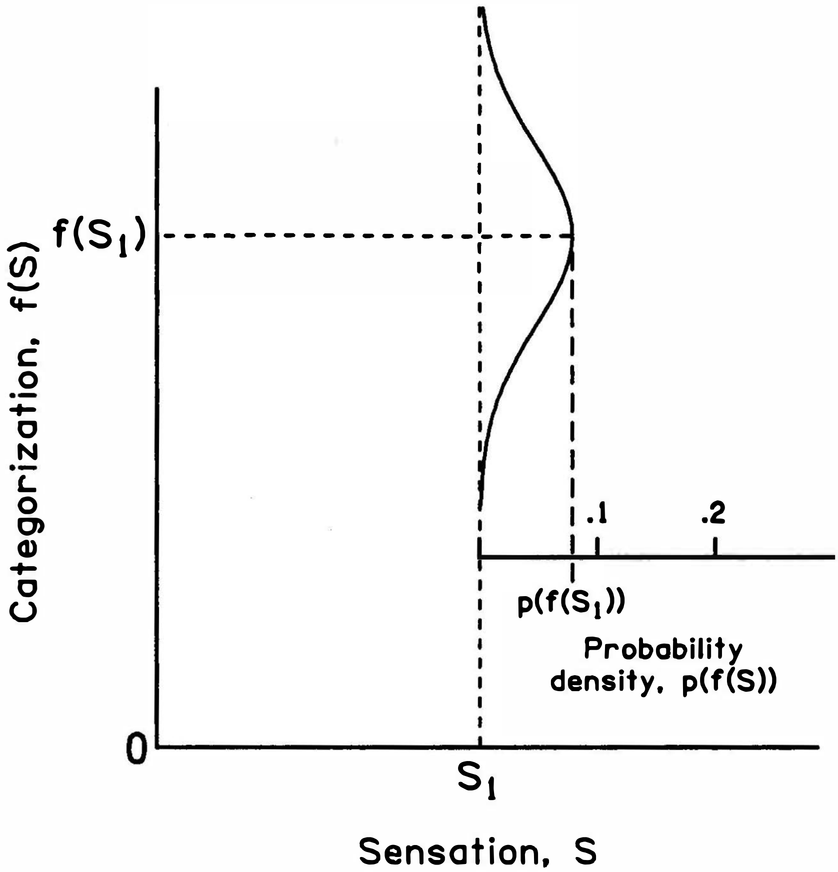

Figure 6. The stochastic nature of categorization. The sensation S_1 seen in Fig. 5, over repeated occurrences, evokes a distribution of categorizations (here assumed Gaussian, in the absence of data to the contrary) having a mean value $f(S_1)$ which occurs with a probability density $p(f(S_1))$. Probability density should be thought of as an axis that extends perpendicularly out of the page. The pictured Gaussian distribution is broader than that of Fig. 5, to suggest that variability in memory greatly exceeds variability in sensation (see text) in absolute identifications.

