

Signalling systems for individual recognition: an information theory approach

MICHAEL D. BEECHER

Department of Psychology, University of Washington, Seattle, Washington 98195, U.S.A.

Abstract. This paper develops a quantitative method for measuring the information capacity of an animal's 'signature system', i.e. the set of cues by which individuals are identified. The information measure (H_s) is derived by applying Shannon's measure for the information in a continuous variable to a simple linear model. The model is essentially the analysis of variance model II (random effects), and is implicit in the many ANOVAs and discriminant function analyses that have been done on the signature systems of animals. For multivariate measurements, a principal components transformation of the data permits the information in the independent components to be added to give the total information. An analysis of illustrative data sets reveals a close correlation between H_s and the probability of a correct classification of an individual (P) obtained by discriminant function analysis. H_s has the advantage, however, that it is a population estimate whereas the value of P is tied to the number of individuals in the sample. The information analysis approach may prove valuable for comparative analyses where evolutionary hypotheses predict one species to have a better developed signature system than another.

There has been considerable interest in individual recognition in recent years (see, e.g. reviews in Falls 1982; Colgan 1983). To date, however, most studies have not gone beyond suggesting that individual recognition is possible because sufficient variation exists in presumptive cues such as calls or visual markings, or showing that recognition occurs, via cross-fostering, playback or other type of experiment. While comparisons have sometimes been made between one species and another, they have typically been confined to present-absent comparisons, as in the well-known generalization that parent-offspring recognition occurs in herring gulls but not in kittiwakes (Cullen 1957). What has been lacking in these studies is quantitative description of the recognition system. Quantitative description would be valuable because evolutionary logic dictates that natural selection will act to differing degrees on recognition systems. For example, in a colonial species such as the Mexican free-tailed bat, *Tadarida brasiliensis mexicana*, in which parents must find their offspring among hundreds of young of similar age (McCracken & Gustin 1987), we would expect to find a more highly developed system than in a less colonial species in which parents do not face a recognition problem of such magnitude (Beecher 1982; Jouventin 1982; Colgan 1983).

In this paper, I develop a method for analysing the signals by which animals are recognized. The

method is based on information theory (Shannon & Weaver 1949) and it has the following features. (1) It quantitatively describes the signal (or signature) system. (2) It has inherent meaning in the recognition context. For example, it is directly translatable into the size of the group in which a particular individual could be recognized with a given degree of accuracy. (3) As is true of information measures generally, it allows 'apples and oranges' comparisons. Thus we can make comparisons across species and across recognition cue modalities. This is the key characteristic, for it makes possible a truly comparative approach. A preliminary version of this analysis has been presented in Beecher (1982).

Some recent studies have used discriminant function analysis to quantify the extent to which individuals can be classified on the basis of signal measurements (e.g. Hafner et al. 1979; Smith et al. 1982; Gelfand & McCracken 1986). Although the discriminant function technique can give an overall measure of classification success, this measure has no general meaning, being tied to the sample size of the data set. The discriminant function analysis is logically very similar to the information analysis described in this paper, however, and I discuss their relationship below.

The model developed in this paper applies to signatures that are multivariate in nature, i.e. consist of several, intercorrelated variables, and

that vary within individuals. I will first develop the argument, however, in the simpler context of discrete, unitary signatures that are invariant within individuals.

THEORY

Biological Context and General Perspective

Consider the imposing recognition problem found in the 'maternity caves' of Mexican free-tailed bats (McCracken 1984; Gelfand & McCracken 1986; McCracken & Gustin 1987). A mother leaves her pup in a mass of similarly aged young ('creche') and returns twice a day to nurse. As she searches for her offspring, she encounters many unrelated pups which will attempt to nurse from her. Although the caves contain 1–20 million bats, the magnitude of her recognition problem is reduced by the pup's fidelity to a relatively circumscribed area. Once the mother has homed to this limited area, she must still screen some 1500 pups on average, according to estimates of McCracken & Gustin (1987). Thus, in the absence of signature cues, the chance of a mother finding her pup would be approximately 1/1500. Our general prediction is that in species such as free-tailed bats, signature systems will have evolved to facilitate recognition. Observational and experimental studies of species with strong selection for recognition have generally revealed recognition based on signature cues (in the case of free-tailed bats, olfactory and acoustic; see reviews in Falls 1982; Colgan 1983). These studies do not permit us, however, to evaluate the relative contributions of signature, perceptual and behavioural adaptations to the recognition process. As part of an effort to dissect out the specific actions of selection in the evolution of recognition systems, I developed the model to be described in this paper. Its purpose is to quantify the extent to which a signature system reliably identifies individuals within a recognition group such as a creche.

Recognition will be treated here as a communication problem. The sender provides cues, 'signature' cues, which identify it, uniquely in the ideal case. Although senders may not always be favoured to identify themselves (see Beecher 1988; Beecher & Stoddard, in press), this paper considers only the general case where reliable identification is favoured. The receiver processes these signature cues, presumably comparing them to some expect-

ation, and behaves in accordance with some decision rule, either accepting or rejecting the individual as its mate, offspring or whatever. Selection could act on such a recognition system in three general ways: (1) by elaborating the signature cues, (2) by elaborating the sensory-perceptual system, and/or (3) by modifying the decision rules and behaviours by which recognition is expressed. This paper focuses on the first type of adaptation, and the information measure derived herein describes only the signature system. I will consider the implications of the sensory-perceptual system for this analysis in the Discussion.

The communication perspective just outlined contains rather specific meanings for several terms that are sometimes used interchangeably. The key distinction is between the process of the sender signalling its identity ('identification'), and the process of the receiver extracting information about identity ('discrimination' or 'recognition'). That is, I use 'identification' as does a guard requesting an unknown individual to identify himself. I define 'recognition' as discrimination between individuals or classes of individuals based on signature information. That is, I use 'recognition' in the conventional, operational sense, and 'not as a theoretical term for some process independent of stimulation and subsequent response' (Colgan 1983, page 2). Recognition varies from simple discrimination of one or a few individuals (e.g. offspring) from all other individuals, to discrimination of each individual in the group from every other individual: I reserve 'individual recognition' for the latter extreme.

