

# Identifying the structure in cuttlefish visual signals

Anne C. Crook<sup>1</sup>, Roland Baddeley<sup>2\*</sup> and Daniel Osorio<sup>2</sup>

<sup>1</sup>Department of Zoology and Animal Ecology, University College, Lee Maltings, Cork, Ireland (a.crook@ucc.ie)

<sup>2</sup>Laboratory of Experimental Psychology and School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK

The common cuttlefish (*Sepia officinalis*) communicates and camouflages itself by changing its skin colour and texture. Hanlon and Messenger (1988 *Phil. Trans. R. Soc. Lond. B* **320**, 437–487) classified these visual displays, recognizing 13 distinct body patterns. Although this conclusion is based on extensive observations, a quantitative method for analysing complex patterning has obvious advantages. We formally define a body pattern in terms of the probabilities that various skin features are expressed, and use Bayesian statistical methods to estimate the number of distinct body patterns and their visual characteristics. For the dataset of cuttlefish coloration patterns recorded in our laboratory, this statistical method identifies 12–14 different patterns, a number consistent with the 13 found by Hanlon and Messenger. If used for signalling these would give a channel capacity of 3.4 bits per pattern. Bayesian generative models might be useful for objectively describing the structure in other complex biological signalling systems.

**Keywords:** cephalopod; communication; auticlass; mixture models; Bayesian methods; cluster analysis

## 1. INTRODUCTION

In common with their cephalopod relatives squid and octopus, cuttlefish display a wide range of visual patterns, which are used for crypsis, interspecific and intraspecific communication (Hanlon & Messenger 1996). These patterns can be defined as signals, insofar as they affect the behaviour of other animals (Maynard Smith & Harper 1995), but note that this is a different use of the term ‘signals’ from that employed by Hanlon & Messenger (1996), where the term is restricted to patterns used for social communication. The expression of these signals is probably not random, but structured in some way, and can be assumed to reflect the underlying signalling state of the animal. One proposal is that the structure is hierarchical (Packard & Hochberg 1977), with the visual signals falling into a set of major categories called ‘body patterns’. Although the term ‘body pattern’ is not mathematically defined in the literature, a reasonable definition might be that two signals within a body pattern are more similar to one another than are any two signals from different patterns.

To produce signals cephalopods vary the expression of four types of ‘component’ (Hanlon & Messenger 1988): (i) chromatic (i.e. skin coloration); (ii) skin texture (e.g. rough or smooth); (iii) postural; and (iv) locomotor. For the common cuttlefish, *Sepia officinalis*, Hanlon & Messenger (1988, 1996) identify 34 chromatic, six textural, eight postural and six locomotor components, each of which they assign a unique number (figure 1). Thus, light and dark ‘zebra’ stripes are components 9 and 26, respectively.

Shallow water cephalopods, such as *S. officinalis*, are thought to communicate primarily with chromatic and postural signals (Hanlon & Messenger 1996). ‘Colour’ itself (hue and saturation as recognized by humans) is relatively unimportant as most cephalopods, including *S. officinalis*, are believed to be colour-blind (Messenger 1979; Marshall & Messenger 1996).

Cephalopods are remarkable for the speed and flexibility with which they generate visual signals. For the Caribbean reef squid (*Sepioteuthis sepioidea*), the richness of the signalling repertoire has led to the suggestion that it has a ‘language’ with up to 35 distinct signalling states (i.e. words; Moynihan & Rodaniche (1982); but cf. Hanlon & Messenger (1996, pp. 129–130)). More generally, body patterns have been qualitatively described for several species (e.g. *S. officinalis*—Holmes (1940); Hanlon & Messenger (1988); *Sepioteuthis sepioidea*—Moynihan & Rodaniche (1977, 1982); *Metasepia pfefferi*—Roper & Hochberg (1988); *Euprymna scolopes* and *Idiosepius pygmaeus*—Moynihan (1983a,b); review—Hanlon & Messenger (1996)), but so far a quantitative method for identifying even basic characteristics of these signalling systems is lacking. To begin a systematic study relating signals to behavioural context we give a formal definition of a body pattern, and go on to suggest how these patterns can be classified.

Just as there are field guides for animal identification, a guide to signalling states of a given species would be useful. The method we use for *S. officinalis* generates taxonomies of signals based on putative underlying signalling states (figure 2). The method will suit any communication system characterized by: (i) individual, low-level constituent components of the signals that can be clearly defined; and (ii) these low-level components that can be described as several discrete modes (e.g. body patterns). By applying a method previously used for classification of the spectral classes of stars and genetic sequences (Goebel *et al.* 1989;

\* Author for correspondence (rjb@biols.susx.ac.uk).