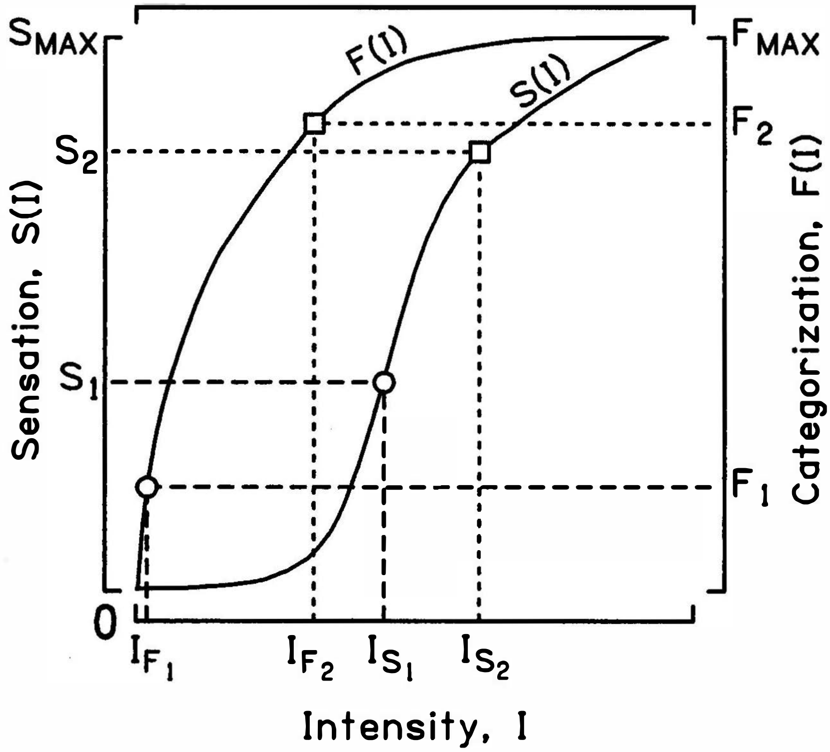

Figure 7. A graphical version of the proof (see text) that given $F(I)=f(S(I))$, and if $f(S)\neq S$, then the declaration $F(I')=S(I)$ implies that $I\neq I'$, that is, that the stimulus intensity "event" I' of $F(I')$ is not the stimulus intensity "event" I of $S(I)$. All functions here represent mean values, and for the sake of illustration, $f(S)$ is assumed to be nonlinear. We assume that $S(I=0)=0$ and $f(S=0)=0$ (see Fig. 4), hence $F(I=0)=0$. We also assume that $f(S_{MAX})=f_{MAX}=F_{MAX}$. The mean sensation S_1 evokes the mean categorization F_1; due to $f(S)\neq S$, the dashed horizontal line touching S_1 will not intercept the F-axis at F_1, hence the respective intensities I_{S_1} and I_{F_1} differ (follow the long dashed lines). The same sort of argument applies for the mean sensation S_2 that evokes the mean categorization F_2, and the respective intensities I_{S_2} and I_{F_2} (follow the short dashed lines).

Overall Summary

Shannon's Information Theory offered an estimate of "information transmitted" from the data cumulated in the "confusion matrix". Originally, the columns of the matrix were labeled by symbol sent ("event") and the rows by symbol received ("outcome"). However, sensory psychologists broke from this tradition, instead labeling the matrix's columns by stimulus category and its rows by response category. The usefulness of the psychologists' approach has been doubted by theorists. The present paper reveals that those doubts were justified, not only because the psychologists' approach uses arbitrarily defined "events" and "outcomes", but also because its "event" and "outcome" continua are arbitrarily partitioned.

Shannon's "symbols received" (Fig. 2) are elements of the set of "symbols sent". But category judgments, used by sensory psychologists as "outcomes", are nonlinear transformations of stimulus intensity "events" and, as such, are not elements of Shannon's "general communication system". In principle, this incongruity might be compensated for by algebraic conversion of categorizations back into outcomes-as-events – that is, back to stimulus intensities. But such a conversion has not been provided by sensory psychologists, which, in retrospect, is no surprise; categorization should be recognized as a two-step process involving transformation of intensity to sensation, followed by transformation of sensation to a category judgment. Both steps are stochastic. It is presently impossible to quantify the mean sensation with intensity, or the relation linking mean sensation to mean category judgment. Hence, it is not possible to work backwards from a mean category judgment to a mean sensation to a stimulus intensity. And indeed, thanks to stochasticity, it is *always* impossible to work backwards from a particular category judgment to a particular sensation to a particular stimulus intensity. Altogether, categorizations of stimulus intensity, made by human subjects cannot be reduced to elements of the set of possible "events" (stimulus intensities), and therefore cannot be used to compute "information transmitted" using Shannon Information Theory. Without that correction, we are left with "false noise" which reduces the possible computed magnitude of "information transmitted".

Final Analysis

The mistakes just mentioned all have a common root cause. That root cause is the misapplication of Shannon Information Theory to things that it was not designed for, as follows.

To compute Shannon "information transmitted" requires an *observer* (sometimes called "the engineer"). The observer must perforce know what the "events" are, and what their *a priori* probabilities of occurrence are. The observer records the actual frequency of occurrence of the outcomes (as a subset of the events). Indeed, it is observers-as-engineers who construct the "general communication system" itself.