The distinction between identification (focus on senders) and discrimination/recognition (focus on receivers) is critical for this paper, as the method described herein applies only to identification systems. On the other hand, this focus on the identification system means that the distinction between simple discrimination (i.e. one individual discriminated from all the rest) and true individual recognition is not critical for this paper, for however different these tasks may be for the receiver, they impose the same minimal requirements on an identification (signature) system. For example, a particular mother looking for her offspring in the creche needs only to discriminate her offspring from all other young; she need not discriminate among unrelated young. From the recognition perspective this discrimination of one versus many is certainly simpler than true indi-

vidual recognition. From the perspective of the identification system, however, every mother with an offspring in the creche must make her own particular discrimination of one versus many. Thus, the requirement for the signature system that any individual in the group be distinguishable from all others is equivalent to the requirement that each individual in the group be distinguishable from every other: both requirements could be met, minimally, by N distinct signatures for N individuals.

As a final note, this discussion of the requirements of the signature system 'as a whole' is not intended to imply group selection. If individuals benefit by having distinctive signatures, natural selection should give us a signature 'system' which, when viewed as a whole, appears to have 'solved the requirement of providing distinctive signatures for all individuals'. As the method described in this paper is essentially independent of these theoretical considerations, I refer the reader elsewhere for further discussion of them (Beecher 1982, 1988; Beecher & Stoddard, in press).

Discrete Signature Model

Since identification and recognition are inherently quantitative concepts, they can be readily analysed from the perspective of information theory. The application is particularly straightforward in the discrete signature model considered first. For a general treatment of information theory see Shannon & Weaver (1949), Quastler (1958) or Attneave (1959). The application of information theory to animal communication is well described in Wilson (1975), Hailman (1977) and Losey (1978).

The information quantity examined in this paper is the information capacity of the signature system, by which I mean its ability to identify individuals uniquely, expressed in terms of how many individuals it can identify under certain fixed assumptions about the receiver, error levels and so forth. Our goal is to be able to compare the signature systems of different species or populations, or even different signature systems within one group of animals (e.g. the olfactory and acoustic signature systems of Mexican free-tailed bats). To make these comparisons, we need a method for estimating the information capacity of a signature system from measurements in real populations. At the outset, I should emphasize that

the information measure describes only the signature system, and in no way implies that this amount of information is actually extracted by any particular receiver. Indeed, it is highly unlikely that any receiver extracts all of the information in a signature system, since the receiver generally is interested in only a small portion of it (e.g. in whether the signaller is its offspring or not).

In the discussion that follows, I simplify by assuming recognition is purely one-way (e.g. parents searching for offspring, with offspring indifferent as to who feeds them), and that each receiver has a single target individual within the group (as in a free-tailed bat creche, where each mother has a single offspring). Both two-way recognition and multiple target individuals could be added to the model without affecting the general argument (for a discussion of the complications of reciprocal parent-offspring recognition, see Beecher et al. 1985; Beecher 1988).

I begin by characterizing the recognition group in terms of its effective size, N , which is the number of individuals that, on average, are equally likely to be confused with the target individual. The effective size of the group will inevitably be smaller than the actual size of the group (creche, troop, colony, etc.). Consider a parent searching for its young in a creche. I suppose that the parent first applies a 'preliminary screen' using non-signature cues. For example, the parent goes to a location where its young is likely to be, rejects individuals that are much younger or much older than its offspring, and so forth. When all the non-signature evidence has been exhausted, the parent is left with N individuals, one of whom is the target individual. In practice, N can be estimated at least crudely from careful observational studies (e.g. the 1500 estimate for Mexican free-tailed bat creches given above).

I next suppose that each individual in the recognition group is identified by a signature, not necessarily unique. The signature set is conceived as existing independently of the particular individuals in the recognition group. That is, it is the set of the S possible signatures, each with its associated relative frequency. Attention is restricted to the case where $S > N$, for the following reason. It is unlikely that there is a biologically realistic mechanism of signature determination (such as a simple genetic mechanism) which could guarantee unique signatures (contrasted with, say, the mechanism of assigning jersey numbers to members of an athletic team). Therefore, if individuals 'select' their signatures independently, S would have to be con-

siderably larger than N or signature duplications within the group would be common.

I define three information measures: (1) H_0 , the inherent uncertainty as to identity within the group of N individuals; (2) H_G , the potential signature information present within the set of M signatures observed within a particular group of N individuals; and (3) H_S , the potential signature information present in the entire set of S signatures ($M \leq N < S$).

H_0 , the initial or inherent uncertainty as to identity, is defined purely in terms of the number of individuals in the recognition group, i.e. the number of individuals requiring identification

$$H_0 = \log N \quad (1)$$

where the log here and throughout is to the base 2 and is measured in bits (here bits/individual). H_0 is the minimum number of binary decisions the recognizer would need to narrow the search down to the target individual (assuming that all individuals are uniquely identified).

Considering either the larger pool of S signatures or the smaller set of M observed signatures, the information value of a given signature in a set is

$$h_i = -\log p_i$$

where p_i is the probability of the i th signature within the set. Here, h_i is the minimum number of binary decisions the recognizer would need to narrow the search down to the i th signature. The lower case h indicates that our information measure pertains to a single signal (signature), and not the entire signal set. Some authors refer to the information value of a particular signal as its 'surprisal', and reserve the term 'information' for the full signal set (see Attneave 1959, page 6; Hailman 1977, page 32).

The average information value of a signature is then the sum of the h_i , weighted by their relative frequencies of occurrence, or

$$H_S = -\sum^S p_i \log p_i \quad (2)$$

bits/signal, if we are evaluating the entire signature set or

$$H_G = -\sum^M p_i \log p_i$$

bits/signal if we are evaluating only the M signatures of the N individuals in a particular group. Note that where signatures within the full set are equiprobable, $H_S = \log S$ and when $M = N$, $H_G = H_0$. Figure 1 provides a simple example illustrating calculation of H_0 , H_S and H_G .

(A) Recognition group ($N=16$ individuals)	
01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16	$H_0 = 4$
(B) Signature system ($S = 26$ signatures)	
a through z (equiprobable)	$H_S = 4.70$
(C) Observed signatures in the group ($N=16$ individuals, $M=12$ signatures)	
a a b c c d e f g h i j k k k l	$H_G = 3.45$
(D) Re-classification of group into one target individual and all others	
a x x x x x x x x x x x x x x x	$H_2 = 1.56$

Figure 1. Simple example to illustrate meanings of the different information measures. Individuals 01 through 16 (A) can have signatures a through z, each of which is equiprobable (B). One random draw of signatures is shown in C. Note the duplications in C: two individuals share 'a', two individuals share 'c', three individuals share 'k'. One 'egocentric' re-classification of signatures (with respect to individual 01) is shown in D.