One contribution of 12 to a Theme Issue ‘Information and adaptive behaviour’.

body patterns and their components in *Sepia officinalis*

<p>light</p> <ul style="list-style-type: none"> <li>(1) white posterior triangle</li> <li>(2) white square</li> <li>(3) white mantle bar</li> <li>(4) white lateral stripe</li> <li>(5) white fin spots</li> <li>(6) white fin line</li> <li>(7) white neck spots</li> <li>(8) iridescent ventral mantle</li> <li>(9) white zebra bands</li> <li>(10) white landmark spots</li> <li>(11) white splotches</li> <li>(12) white major lateral papillae</li> <li>(13) white head bar</li> <li>(14) white arm triangle</li> <li>(15) pink iridophore arm stripes</li> <li>(16) white arms spots (males only)</li> </ul>	<p>chromatic components</p> <p style="margin-left: 40px;">mantle</p> <p style="margin-left: 40px;">head or arms</p>	<p>dark</p> <ul style="list-style-type: none"> <li>(17) anterior transverse mantle line</li> <li>(18) posterior transverse mantle line</li> <li>(19) anterior mantle bar</li> <li>(20) posterior mantle bar</li> <li>(21) paired mantle spots</li> <li>(22) median mantle stripe</li> <li>(23) mantle margin stripe</li> <li>(24) mantle margin scalloping</li> <li>(25) dark fin line</li> <li>(26) black zebra bands</li> <li>(27) mottle</li> <li>(28) latero-ventral patches</li> <li>(29) anterior head bar</li> <li>(30) posterior head bar</li> <li>(31) pupil</li> <li>(32) eye ring</li> <li>(33) dark arm stripes</li> <li>(34) dark arms</li> </ul>
<p>textural components</p> <ul style="list-style-type: none"> <li>(35) smooth skin</li> <li>(36) coarse skin</li> <li>(37) papillate skin</li> <li>(38) wrinkled first arms</li> <li>(39) white square papillae</li> <li>(40) major lateral papillae</li> </ul>	<p>postural components</p> <ul style="list-style-type: none"> <li>(41) raised arms</li> <li>(42) waving arms</li> <li>(43) splayed arms</li> <li>(44) drooping arms</li> <li>(45) extended 4th arm</li> <li>(46) raised head</li> <li>(47) flattened body</li> <li>(48) flanged fin</li> </ul>	<p>locomotor components</p> <ul style="list-style-type: none"> <li>(49) sitting</li> <li>(50) bottom suction</li> <li>(51) buried</li> <li>(52) hovering</li> <li>(53) jetting</li> <li>(54) inking</li> </ul>

body patterns

chronic patterns

- 1. uniform light
- 2. stipple
- 3. light mottle
- 4. dark mottle
- 5. disruptive
- 6. weak zebra

acute patterns

- 7. uniform blanching
- 8. uniform darkening
- 9. acute disruptive
- 10. flamboyant
- 11. deimatic
- 12. intense zebra
- 13. passing cloud

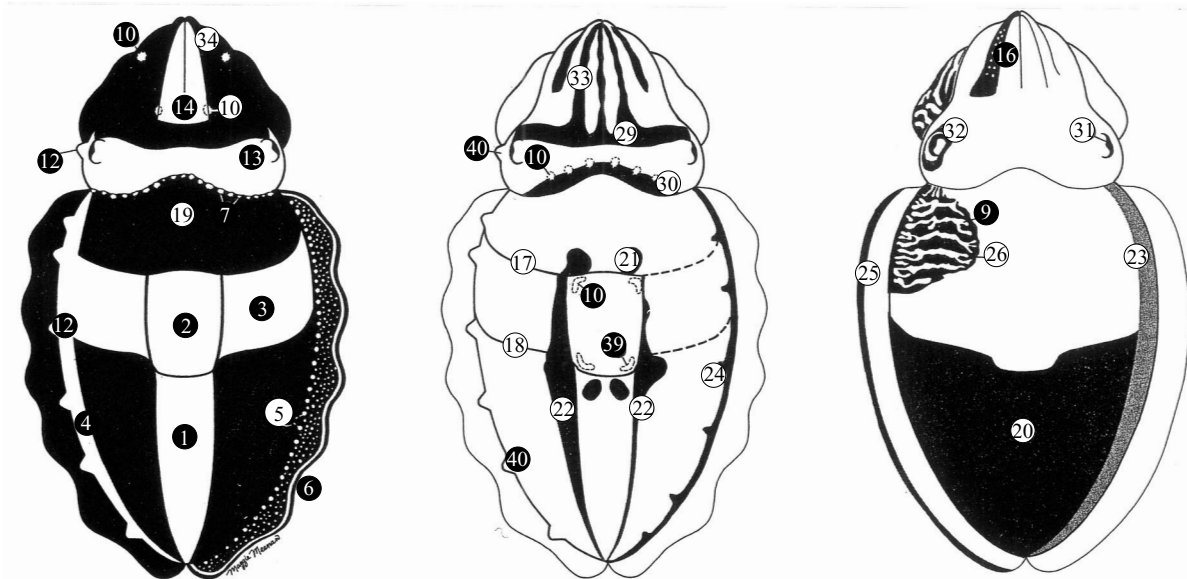


Figure 1. Behavioural components and body patterns of juvenile cuttlefish (*Sepia officinalis*). From Hanlon & Messenger (1988), reproduced with permission.