Remarkably, sensory psychologists have never acknowledged the role of the observer. Shannon himself may be partly to blame, because the observer did not appear in Shannon's (1948) original illustration of his system, and hence not in Fig. 1 here. Regardless, sensory psychologists assumed that what they were measuring, in applying Shannon Information Theory to absolute identification, was the neural noise of a sensory system. Seemingly, they did not think that they were measuring aspects of a more complicated system, one that generates human behavior under other kinds of noise. Indeed, the sensory psychologists treated sensory response as if it was isolated from all other neural activity within the human subject, hence examinable independently of the sensory system's role as part of the whole human, the whole human who was always a potential *observer*.

That is, sensory psychologists, themselves observers, tried to quantify the noisiness of a "system" that was part of a larger system, the potential observer. By indicating their own responses, rather than having them recorded in an unbiased manner by a third party, the experimental subjects themselves introduced a bias, in fact, a memory bias. That bias occurred because subjects could not accurately indicate their own sensations, or work backwards from those to indicate the evoking stimuli (even if sensation were not stochastic, or if the stochasticity were ignorable). In contrast, in Shannon Information Theory the observer is utterly apart from the "general communication system", thus being unbiased, a necessary attribute.

In short, it may have been sensible to apply Shannon Information Theory to things made by humans, such as transmission lines, but not to humans themselves.

Acknowledgements

Study supported by the author. I thank Prof. Claire S. Barnes PhD (VA Palo Alto HCS) for her insightful comments.

Notes

1. Earlier "reviews" of the results of the Garner-Hake method in psychology (e.g., Broadbent, 1975; Baddeley, 1994; Shiffrin & Nosofsky, 1994) were too brief, and focused on the respective authors' own connections to the field rather than on the evidence in general.

2. There are many researchers who believe that $S(I)$ is already well-known. Indeed, those researchers believe that $f(S)=S$ and that $S(I)$ follows Stevens' Law, named after Professor S.S. Stevens of Harvard University. Stevens' Law was inferred from magnitude judgments, which are self-quantifications of sensations. Stevens treated magnitude judgments as "reflexes" (e.g., Stevens, 1959), i.e., $f = Identity$. However, the latter is unlikely, because humans cannot accurately self-quantify their sensations. And regardless, Stevens' Law lacks credibility on an enormous number of other grounds, for which the supporting references are far too numerous to mention.

3. Nizami (2011) did not consider the skewing of the distribution of sensation, but perhaps should have.

References

Attneave, F. (1962); Perception and Related Areas; Psychology: a Study of a Science, Vol. 4, Biologically Oriented Fields: Their Place in Psychology and in Biological Science (ed. S. Koch), McGraw-Hill, New York, NY (pp. 619-659).

Baddeley, A. (1994); The Magical Number Seven: Still Magic After All These Years?; Psychological Review, Vol. 101, No. 2 (pp. 353-356).

Beebe-Center, J.G., Rogers, M.S. and D.N. O'Connell (1955); Transmission of Information about Sucrose and Saline Solutions through the Sense of Taste; Journal of Psychology, Vol. 39, No. 1 (pp. 157-160).

Blamey, P.J. and G.M. Clark (1987); Psychophysical Studies Relevant to the Design of a Digital Electrotactile Speech Processor; Journal of the Acoustical Society of America, Vol. 82, No. 1 (pp. 116-125).

Broadbent, D.E. (1975); The Magic Number Seven After Fifteen Years; Studies in Long-Term Memory (eds. A. Kennedy and A. Wilkes), Wiley, New York (pp. 3-18).

Collins, A. (2007); From H=*log* s^n to Conceptual Framework: a Short History of Information; History of Psychology, Vol. 10, No. 1 (pp. 44-72).

Elvevag, B., Brown, G.D.A., McCormack, T., Vousden, J.I. and T.E. Goldberg (2004); Identification of Tone Duration, Line Length, and Letter Position: an Experimental Approach to Timing and Working Memory Deficits in Schizophrenia; Journal of Abnormal Psychology, Vol. 113, No. 4 (pp. 509-521).