I advanced the argument earlier that selection will favour larger H_S in species (or populations) with greater identification/recognition needs (larger H_0). I now use our discrete signature model to demonstrate a second relationship relating identification needs to the information capacity of the signature system. For simplicity, the S signatures in the set are assumed to be equiprobable. Then $H_S = \log S$ and since $S > N \geq M$, $H_S \geq H_0 > H_G$.

If each of the N individuals within the group draws its signature randomly from the pool of S equiprobable signatures, then we can specify the value of H_S that would allow discrimination at a certain error level. Receivers are assumed to be 'ideal receivers', with an error arising only when two individuals have drawn duplicate signatures. If signatures are equiprobable and drawn at random (with replacement), then the probability of the same signature being drawn more than once is

$$p = 1 - \left(1 - \frac{1}{S}\right)^{N-1}$$

(see Beecher 1982). For $N/S < 0.1$, this simplifies to

$$p = \frac{N}{S}$$

approximately. Rearranging, we see that

$$H_S - H_O = -\log p$$

That is, for relatively error-free identification (and recognition), the information capacity of the signature system must be considerably greater than the initial uncertainty. Note that we have not even considered the additional problems posed by non-ideal receivers, which should favour signature redundancy and further increase the necessary $H_S - H_O$ difference.

A caution is in order here. None of the equations above imply that any receiver actually individually recognizes each individual in the group. As mentioned above, in the most common case, the recognizer has an interest only in discriminating the target individual from all other individuals in the group, and none in discriminating among the remaining individuals (for an exception, see Cheney & Seyfarth 1980). One can conceptualize the signature system from the narrow perspective of one particular recognizer by reducing the signature set to two classes of signatures, the signature of the target individual and the class of signatures of the remaining individuals. Then $M=2$, the reference signature has the information value (surprisal) of $\log N$ and the remaining $N-1$ signatures have the information value of $\log (N/N-1)$ and so the average signature has the information value of

$$H_2 = \frac{1}{N} \log N + \frac{N-1}{N} \log \frac{N}{N-1} \quad (3)$$

That '2' subscript denotes that we have arbitrarily re-classified signatures into two categories. This perspective is useful primarily when the focus is on information transmitted to a particular receiver. For example, discussing an analogous problem in species recognition, Hailman (1977, page 30) has pointed out that a particular duck undoubtedly extracts less information from the plumage traits of the different duck species on a lake than does an experienced bird-watcher, in that the duck is concerned only with distinguishing its species from all the others. Our focus here, however, is not on the information extracted by one receiver from the system, but on the information available to all receivers. In the case of this duck analogy, we are interested in the signature system that would permit a mallard, or a goldeneye, or an individual of any of the other species, to discriminate correctly among species. In the case of our context of

interest, we are interested in the signature system that would permit any receiver (not one particular receiver) to discriminate its target individual from all the rest. Thus for our purposes, equation 2, referring to the average signal, is more appropriate than equation 3, which refers, by implication, to the perception of one particular receiver. As a final note, it is instructive to compare the general meaning of equations 2 and 3. Equation 2 indicates that uncertainty concerning which individual in the group is the target individual increases with N . Equation 3 gives the other side of the coin: uncertainty concerning whether or not a given individual is the target individual decreases with N .

Continuous Signature Model: One Variable Case

Shannon has shown that the average information in a continuous variable is

$$H(x) = -\int p(x) \log p(x) dx$$

where $p(x)$ is the probability density function of x (Shannon & Weaver 1949). Note the analogy to equation 2. Again, this information measure refers to the average value of a signal in the set. Shannon has shown further that

$$H(x) = \log c\sigma \quad (4)$$

where σ is the standard deviation and c is a constant given by the form of the distribution (c ranges from 3.46 for a rectangular distribution to 4.13 for a normal distribution). Note that if we use a rectangular (uniform) distribution to approximate the discrete equiprobable case, equation 4 reduces to $\log S$. In this approximation each signature is assigned a number 1, 2, ..., S (i.e. the width of a category = 1). Thus, S is equivalent to the range of the distribution. Since for a rectangular distribution $\sigma = \text{range}/3.46$, substitution into equation 4 gives $\log S$.

Unlike the discrete variable H , the continuous variable H is a relative, not an absolute measure, the value of H depending on the units of measurement (e.g. it would depend on whether our variable were measured in inches or cm). Related to this problem, zero in this scale of measurement is totally arbitrary; it simply occurs when $\sigma = 1/c$. This means that the information in two continuous variables could not be compared unless they were measured on the same scale. Both of these problems are eliminated, however, when we use the simple linear model to be described next.

The linear model I will develop here is essentially identical to the analysis of variance model II or

random effects model (e.g. Sokal & Rohlf 1981). As we shall see, this model has been used many times in the past in analysing signature traits, although to my knowledge this has never been explicitly recognized. Rather, the model is implicit in the many analyses that have used either linear discriminant functions to classify individuals or have carried out ANOVAs on these data.

Suppose we are measuring a single variable trait, such as the duration of a call, and have n observations each on k individuals. Then by the model a particular observation, X_{ij} , is assumed to be composed of two independent components: a component B_i , reflecting true differences between individuals, and a 'within-individual' or 'error' component, W_{ij} . I treat this last component as originating within the signaller (hence its name) but in fact it could equally well be considered as originating within the receiver (this alternative viewpoint is taken up in the Discussion). Therefore,

$$X_{ij} = B_i + W_{ij} \quad (5)$$

assuming that the means are zero. Because B_i and W_{ij} are independent, the variances have the simple relationship

$$\sigma^2_T = \sigma^2_B + \sigma^2_W \quad (6)$$

where σ^2_T is the total variance in X and σ^2_B and σ^2_W are the variances in B and W , respectively.