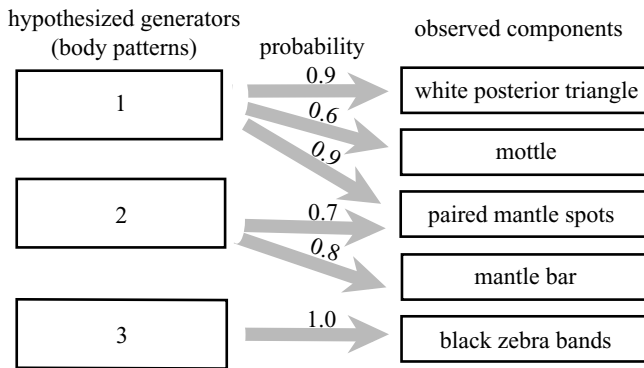


Figure 2. Summary of the experimental strategy. At any one time a cuttlefish displays several pattern components. From the expression of these components, a clustering algorithm infers the number and properties of underlying generators or body patterns. Each pattern is defined by the probabilities of expression of each of the components. Arrows are not shown when there is zero probability of a component being expressed by a given generator. The expression pattern shown here is diagrammatic and is not based on real cuttlefish behaviour.

Cheeseman & Stutz 1996), we propose that it is possible to automatically identify the number and kind of signals available to a receiver.

Why is automating signal classification necessary? An experienced field worker can reliably identify distinct body patterns, but for the new researcher, and to compare results between different laboratories, an automated classification system using a statistical method has advantages. In any case, human observers are notorious for identifying strong regularities in random variation (e.g. finding constellations in the stars), and for missing regularities in data that have many dimensions of variation. Also, when the number of components is large, it is important to separate genuine structure in signals from apparent regularities caused by a limited number of samples generated by a stochastic process.

Our approach is to propose a very large set of models (hypotheses) for how the observed signals were generated (a generative model approach; Roweis & Ghahramani (1999)), and then to use probability theory to calculate the relative (posterior) probability that these signals were generated by a given model. Each hypothesis consists of: (i) the number of distinct body patterns; (ii) the probability of displaying each of these patterns; and (iii) the probability of expressing each of the different components when displaying any given body pattern (54 for cuttlefish, *S. officinalis*; Hanlon & Messenger (1996)). Bayesian statistical methods (see later) give the probability that a given hypothetical generator had generated the observed data. This allows us explicitly to: (i) test the hypothesis that there are distinct 'body patterns', against a hypothesis that the observations are a small sample of patterns from a dataset with random variation; and (ii) to test hypotheses about the number of patterns.

## 2. METHODS

### (a) Recording

Four juvenile (or sub-adult) *S. officinalis* (mean mantle length 73 mm) were maintained in laboratory aquaria at the University

of Sussex between June and September 2000. Animals were filmed in isolation, except for recording social interactions. We took care to ensure that body patterns were not attributable to unintentional disturbance. Animals were recorded in a range of behavioural contexts as follows:

- (i) No stimulus (318 examples). At the start of each recording session, the focal animal was recorded in the absence of additional stimuli. This provided background data relating to body patterns in untested, experimentally naive individuals.
- (ii) Feeding (116 examples). *S. officinalis* were filmed hunting and eating shore crab (*Carcinus maenas*) and glass prawn (*Palaemonetes pugio*) that were present in the aquarium before the subject was introduced.
- (iii) Freely moving prey (184 examples). Prey as above, but separated from the focal cuttlefish by plastic mesh and therefore inaccessible to it.
- (iv) Prey presented in transparent tubes (135 examples). The focal *S. officinalis* was given the same prey as above held captive within sealed and weighted transparent plastic tubes.
- (v) Social (35 examples). Focal *S. officinalis* were filmed in the presence of either one or two conspecifics.
- (vi) Squid/fish model or *S. officinalis* image (20 examples). Models of fish (fishing lures), squid and juvenile conspecifics were presented to the focal *S. officinalis*. Models were held outside the aquarium on a piece of rigid wire. Care was taken to ensure that the model alone was visible. Since cuttlefish quickly habituate to models, only data relating to the first two presentations of each model were analysed.

Aquaria were illuminated with diffuse overhead lighting and surrounded by cloth to minimize disturbance, and were maintained at  $17 \pm 2$  °C, 12 L : 12 D diurnal cycle. One aquarium (800 mm  $\times$  800 mm, water depth 150 mm) was designed for filming with a horizontally placed camera (Canon XL-1, digital video), having a large mirror positioned above it at an angle of 45°. Focal individuals were allowed to acclimatize to the filming aquarium for a minimum of 2 h before recording. A total of 808 individual frames from the digital video were captured with Canon Data Commander software, and constituent components of patterns classified according to Hanlon & Messenger's (1988) scheme (figure 1).

### (b) Computation

#### (i) Data format

Each cuttlefish image was represented by a vector of component expressions, consisting of ones and zeros, according to whether or not each of the 54 body-pattern components recognized by Hanlon & Messenger (1988) was present (figure 1).

#### (ii) Modelling the structure of cuttlefish patterns

Is it best to describe the cuttlefish signals as coming from a continuum or from discrete signalling states (with possible variation within each state)? We use Bayesian statistical methods to turn this 'cluster analysis' problem into one of probabilistic inference. The objective is to estimate the relative probabilities that the observed patterns were created by differing numbers of unknown hypothetical pattern 'generators' (figures 2 and 3). In statistics, the analysis of data by finding the model most likely to have generated the data is known as generative modelling (Roweis & Ghahramani 1999).