Engen, T. and C. Pfaffmann (1959); Absolute Judgments of Odor Intensity; Journal of Experimental Psychology, Vol. 58, No. 1 (pp. 23-26).

Eriksen, C.W. and H.W. Hake (1955); Multidimensional Stimulus Differences and Accuracy of Discrimination; Journal of Experimental Psychology, Vol. 50, No. 3 (pp. 153-160).

Fulgosi, A., Bezinovic, P., Mimica, I. and V. Taksic (1987); Channel Capacity in the Magnitude Estimation of Weight; Perception and Psychophysics, Vol. 42, No. 4 (pp. 328-332).

Garner, W.R. (1953); An Informational Analysis of Absolute Judgments of Loudness; Journal of Experimental Psychology; Vol. 46, No. 5 (pp. 373-380).

Garner, W.R. and H.W. Hake (1951); The Amount of Information in Absolute Judgments; Psychological Review, Vol. 58, No. 6 (pp. 445-459).

Gregory, R.L. (1980); Whatever Happened to Information Theory? (2); Perception, Vol. 9, No. 5 (pp. 489-491).

Luce, R.D. (2003); Whatever Happened to Information Theory in Psychology?; Reviews of General Psychology, Vol. 7, No. 2 (pp. 183-188).

McCormack, T., Brown, G.D.A., Maylor, E.A., Richardson, L.B.N. and R.J. Darby (2002); Effects of Aging on Absolute Identification of Duration; Psychology and Aging, Vol. 17, No. 3 (pp. 363-378).

Miller, G.A. (1956); The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information; Psychological Review, Vol. 63, No. 2 (pp. 81-97).

Miyazaki, K. (1988); Musical Pitch Identification by Absolute Pitch Possessors; Perception and Psychophysics, Vol. 44, No. 6 (pp. 501-512).

Mori, S. and L.M. Ward (1995); Pure Feedback Effects in Absolute Identification; Perception and Psychophysics, Vol. 57, No. 7 (pp. 1065-1079).

Murphy, D.R., Schneider, B.A. and H. Bailey (2010); The Effects of Age on Channel Capacity for Absolute Identification of Tonal Duration; Attention, Perception and Psychophysics, Vol. 72, No. 3 (pp. 788-805).

Neath, I., Brown, G.D.A., McCormack, T., Chater, N. and R. Freeman (2006); Distinctiveness Models of Memory and Absolute Identification: Evidence for Local, not Global, Effects; Quarterly Journal of Experimental Psychology, Vol. 59, No. 1 (pp. 121-135).

Nizami, L. (2010); Interpretation of Absolute Judgments using Information Theory: Channel Capacity or Memory Capacity?; Cybernetics and Human Knowing, Vol. 17, Nos. 1-2 (pp. 111-155).

Nizami, L. (2011); Memory Model of Information Transmitted in Absolute Judgment; Kybernetes, Vol. 40, Nos. 1-2 (pp. 80-109).

Petrov, A.A. and J.R. Anderson (2005); The Dynamics of Scaling: a Memory-Based Anchor Model of Category Rating and Absolute Identification; Psychological Review, Vol. 112, No. 2 (pp. 383-416).

Phillips, J.P.N. (1964); A Note on Treisman's Model; Quarterly Journal of Experimental Psychology, Vol. 16, No. 4 (p. 386).

Shannon, C.E. (1948); A Mathematical Theory of Communication; Bell System Technical Journal, Vol. 27, No. 1 (pp. 623-656).

Shiffrin, R.M. and R.M. Nosofsky (1994); Seven Plus or Minus Two: a Commentary on Capacity Limitations; Psychological Review, Vol. 101, No. 2 (pp. 357-361).

Stevens, S.S. (1959); The Quantification of Sensation; Daedalus, Vol. 88, No. 4 (pp. 606-621).

Terman, M. (1965); Improvement of Absolute Pitch Naming; Psychonomic Science, Vol. 3, No. 6 (pp. 243-244).

van Krevelen, A. (1951); The Ability to Make Absolute Judgments of Pitch; Journal of Experimental Psychology; Vol. 42, No. 3 (pp. 207-215).