H_S is then defined as the amount of information needed to reduce the total uncertainty to the within-individual uncertainty, i.e. by equation 4.

$$H_S = \log c_T \sigma_T - \log c_W \sigma_W$$

Hence, assuming c is the same for total and within distributions

$$H_S = \log \frac{\sigma_T}{\sigma_W} \quad (7)$$

Thus from equation 6

$$H_S = \log \sqrt{\frac{\sigma^2_B + \sigma^2_W}{\sigma^2_W}} \quad (8)$$

H_S so defined has all the properties an information measure should have (see Shannon & Weaver 1949), including the following. (1) Signature information increases directly with σ_B and inversely with σ_W . (2) $H_S = 0$ when $\sigma_B = 0$. (3) H_S is an absolute measure with a non-arbitrary zero, the unit of measure being the within-individual uncertainty. The original units of measurement are immaterial. We can compare, say, the amount of signature information conveyed by the amount of dark

feathering on the face with that conveyed by the average frequency of a call.

Because the linear model leading to equation 8 is formally identical to the ANOVA Model II, we immediately have an appropriate statistical test for the presence of signature information. Note that between 'groups' here is between individuals and within 'groups' is within individuals. The expectations for the between mean square and within mean square are then

$$MS_B = n \sigma^2_B + \sigma^2_W \quad (9)$$

$$MS_W = \sigma^2_W \quad (10)$$

The null hypothesis is that there is no source of variation beyond the inherent within-individual 'noise', σ_W , i.e. $H_0: \sigma_B = 0$. By hypothesis, then, the ratio

$$F = \frac{MS_B}{MS_W} \quad (11)$$

should equal 1. As mentioned in the introduction, it has been common practice to test for signature variation by precisely this statistical test, which implies the assumption of this particular linear model. None of these studies, however, after rejecting the null hypothesis, has proceeded to the next step of evaluating how much signature information is present. In the method I describe here, the same data are used to estimate the available information via equation 8. From equations 8–11 we have the convenient computational formula

$$H_S = \log \sqrt{\frac{MS_B + [n-1]MS_W}{n MS_W}} \quad (12)$$

and see that F and H_S are closely related:

$$H_S = \log \sqrt{\frac{F + n - 1}{n}} \quad (13)$$

Continuous Signature Model: General Multivariate Case

The signature traits typically measured by investigators, usually vocal or visual signals, are inherently multivariate. That is, they can be analysed into a number of intercorrelated variables. Studies in this area have generally overlooked the intercorrelations, doing a separate ANOVA on each variable. It is sometimes assumed that the larger the number of significant F s obtained in such an analysis, the greater the potential signature information. Such an assumption is incorrect, of course, since much of the information may be shared by variables, i.e. may be redundant.

Consider first the hypothetical case where the variables are not intercorrelated. Then the total information H_S is simply the sum of the information H_i in each of the independent variables, and

$$H_S = \sum H_i = \sum \log \frac{\sigma_{Ti}}{\sigma_{wi}} \quad (14)$$

where σ_i is the standard deviation of the i th trait.

In the typical case, however, the variables will be intercorrelated in varying degrees, and equation 14 would be inappropriate for such data. The most direct solution to this problem is to transform the original variables to give a second set of variables which meet the following two criteria: (1) that they be independent, and (2) that they contain the precise amount of non-redundant variance contained in the original set. These criteria are met by a principal components transformation (e.g. Pimentel 1979; Manly 1986). In a principal components transformation, the original variance-covariance matrix \mathbf{V} is transformed into a variance-covariance matrix \mathbf{L} in which all covariances are zero (criterion 1). The product of the variances (eigenvalues, λ_i) of the transformed matrix is equal to the determinant or generalized variance of the original variance-covariance matrix, i.e.

$$\prod \lambda_i = |\mathbf{L}| = |\mathbf{V}| \quad (15)$$

where $|\mathbf{L}|$ and $|\mathbf{V}|$ are the determinants of the respective variance-covariance matrices. Since the generalized variance is the total non-redundant variance of the original variables, the principal components transformation meets our second criterion. The variance estimates from the principal component data are thus the independent variances we wish to substitute into equation 14.

Before submitting the data to the principal component analysis, the variables X_1, X_2, \dots must be reduced to comparable form. In the present case, our theory dictates that we transform the raw scores by the within-individual standard deviation σ_w . That is, we obtain

$$X'_{ij} = \frac{X_{ij}}{\sigma_w} \quad (16)$$

so that

$$\sigma_{T'} = \frac{\sigma_T}{\sigma_w}$$

and

$$\sigma_{w'} = 1$$

where primes designate σ_w -transformed scores. Thus if σ_{T_1}' is say twice σ_{T_2}' then it will be so weighted in the principal components transfor-

mation. (Note that the principal components transformation is done on the variance-covariance matrix, not the correlation matrix, which would weight all variables equally.) That is, the variables are weighted according to the amount of information they contain when considered separately. Without this step, the unit of measurement would be the main determinant of the weighting variables received in the principal components transformation.

The variables X'_1, X'_2, \dots are then submitted to a principal components transformation to give the new, independent variables U_1, U_2, \dots which can then be analysed in separate ANOVAs. Mean square estimates of σ^2_T and σ^2_w are obtained as described above from equations 9 and 10 and the total information is computed from equation 14.

A few points should be made about the relationship between the transformed U variables (principal components) and the original X' variables. First, the numerator of equation 14 could be obtained directly from the original total variance-covariance matrix since the product of the eigenvalues will equal the determinant of the variance-covariance matrix from which they were obtained (equation 15). In general, however, this same relationship will not apply to the denominator of equation 14, since the principal components transformation is based on the total scores, not on the components B or W . Thus although the total variance-covariance matrix based on principal components scores will have zero covariances, the between variance-covariance matrix (based on individual means) and the within variance-covariance matrix (based on residuals) will not.

Two further points may be made about the relationship of the original and transformed variances by considering two special cases. The first case is when all correlations equal zero. While the calculation of the determinant of a matrix is generally complicated, it is the sum of the product of the diagonal elements (variances in a variance-covariance matrix), plus or minus various products involving off-diagonal elements (covariances). Thus, when all correlations equal zero, all covariances equal zero, and the determinant is simply the product of the variances and $|\mathbf{V}| = \log \prod \sigma_i^2$.

The second instructive case is the two-variable case. For a 2×2 matrix, the determinant is simply the product of the two diagonal elements (variances, σ^2) minus the product of the two off-diagonal elements (covariances, $\rho\sigma_1\sigma_2$), or

$$|V| = \sigma^2_1 \sigma^2_2 - \rho^2 \sigma^2_1 \sigma^2_2 \\ = (1 - \rho^2) \sigma^2_1 \sigma^2_2$$

where ρ is the correlation coefficient. In this two-variable case it is readily apparent that the generalized variance is simply the product of the variance in variable 1 and the variance in variable 2 which is not explainable by the correlation between the two variables (the residual variance).