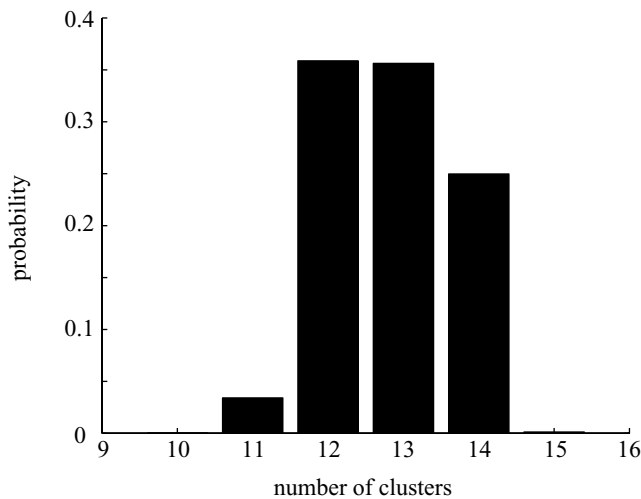


Figure 3. The relative probability of the number of distinct body patterns being from 1 to 20. Only the probabilities between 9 and 16 are shown, as all others are infinitesimally small.

For the dataset of cuttlefish images, an appropriate definition of a single generator is as a set of probabilities of each of the components being expressed when the generator is present (figure 2). Further, we assume that only one generator is ever active at one time, and the components are conditionally independent. For example, consider a system with only two components; say a left and right eyespot. A single generator with probabilities of 0.1 and 0.9 would mean that the left eyespot had a 10% chance of being expressed compared with a 90% chance of the right eyespot, when this generator was present.

The simplest model is therefore one where a single generator generates all patterns. If so, the only structure to the patterns would be that some components were more probable than others (though a given dataset could appear to have more structure simply from the effects of sampling). More complex models can be constructed by combining multiple generators in the following manner: (i) a generator is selected at random from a pool of possible generators, each associated with a different probability of being selected; (ii) the chosen generator expresses the 54 components with their associated probabilities. This type of statistical model is known as a mixture of Bernoullis (Ripley 1996).

In the framework of Hanlon & Messenger (1988) a single type of body pattern comprises a specific set of components being displayed. Our notion of body pattern is more flexible in that for a given pattern components are neither present or absent, but are expressed with a specific probability (0: never to 1: always present in the pattern displayed). Identifying different body patterns by finding the most probable model of data generation has the important advantage that it carries out cluster analysis on a well-founded probabilistic basis.

We then need to estimate three sets of parameters: (i) the number of generators (i.e. body patterns); (ii) the probability that each generator occurs; and (iii) the probability of each component being expressed in a given generator (figure 2). The objective is to find a set of parameters that maximizes the probability of generating the observed data.

Simply maximizing the probability of generating the data (maximum likelihood) would produce the trivial result that the number of generators is equal to the number of distinct patterns in the image database. Therefore, we need to place a principled

cost on model complexity. As complicated models are intrinsically less probable than simple ones any additional complexity in a model (additional generators, in our case) must be justified by an improved fit to observations. We evaluated models that postulated from one to 19 body patterns, as initial shorter runs showed that all high probability solutions lay within this range.

Modelling was done with the AUTOCLASS system (Stutz & Cheeseman 1995), which is widely available (<http://ic.arc.nasa.gov/ic/projects/bayes-group/autoclass/>), and Bayesian methods were then used to estimate the number of different generators (Bishop 1995). AUTOCLASS uses a combination of multiple restarts, an expectation-maximization algorithm, and a split and merge process to maximize the probabilities for a given mode (there are similar alternatives such as minimum message length based SNOB; Wallace & Dowe (2000)). AUTOCLASS was run with default parameters, except that there were 1000 optimization steps per model (to improve the quality of the solutions at the expense of computer time). The high dimensionality (54) of the dataset of cuttlefish images means that to ensure a good fit many models must be evaluated. With the conservative parameters used, the main simulation took 19 days (450 h) on an Alpha EV6 500 MHz RISC workstation.

### 3. RESULTS

We photographed body patterns of four *S. officinalis* individuals in laboratory aquaria under a variety of controlled behavioural situations. Images were classified according to the presence or absence of the 54 known possible body pattern components (chromatic, textural, postural and locomotor; Packard & Hochberg (1977), Hanlon & Messenger (1988)). A database of 808 images was classified, and from this we investigated the statistical structure of body patterns. Specifically, the number of distinct body patterns within the dataset, and the structure of each pattern in terms of the component probabilities of each of the 54 components being expressed.

A 'body pattern' is defined as a cluster in the 54-dimensional space given by the expression of the basic behavioural components. A total of 194 688 runs of the AUTOCLASS algorithm evaluated and optimized different models, testing the probability of the number of body patterns being 1–19. The most probable number of body patterns (i.e. clusters) was 12, with 13 or 14 being within the same range (figure 3). Other possible numbers of body patterns were much less likely. This result can be compared with 13 basic types of pattern identified by Hanlon & Messenger (1988), but no direct correspondence can be made. For example we recorded only captive juveniles; however, adults and juveniles in their natural habitat may show (possibly many) more types of pattern. For example, only 41 of the 54 components were expressed in our image database (table 1). We think it unlikely that increasing the number of images would have yielded further clusters within the range of patterns observed in our laboratory, but that recording from more naturalistic conditions would.

Individual images could be assigned to a cluster (body pattern) with confidence. The average probability of an image belonging to the most likely cluster (of 12) exceeded 97%, and 95% of images could be assigned to one of the 12 clusters with greater than 84% confidence.

Table 1 shows graphically the probabilities associated

Table 1. The 12 most probable clusters in order of decreasing probability of occurrence.

(Each column represents a given cluster, and each row one of the 54 different components in the scheme of Hanlon & Messenger (1988). Only components present in at least one of our images are shown. A five-level coding scheme is employed: probabilities of occurrence between 0 and 0.2 are shown as '--'; between 0.2 and 0.4 as '-'; between 0.4 and 0.6 as '.'; between 0.6 and 0.8 as '+' and between 0.8 and 1.0 as '++'.)

	1	2	3	4	5	6	7	8	9	10	11	12
1	--	--	++	--	--	--	--	--	--	--	-	--
2	--	--	++	--	--	--	.	--	--	+	-	--
3	--	--	--	--	--	--	--	--	--	--	--	--
5	+	++	++	+	++	+	++	-	++	--	++	--
7	--	--	.	--	--	--	--	--	--	--	.	--
9	--	++	--	--	++	--	++	--	++	++	++	--
10	--	+	--	-	--	--	.	--	-	++	++	--
11	--	--	--	--	--	--	--	--	--	--	--	.
12	++	++	++	++	--	++	--	--	.	-	++	--
13	--	-	++	.	--	--	.	--	++	--	++	--
14	--	--	--	--	--	--	--	--	--	--	--	--
16	--	--	--	--	--	--	--	--	--	--	--	--
17	--	--	++	.	--	--	--	--	--	--	++	--
18	--	--	++	.	--	--	--	--	--	--	.	--
19	--	--	--	--	--	--	--	--	--	--	--	--
20	--	--	--	--	--	--	--	--	--	--	--	--
21	.	.	-	-	-	+	.	.	--	.	--	.
22	-	--	++	.	--	--	-	--	+	--	++	.
23	-	--	--	--	--	--	--	--	--	--	--	--
24	--	--	--	--	--	--	--	--	--	--	--	--
25	--	--	--	--	--	--	--	--	--	--	--	--
26	--	++	--	--	++	--	++	--	++	++	+	--
29	-	+	.	-	.	--	++	--	++	++	++	+
30	++	++	-	-	++	+	++	-	-	++	+	++
32	+	+	--	--	+	+	++	--	--	+	--	++
33	--	-	--	--	-	-	--	--	--	--	.	.
34	.	.	--	--	--	-	++	--	--	.	--	.
35	++	++	++	++	++	++	++	++	++	++	++	++
36	--	--	--	--	--	--	--	--	--	--	--	--
37	++	++	++	++	--	++	--	--	.	.	++	--
39	--	--	--	-	--	--	--	--	--	--	-	--
40	.	.	+	++	--	+	--	--	-	--	++	--
41	--	--	.	-	--	--	--	--	-	--	-	--
42	--	--	--	--	--	--	--	--	--	--	--	--
43	--	--	--	--	--	--	--	--	--	-	--	--
44	--	--	--	--	--	--	--	--	--	--	--	--
47	++	++	++	++	++	++	++	++	++	++	++	++
49	--	--	.	.	--	--	--	-	--	--	-	--
52	++	++	.	-	++	--	++	.	-	++	+	++
53	--	--	--	--	--	++	--	--	.	--	--	--
54	--	--	--	--	--	--	--	--	--	++	--	++

with the various body patterns for the most probable solution found (with 12 distinct body patterns). Figure 4 illustrates representative examples of each cluster. Table 2 describes the images together with differences from them and the prototypical body pattern (i.e. the pattern with the highest probability of being expressed for each body pattern).

A detailed investigation of the relationship between the body pattern and the context is beyond the scope of this article (and the subject of a document in preparation), but one result is relevant. This is the strong relationship between the body pattern displayed, and the context the animal was recorded in. Given previously unseen body patterns, the correct context could be predicted *ca.* 60% of the time, and this was approximately constant across all contexts (Baddeley *et al.* 2003).

#### 4. DISCUSSION

Hanlon & Messenger (1988) provide a detailed list of the chromatic and other components used by juvenile *S. officinalis* and, following Packard & Hochberg (1977), suggest how their expression is coordinated to produce several body patterns (figure 1). We have given an objective and quantitative method for specifying the number of distinct body patterns (signalling states) used by the cuttlefish *S. officinalis*. Finding such structure in data is notoriously difficult for human observers, who often detect structure when there is none, or, especially in high-dimensional data, overlook strong regularities in the less perceptually salient features. The AUTOCLASS method has a successful history in identifying distinct clusters in the infrared spectra of stars (Goebel *et al.* 1989), and in the