AN ILLUSTRATIVE EXAMPLE

Introduction and Methods

An example, based on a data set created by simulation, is presented here to illustrate features of the information analysis. The simulated data set and all statistical analyses were done using the SYSTAT statistical package (Wilkinson 1986). I will be happy to provide interested persons with the data set on request.

The simulated data set was based on seven independent, normally distributed variables fitting the description of equation 5, i.e. each variable was the sum of two independent variables, B_i and W_{ij} . For each of the seven composite variables, $\sigma^2_w = 1$ while σ^2_b ranged from 2 to 1024. All ρ s were zero. The final data set consists of 10 measurements on each of seven variables from each of 20 'individuals'. The data set was designed to resemble data sets one is likely to obtain with real animals. First, the sample size is quite realistic. Second, if we endeavour to extract the minimum number of variables necessary to characterize the composite signature trait, these variables should have low intercorrelations. Third, the range of σ_T/σ_W is representative of the range I have encountered in real data sets, such as the swallow calls we have analysed, and which will be considered in the Discussion (Beecher 1982; Beecher et al. 1986; Medvin, Stoddard & Beecher, unpublished data).

The major purpose of this example, apart from illustrating the mechanics of the analysis, is to compare our principal component/information analysis with discriminant function analysis of the same data. As mentioned in the introduction, discriminant function analysis has been used in recent studies as a way of quantifying the ability of the signature traits to identify individuals (e.g. Hafner et al. 1979; Smith et al. 1982; Gelfand & McCracken 1986). Typically the data set is split, with one subset used to derive the discriminant functions which are then used to classify the second

subset. How well the second subset is classified is thus a measure of the signature capabilities of the measured variables, at least for the sample considered. Discriminant function analysis resembles the principal component analysis which we have used here in that we derive new variables based on linear combinations of the original variables; both transformations give a number of factors/functions equal to the original number of variables (though they need not all be significant). The criteria for the choosing of the coefficients in the two procedures, however, are somewhat different. In principal component analysis, the coefficients are chosen so that the original variables are transformed into principal components having zero covariances. In discriminant function analysis, the coefficients are chosen so that the original variables are transformed into canonical discriminant functions which reflect differences between the groups as much as possible ('groups' = individuals here). That is, the discriminant functions are chosen so as to maximize the ratio MS_B/MS_W for each of the functions successively. The principal component analysis is constrained so as not to produce any 'new' variance, and so is an appropriate first step for our information measure, which is intended to characterize the original total non-redundant variance. The discriminant function analysis is not so constrained, its purpose being only to separate the groups (individuals) maximally. Any particular observation is classified as to group membership on the basis of its Mahalanobis distance from the group centroid (Pimentel 1979; Manly 1986; Wilkinson 1986).

Analysis and Results

Population and sample values for the simulated data set are shown in Table I (obtained r s, not shown, were all small and non-significant). The data analysis proceeded as follows.

(1) A simple ANOVA was done on each variable in the original data set. Only variables giving a significant F are kept (although in fact any variable not doing so would have little impact on the subsequent analysis). In this case, F for each of the seven variables was highly significant ($P < 0.00001$). The between and within mean squares were used to estimate σ_B and σ_W (equations 9 and 10) for the seven variables (Table I, 'sample'). (2) Each variable in the original data was transformed by its σ_W estimate to give the X' of equation 16. Note that if

Table I. Population parameters and sample estimates (10 observations from each of 20 'individuals') for data set obtained via simulation

Variable	Population*				Sample*				Principal components†					
	σ^2_B	σ^2_W	σ_T/σ_W	H_i	s^2_B	s^2_W	s_T/s_W	H_i	s^2_B	s^2_W	s_T/s_W	H_i		
1	2	1	1.73	0.79	1.81	0.86	1.76	0.82	1.16	0.95	1.49	0.58		
2	4	1	2.24	1.16	6.20	1.02	2.66	1.41	4.32	0.95	2.36	1.24		
3	9	1	3.16	1.66	7.96	0.96	3.05	1.61	6.68	1.06	2.70	1.43		
4	25	1	5.09	2.35	38.4	1.20	5.75	2.52	30.34	1.01	5.57	2.48		
5	64	1	8.06	3.01	64.5	1.03	7.99	3.00	60.92	1.00	7.88	2.98		
6	225	1	15.03	3.91	359.1	1.08	18.3	4.19	299.6	0.98	17.52	4.13		
7	1024	1	32.02	5.00	1232	0.89	37.0	5.21	1411	1.05	36.62	5.19		
			$H_S = 17.88$				$H_S = 18.76$				$H_S = 18.03$			

* Each variable was the sum of two random, normally distributed variables $B_i + W_{ij}$ with means = 0 and variances as indicated under Population.

† Each principal component is listed in the row with the original variable which loaded most heavily on it.

Symbols: σ s represent population parameters, s s represent estimates of those parameters via equations 9 and 10.

the ANOVA were repeated on these transformed scores, the F s would be identical to those of step 1. (3) The X' data were submitted to a principal components transformation. As might be expected for these data, the resulting principal components were similar to the original variables because of the low correlations among variables. (4) A simple ANOVA was done on the principal components data. All seven factors were significant at $P < 0.00001$. The between and within mean squares were used to estimate σ_B and σ_W (equations 9 and 10) for the seven factors (Table I, 'Principal components'). (5) The individual H_i and the overall H_S for each population variable, sample variable and principal component were computed via equations 7 and 14.

Comparing the H_S estimates from Table I, it can be seen that the sample H_S is too high (18.76 versus the true value of 17.88), as expected, since this estimate contains redundant variance (i.e. the variable intercorrelations have not been removed). The H_S obtained from the principal components analysis data (variable intercorrelations removed), however, is close to the true population value (principal components analysis H_S estimate = 18.03 versus the true value of 17.88). Given that we have only a single sample, I will make only two remarks about sampling error. First, in this case it is obviously quite small. Second, in general H_S does not present special problems for evaluating sampling error (see, e.g. Losey 1978), since H_S is a simple derivative of variance estimates for which

there are well-known statistical tests. Additionally, variables which are marginally significant have little effect on the value of H_S .