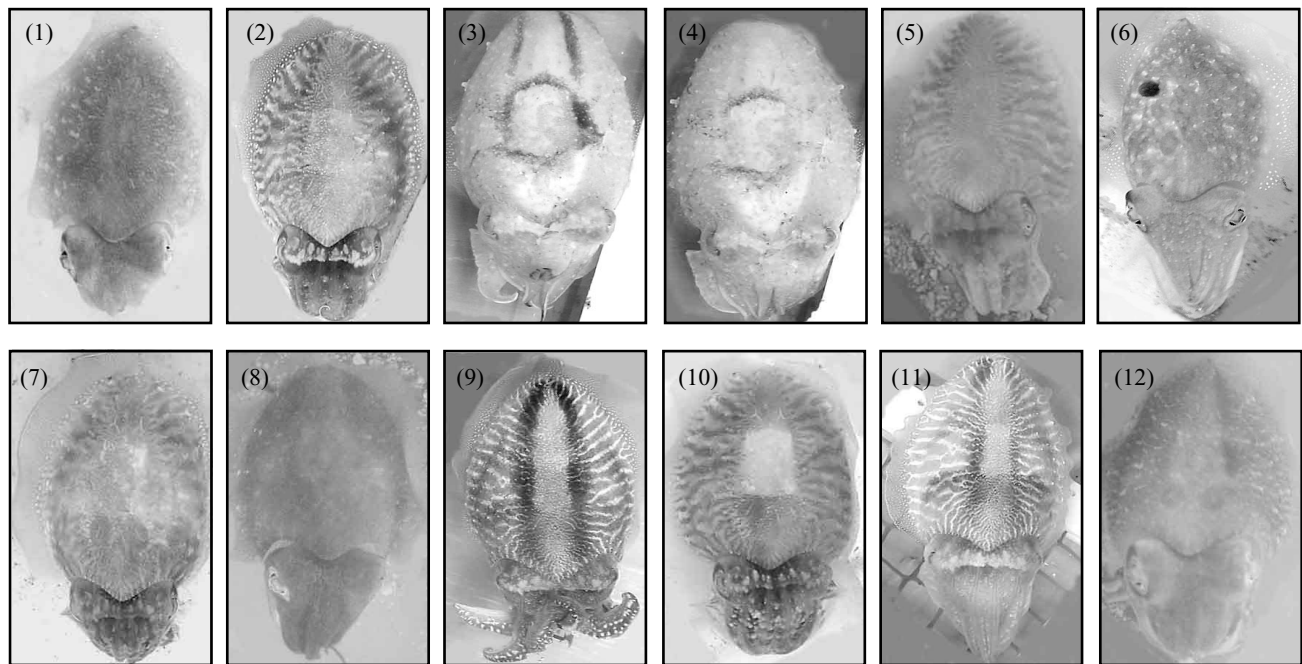


Figure 4. Images typical of the 12 clusters of body patterns produced by *Sepia officinalis* juveniles in our image database. They are ordered by decreasing probability of expression (figure 5). Each image was chosen as a compromise between closeness of the most obvious visual features to the cluster centre and image quality. Tables 1 and 2 give further information on the patterns.

Table 2. Description of the 12 identified body patterns, together with any differences between the prototypical pattern, and the representative image shown in figure 4.

cluster number	description	comments
1	uniform light or light mottle	component 40 missing
2	zebra stripes + mottle	additional component 13
3	blanched with white square and triangle	additional component 41
4	blanched and papillate	additional components 17 and 18
5	zebra stripes	additional components 12 and 37
6	deimatic	additional component 34; component 21 is presented unilaterally
7	light mottle	additional component 37, missing component 9
8	uniform	additional component 52
9	intense zebra	—
10	zebra stripes + white square	additional component 42, missing component 54
11	zebra stripes + headbar	additional components 1 and 2, missing component 33
12	eye ring	missing component 33 additional components 1 and 2

discovery of classes of proteins, introns and other patterns in DNA/protein-sequence data (Cheeseman & Stutz 1996). Although we are unable to make a precise comparison between the 12 body patterns that we identify (figure 4) and those of Hanlon and Messenger (figure 1), some of their similarities and differences can be seen in the descriptions given in table 2.

To tackle the problem of identifying the number of distinct signalling states of a cuttlefish, we endeavoured to estimate the number of their visual signal generators. We give explicit probabilities for the number of states being a given number. The most probable interpretation was that there were 12 states, but solutions based on 13 and 14 clusters also had high probability. Although the result is compatible with the 13 body patterns described by Hanlon & Messenger (1988), we can rule out the much larger number of body patterns claimed for other cephalopods

(Hanlon & Messenger 1996). This is not to say that these animals do not have these much larger signalling repertoires, but given no objective means of associating probabilities with different interpretations, it is impossible to compare estimates of the number of different signalling states derived from different laboratories. The AUTOCLASS system is freely available, and simple to use. It would be interesting to measure the number of signalling states in other species objectively and, in particular, find out if there is any systematic relationship between this number and the average group size. We are currently working on the classification of several other *Sepia* species (e.g. *S. lycidas* and *S. pharaonis*).