Because the mathematical bases for the principal components/information analysis and discriminant function analysis are different, the simulated data base was used to derive an empirical measure of the relationship of the two procedures. All seven variables, separately and in several combinations of two or three variables, were analysed by both procedures. The outcome of the principal components/information analysis is H_S . The outcome of the discriminant function analysis is the percentage of observations that are correctly classified as to group ('individual'). In the discriminant function analysis, one-half of the data set was used to derive the discriminant functions used to classify the second half of the data set (i.e. the first subset consisted of five observations each on 20 'individuals', and the second subset consisted of five additional observations on each of the same 20 'individuals').

A comparison of H_S (from the principal component/information analysis) and the probability of correct classification (P , from the discriminant function analysis) is shown in Fig. 2. Twenty data points are shown: all seven variables considered separately, seven pairs of variables, and six trios of variables. Although the data points are not all independent, the function is essentially identical to smaller functions containing all independent points (e.g. each of the seven variables considered separa-

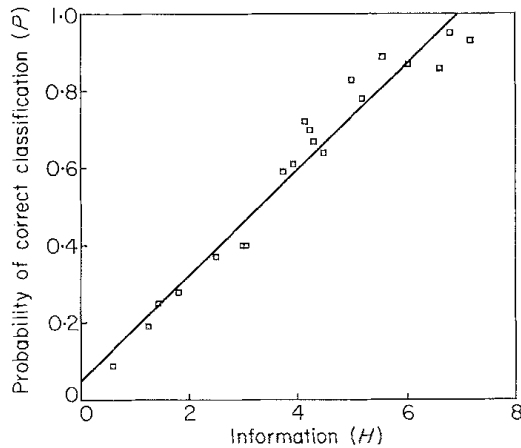


Figure 2. Probability of correct classification (P), obtained from the discriminant function analysis, as a function of the information capacity (H_S) of the variable set. Classification is for the full set of 20 'individuals'. Points are the seven variables considered separately, seven pairs of these variables, and six trios of variables (see text). Best fit line: $P=0.05+0.14H$ ($r=0.98$).

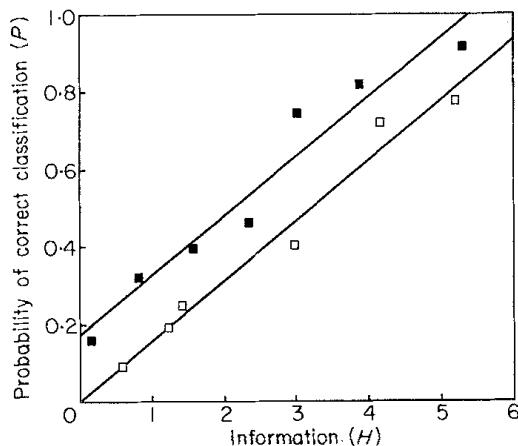


Figure 3. Probability of correct classification (P) as a function of information capacity (H_S); shown for the seven variables taken separately. The parameter is the number of 'individuals' (10 or 20). ■: $k=10$ individuals. The best fit line for these data is: $P=0.17+0.15H$ ($r=0.97$). □: $k=20$ individuals. Best fit line for these data is: $P=0+0.16H$ ($r=0.99$).

tely). Note that $P=0.05$, chance level for 20 individuals, when $H_S=0$. Note also that we 'hit ceiling' ($P \approx 1.0$) at $H_S \approx 8$. Any combination of variables giving $H_S \geq 8$, gave $P \geq 0.99$ (usually 1.0) and these data points are not plotted in Fig. 2. The major point here is that H_S and P are clearly measuring the same thing. The major advantage of H_S is that it provides a measure that is independent of the particular conditions of the sample. In particular, the value of P obtained from the discriminant function analysis depends on the number of individuals in the sample, and has no general meaning outside of this context. For a given H_S , decreasing the number of individuals in the sample will increase P (unless H_S is high enough that P is already at its 1.0 ceiling). For example, variables 4 and 6 together ($H_S=3.72$) do a poor job of allocating observations to individuals for the full sample of 20 individuals (59% correct classification, point shown in Fig. 2) but do better for a smaller sample of 10 individuals (76% correct classification, not shown in Fig. 2). A demonstration of this difference is shown in Fig. 3, which, for each of the seven variables separately, compares a half of our full data set (10 individuals) with the full data set (20 individuals). It can be seen that for a given value of H_S , the probability of correctly classifying an individual averages about 0.2 higher for the smaller data set. The advantage of H_S as a measure is that it is independent of the number of individuals to be classified, while predicting our ability to classify an individual as to identity.

DISCUSSION

Assumptions of the Information Analysis

The information analysis approach that I have described makes a number of assumptions. When the goal of the application is to evaluate the absolute value of the obtained H_S , these assumptions must be fully met. When the goal of the application is to compare the relative values of H_S for two or more species (or two or more modalities within a species), only much weaker assumptions are required. As an example of an absolute-value application, in my preliminary development of this information model, I predicted the information capacity of the chick call of the bank swallow, *Riparia riparia*, necessary for a parent to find its chick reliably in the creche of typical size and I compared this with the value I obtained (Beecher 1982). This exercise requires satisfaction of all of

the assumptions I will list below (and of some problematic guesses, e.g. the 'acceptable' level of error for a parent, the 'typical' size of a creche). I now regard this exercise as naive (although a worthwhile pedagogical enterprise, as it clearly illustrates the general meaning of H_S). As an example of a relative-value application, in the same paper (Beecher 1982) I predicted that the information capacity of the bank swallow chick call should be greater than that of the homologous call of the rough-winged swallow, *Stelgidopteryx serripennis*, as the former species is highly colonial and the latter is not (see similar argument below).

I will detail these assumptions in their strongest form (that are required for absolute evaluations of H_S) and indicate as well the weaker requirements of relative (comparative) evaluations.

Assumption 1: ideal receiver

Our method provides a measure of information capacity of the signals, not information extracted by the receiver. For the absolute value of H_S to have meaning, the receiver must have extracted all the information from the signal that we have extracted. For comparative analyses, the following, weaker version of this assumption is required.

In our approach σ_w is used as the 'error' term in the evaluation of σ_T . In this paper I have treated σ_w as residing within the sender (e.g. its calls vary), but I could just as well treat it as residing within the receiver (its perception of the calls varies), or as a composite of both. The theory is neutral on this point. Here I distinguish these two sources of error variation as the within-sender σ_w and the receiver's 'just noticeable difference' (JND). Although some of our measured within-sender σ_w may actually arise within the measuring instruments (e.g. microphone, tape-recorder, spectrograph, spectrogram measurer, etc.), in practice we will use calibration procedures to show that measurement error accounts for only a small proportion, relative to true sender variation, of the measured σ_w . However, we will often not have information on the receiver's JND. This lack of information could lead to serious misinterpretation if in fact the receiver's JND is considerably larger than the within-sender σ_w , or if the JND and σ_w have an unpredictable relationship, and the species being compared differ in this respect. For comparative analyses, we need only assume that the JND is consistently less than the within-sender σ_w , or that the two have a consistent relationship across variables and spe-

cies. Note that when the receiver's JND is considerably less than the within-sender σ_w , we have an approximation of the 'ideal receiver' case. That is, the limits on identification are not the receiver's ability to distinguish two similar signals, but the sender's ability to present the same signal from one time to the next.

Assumption 2: completeness

To assign significance to our obtained absolute value of H_S , we must have extracted all of the relevant information. Depending on our goal, this may be all of the signature information used by the species, or all of the information in a particular modality (odours versus calls for example). For comparative purposes, however, it is only necessary that we have extracted (1) most of the relevant information, and (2) a similar amount for all of the species being compared (or if not a similar amount, that the error be in the direction opposite that of the hypotheses). A good initial check on this assumption is the 'reconstruction' criterion: can we, from our extracted measurements, reconstruct a good replica or model of the original? From call measurements, can we reconstruct a good replica of the original spectrogram? From measurements of egg colour patterns (see Buckley & Buckley 1972; Shugart 1987), can we make a model egg that looks like the real thing?

The completeness criterion really refers to two things, only the second of which is evaluated by the 'reconstruction' criterion. (1) Are all the relevant variables measured to begin with? (2) Are all the relevant variables extracted in the final data reduction? To take call measurements as an example, it is well known that the sound spectrograph largely fails to represent amplitude information (step 1). Additional information may be lost when we extract measurements from the spectrogram (step 2). In this instance it is relatively easy to evaluate the second step but we can evaluate the first step only if we use an instrument suitable for extracting amplitude information (e.g. an oscilloscope).

Assumption 3: variable weighting

Our method weights each parameter in accordance with σ_T/σ_w . This weighting is central to the approach, of course, but some parameters may be more perceptually salient to the animal than other parameters.

Testing the perceptual assumptions

It can be seen that all three of these assumptions are essentially questions about how the receiver analyses (perceives) the signature traits under consideration. In given instances we may know enough about the perception of the particular group with respect to the particular modality that these perceptual assumptions are not a major problem. For example, the perceptual assumptions problem should not be too severe in a comparison of the extent of facial variation in several species of primates, if for no other reason than that the perception of the species in question is likely to be very much like our own (though undoubtedly not identical). On the other hand, a conclusion that several species of bees differ in the extent of signature variation in odours would certainly require a serious evaluation of these perceptual assumptions.

The most direct way to test these assumptions is via a direct investigation of the animal's perception of the signals in question. Unfortunately, most often it will not be practical to test perception in the same detail as the signal itself. In some cases, however, perceptual studies can be used as a check on conclusions of the signal analysis, or can be used to probe particular interesting conclusions. For example, we have analysed the chick call of the colonial cliff swallow, *Hirundo pyrrhonota*, and the non-colonial (or semi-colonial) barn swallow, *Hirundo rustica*, and found that H_S is approximately five bits greater for the cliff swallow call (preliminary accounts in Beecher et al. 1986; Beecher et al. in press; the full account is in preparation, Medvin, Stoddard & Beecher, unpublished data). This species difference is consistent with the prediction described above for bank swallows and rough-winged swallows, and with field experiments on cliff swallows and barn swallows (Stoddard & Beecher 1985; Medvin & Beecher 1987). To check on the perceptual assumptions, we carried out laboratory studies of the perception of these calls by both cliff swallows and barn swallows, using conditioning procedures (Beecher et al., in press). These laboratory studies showed that both cliff swallows and barn swallows can discriminate more easily among the calls of different cliff swallows than among the calls of different barn swallows. Moreover, birds of both species were able to discriminate among the individual calls in a set of calls, thus showing that true individual recognition is possible, even though not manifested in the field

(where, as per our earlier discussion, the bird is interested merely in the distinction between its offspring and unrelated chicks). Finally, we could predict with some accuracy a bird's ability to discriminate between particular calls on the basis of the measured difference between calls, using the variables and weightings of the information analysis. Thus, these laboratory experiments generally support the perceptual assumptions underlying the information analysis in this case. To derive an information measure directly from perceptual experiments, however, will generally require a substantially greater investment of time than a signal analysis, and in general will not be practical. Nor will it generally be feasible to do perceptual tests in the field, if for no other reason than that animals will not respond to heterospecific signals under normal circumstances. If animals are tested only on conspecific signals, sender characteristics and receiver characteristics will be confounded.

Meaning and Uses of the Information Analysis Approach

As our comparison of the information and discriminant function analysis approach illustrated above, H_S measures the extent to which the signature system permits correct identification of individuals. H_S is ultimately translatable into the size of a group in which an individual could be identified to some particular degree of accuracy. If we take our translation rule from the analysis of Figs 2 and 3, then for a five-bit signature system and 90% accuracy, this group size is somewhere between 10 and 20 individuals. This particular translation rule assumes an 'ideal receiver', i.e. one that extracts all of the available information from the signals and assigns identities according to the optimality rule specified in the discriminant function analysis. It is obvious that one could develop a formal model to predict the probability of correct classification given H_S and the group size. A very simple example was developed for the discrete signature model earlier in the paper. I will not pursue this approach further here, however, as I believe the power of the method lies rather in comparative analyses not requiring prediction of precise values of H_S .