The 12 different body patterns were displayed with different probabilities in our dataset (figure 5). This is relevant for two reasons: (i) the distribution of frequencies can have distinctive patterns in many communication

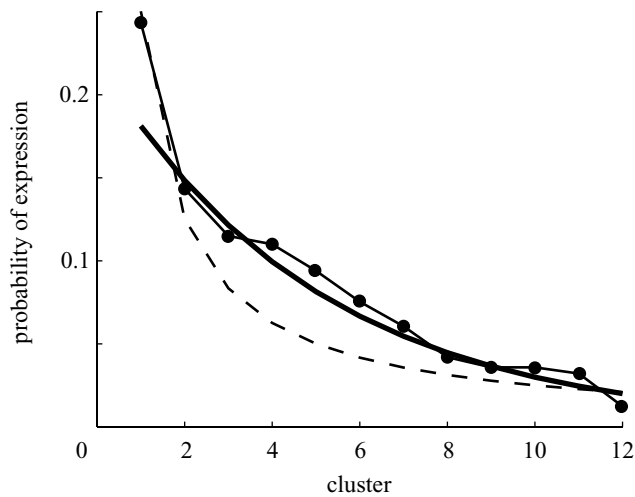


Figure 5. The probabilities of the 12 different clusters (connected black dots). Shown for comparison are the probabilities predicted if these data obeyed Zipf's (1935) law (dashed line), or an exponential decay ( $\alpha = 0.2$ ). As can be seen, apart from the first cluster, the data are much better described as a simple exponential decay rather than the Zipf's law, which characterizes human communication systems.

systems; and (ii) this probability distribution places an upper bound on the channel capacity of cuttlefish visual signals. For natural languages this distribution has a power law form (known as Zipf's (1935) law). Dolphin communication may also follow Zipf's law (McCowan *et al.* 1999), but the expression of cuttlefish body patterns in our dataset does not follow Zipf's law. Instead, an exponential distribution of frequencies provides a much better fit—which is not typical of human communication systems. The distribution does, however, give an estimate of the entropy of the cuttlefish signals, when viewed as a communication channel. This places an upper bound on the information (based on zero noise) transmission capacity of the signals of 3.4 bits per visual signal, far less than that theoretically possible with such a potentially rich medium of communication.

The most similar work to this study on animal communication is by McCowan (1995) who used a clustering algorithm (*k*-means) to categorize the whistles of bottlenose dolphins (*Tursiops truncatus*). There are several problems with this method. These include the non-reproducibility of the results of the *k*-means algorithm (which depends critically on initial conditions). More importantly, as acknowledged by McCowan, no rigorous method was used to determine the number of categories. The methods that we use can be thought of as a substantially improved version of *k*-means clustering. We search for the best solution, rather than one dictated by initial conditions, and utilize a principled method to estimate the number of body patterns displayed by *S. officinalis*.

The present study simplifies the analysis by using binary component expressions. This was done because to measure accurately the surface reflectance of the skin for all the cuttlefish images would require accurate calibration of the digital camera, and would require a very labour-intensive process of measuring the reflectivity for each of the (34) chromatic components. For the remaining non-chromatic

components (textural, postural and locomotor), it is again unclear how one would objectively assign real valued numbers to their 'degree of expression', and statistical models that combine both discrete and continuous measurements are much more difficult to deal with. Despite this, we are currently developing a computerized system that will semi-automatically allow the measuring of chromatic component reflectances. The analogue level of expression almost certainly contains information, and utilizing this should provide greater insight into the use of body patterns in cuttlefish communication, and the application of modern clustering methods such as mixtures of factor analysers (Ghahramani & Hinton 1996).

A further limitation of this study is that animals were only filmed in captivity, and observed in a limited number of distinct behavioural contexts. This raises the issue that this analysis was conducted on multiple different situations, and although many of the body patterns were used for social communication, many more were probably camouflage behaviour. This was done because the methods, and the calculation of channel capacity, require a large amount of data. Despite this, combining across contexts, and the bias in sampling from the different situations, means that the exact information content of the signals and the deviation from Zipf's law should be treated with some caution.

Although the visual nature of the cuttlefish signals simplifies some of the analyses, there is no reason why similar statistical methods cannot be applied to other areas of communication. Clustering, using mixtures of Gaussians rather than the Bernoullis, is the standard recognition model used for artificial speech recognition models. Given that auditory signals are appropriately pre-processed to remove the ambiguity of pitch (a cepstrum representation of the signal is the standard method in artificial speech recognition; Noll (1967)), then the number of distinct vocal signals could be estimated for animals whose communication is sound based. The method requires only (i) that there are large numbers of examples of the signals; (ii) a method for quantifying the different signals; and (iii) a set of statistical models corresponding to the different signalling hypotheses: mixtures of Gaussians or factor analysers would be appropriate. Given this, any communication system could potentially be investigated.

## 5. CONCLUSIONS

Our aim was to analyse the body patterns of the common cuttlefish, *S. officinalis*, by associating probabilistic models with different proposals as to the statistical structure of the signals used. We used statistical model comparison methods to compare these models. We propose a particular quantitative model based on mixtures as a way of estimating the number of distinct body patterns displayed by this species. The computer software we used is widely available, and has been used successfully in such diverse areas as astronomy and DNA-sequence analysis. The method does not produce an absolute answer to the number of clusters, but a probability for each possible interpretation. The number of distinct body patterns identified was similar to that found by Hanlon & Messenger (1988) in cuttlefish, using simple observation.