In conclusion, I suggest that the method I have described here will be most useful for comparative analyses. If, as in the swallow example described above, we can rank several species in terms of some variable affecting recognition (e.g. coloniality),

then we would predict that the information capacities of their signature systems should be ranked similarly. Another major use of the method is in disparate comparisons. For example, we might have reason to compare the individual distinctiveness of the scent mark of a particular mammal with that of the song of a particular bird. Provided we could adequately address the assumptions listed above, so that we had confidence that the analyses were relatively complete and comparable then the information measure would permit this sort of apples-and-oranges comparison. While we have focused on the hypothesis that selection has increased individual distinctiveness, an information analysis may be used to test the contrary hypothesis that selection has decreased individual distinctiveness. For example, several hypotheses have proposed that selection has favoured decreasing the individual distinctiveness of bird songs (e.g. Falls 1982; Beecher & Stoddard, in press). Finally, as suggested above, the information analysis can be based directly on perceptual data if the perceptual assumptions are questionable, or if it is relatively easy to get perceptual data. For example, one might extract the relevant dimensions of the signatures via a multidimensional scaling analysis of perceptual data (e.g. Dooling et al., in press). Provided we could identify the stimulus correlates of these dimensions, we could then use either perceptual JNDs or within-individual σ_{ws} (or both) as our error term.

ACKNOWLEDGMENTS

I thank Mark Pagel, Philip Stoddard and Mandy Medvin for their comments on earlier drafts of this paper. Jack P. Hailman provided some particularly helpful criticism. The work was supported by grants from National Science Foundation.

REFERENCES

- Attneave, F. 1959. *Applications of Information Theory to Psychology*. New York: Holt.
- Beecher, M. D. 1982. Signature systems and kin recognition. *Am. Zool.*, **22**, 477–490.
- Beecher, M. D. 1988. Kin recognition in birds. *Behav. Genet.*, **18**, 465–482.
- Beecher, M. D., Loesche, P., Stoddard, P. K. & Medvin, M. B. In press. Individual recognition by voice in swallows: signal or perceptual adaptation? In: *Comparative Psychology of Audition: Perceiving Complex Sounds* (Ed. by S. H. Hulse & R. J. Dooling), Hillsdale, New Jersey: Erlbaum.
- Beecher, M. D., Medvin, M. B., Stoddard, P. K. & Loesche, P. 1986. Acoustic adaptations for parent-offspring recognition in swallows. *Exp. Biol.*, **45**, 179–193.
- Beecher, M. D. & Stoddard, P. K. In press. The role of bird song and calls in individual recognition: contrasting field and laboratory perspectives. In: *Comparative Perception* (Ed. by M. Berkeley & W. C. Stebbins), New York: John Wiley.
- Beecher, M. D., Stoddard, P. K. & Loesche, P. 1985. Recognition of parents' voices by young cliff swallows. *Auk*, **102**, 600–605.
- Buckley, P. A. & Buckley, F. G. 1972. Individual egg and chick recognition by adult royal terns (*Sterna maxima maxima*). *Anim. Behav.*, **20**, 457–462.
- Cheney, D. L. & Seyfarth, R. M. 1980. Vocal recognition in free-ranging vervet monkeys. *Anim. Behav.*, **28**, 362–367.
- Colgan, P. 1983. *Comparative Social Recognition*. New York: John Wiley.
- Cullen, E. 1967. Adaptations in the kittiwake to cliff-nesting. *Ibis*, **99**, 275–303.
- Dooling, R. J., Park, T. J., Brown, S. D., Okanoya, K. & Soli, S. D. In press. Perceptual organization of acoustic stimuli by budgerigars: II. Vocal signals. *J. comp. Psychol.*
- Falls, J. B. 1982. Individual recognition by sounds in birds. In: *Acoustic Communication in Birds. Vol. 2* (Ed. by D. E. Kroodsma & E. H. Miller), pp. 237–278. New York: Academic Press.
- Gelfand, D. L. & McCracken, G. F. 1986. Individual variation in the isolation calls of Mexican free-tailed bat pups, *Tadarida brasiliensis mexicana*. *Anim. Behav.*, **34**, 1078–1086.
- Hafner, G. W., Hamilton, C. L., Steiner, W. W., Thompson, T. J. & Winn, H. E. 1979. Signature information in the song of the humpback whale. *J. Acoust. Soc. Am.*, **66**, 1–6.
- Hailman, J. P. 1977. *Optical Signals: Animal Communication and Light*. Bloomington: Indiana University Press.
- Jouventin, P. 1982. *Visual and Vocal Signals in Penguins, their Evolution and Adaptive Characters*. Berlin: Verlag Paul Parey.
- Losey, G. S. 1978. Information theory and communication. In: *Quantitative Ethology* (Ed. by P. W. Colgan), pp. 43–78. New York: John Wiley.
- McCracken, G. F. 1984. Communal nursing in Mexican free-tailed bat maternity colonies. *Science, N.Y.*, **223**, 1090–1091.
- McCracken, G. F. & Gustin, M. F. 1987. Batmom's daily nightmare. *Nat Hist.*, **96**, 66–73.
- Manly, B. F. J. 1986. *Multivariate Statistical Methods: A Primer*. London: Chapman & Hall.
- Medvin, M. B. & Beecher, M. D. 1987. Parent-offspring recognition in the barn swallow, *Hirundo rustica*. *Anim. Behav.*, **34**, 1627–1639.
- Pimentel, R. A. 1979. *Morphometrics: The Multivariate Analysis of Biological Data*. Dubuque, Iowa: Kendall/Hall.
- Quastler, H. 1958. A primer on information theory. In: *Symposium on Information Theory in Biology* (Ed. by

- H. P. Yockey & R. L. Platzman), pp. 3–49. New York: Pergamon Press.
- Shannon, C. E. & Weaver, W. 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.
- Shugart, G. W. 1987. Individual clutch recognition by Caspian terns, *Sterna caspia*. *Anim. Behav.*, **35**, 1563–1565.
- Smith, H. J., Newman, J. D., Hoffman, H. J. & Fetterly, K. 1982. Statistical discrimination among vocalizations of individual squirrel monkeys (*Saimiri sciureus*) *Folia Primatol.*, **37**, 267–279.
- Sokal, R. R. & Rohlf, F. J. 1981. *Biometry*. 2nd edn. San Francisco: W. H. Freeman.
- Stoddard, P. K. & Beecher, M. D. 1985. Parental recognition of offspring in the cliff swallow. *Auk*, **100**, 795–799.
- Wilkinson, L. 1986. *SYSTAT: The System for Statistics*. Evanston, Illinois: SYSTAT, Inc.
- Wilson, E. O. 1975. *Sociobiology*. Cambridge, Massachusetts: Harvard University Press.

(Received 18 May 1988; revised 29 August 1988;
MS. number: A5193)