Importantly, the methodology used here allows

quantitative investigation of such claims that, for example, the social squid, *Sepioteuthis sepioidea*, can express as many as 35 different displays (Moynihan & Rodaniche 1982; see also Hanlon & Messenger 1996). Without a statistical method, and when the classifications are performed by eye (a task that is remarkably difficult), such claims can be difficult to interpret. An automatic system helps systematic studies of animal signalling repertoires, and in relating them to their social and physical environment.

Finally, two aspects of these data are relevant to the claim that cuttlefish signals constitute a language (Moynihan & Rodaniche 1982). First, the distribution of signals is not distributed in the same way as that found in human and claimed for dolphin signals (McCowan *et al.* 1999); it does not obey the characteristic Zipf's law. Although this is not a definitive characteristic of a language, it is a characteristic of most human languages. One possible reason why cuttlefish signals do not resemble natural languages is that many patterns are used primarily for camouflage.

More importantly, with 54 different components, it is potentially possible that each signal could communicate up to 54 bits of information, more than enough for very sophisticated communication. By contrast, rather than 54 bits, because of redundancy, we found an estimate of 3.4 bits per signal. Although English only has a channel capacity of *ca.* 1.5 bits per letter, the letters can be added to provide an effectively infinite channel capacity for communication. By contrast, the cuttlefish, even with the capability of combining multiple different components to increase channel capacity, only signals just over 3 bits. Although small, for a relatively non-social species, this capacity may be all that *S. officinalis* requires.

We thank T. Cronin, R. Hanlon and N. Shashar for much advice and encouragement, and J. Messenger for extensive and very helpful comments on the manuscript. J. Rundle and R. Williamson of Plymouth Marine Laboratory supplied and gave essential advice on maintaining cuttlefish. E. Kelman gave invaluable help in the laboratory. Financial support was from the Association for the Study of Animal Behaviour and Enterprise Ireland to A.C.

## REFERENCES

- Baddeley, R. J., Crook, A. C. & Osorio, D. 2003 Relating cuttlefish signalling to its environment: insights from non-linear regression. (In preparation.)
- Bishop, C. M. 1995 *Neural networks for pattern recognition*. Oxford: Clarendon Press.
- Cheeseman, P. & Stutz, J. 1996 Bayesian classification (AUTOCLASS): theory and results. In *Advances in knowledge discovery and data mining* (ed. U. M. Fayyad), pp. 153–180. Boston, MA: AAAI/MIT Press.
- Ghahramani, Z. & Hinton, G. E. 1996 Parameter estimation for linear dynamical systems, University of Toronto technical report CRG-TR-96-2.
- Goebel, J., Volk, K., Walker, H., Gerbault, F., Cheeseman, P., Self, M., Stutz, J. & Taylor, W. 1989 A Bayesian classification of the IRAS LRS atlas. *Astron. Astrophys.* **222**, L5–L8.
- Hanlon, R. T. & Messenger, J. B. 1988 Adaptive coloration in young cuttlefish (*Sepia officinalis* L.): the morphology and development of body patterns and their relation to behaviour. *Phil. Trans. R. Soc. Lond. B* **320**, 437–487.
- Hanlon, R. T. & Messenger, J. B. 1996 *Cephalopod behaviour*. Cambridge University Press.
- Holmes, W. 1940 The colour changes and colour patterns of *Sepia officinalis* L. *Proc. Zool. Soc. Lond. A* **110**, 2–35.
- McCowan, B. 1995 A new quantitative technique for categorising whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, *Tursiops truncatus*). *Ethology* **100**, 177–193.
- McCowan, B., Hanser, S. F. & Doyle, L. R. 1999 Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Anim. Behav.* **57**, 409–419.
- Marshall, N. J. & Messenger, J. B. 1996 Colour-blind camouflage. *Nature* **382**, 408–409.
- Maynard Smith, J. & Harper, D. G. C. 1995 Animal signals: models and terminology. *J. Theor. Biol.* **177**, 305–311.
- Messenger, J. B. 1979 The eyes and skin of *Octopus*: compensating for sensory deficiencies. *Endeavour* **3**, 92–98.
- Moynihan, M. 1983a Notes on the behavior of *Euprymna scolopes* (Cephalopoda: Sepiolidae). *Behaviour* **85**, 25–41.
- Moynihan, M. 1983b Notes on the behavior of *Idiosepius pygmaeus* (Cephalopoda: Idiosepiidae). *Behaviour* **85**, 42–57.
- Moynihan, M. & Rodaniche, A. F. 1977 Communication, crypsis, and mimicry amongst cephalopods. In *How animals communicate* (ed. T. A. Sebeok), pp. 293–302. Bloomington, IN: Indiana University Press.
- Moynihan, M. & Rodaniche, A. F. 1982 The behaviour and natural history of the Caribbean reef squid, *Sepioteuthis sepioidea*, with a consideration of social, signal, and defensive patterns for difficult and dangerous environments. *Adv. Ethol.* **25**, 1–151.
- Noll, A. M. 1967 Cepstrum pitch determination. *J. Acoust. Soc. Am.* **41**, 293–309.
- Packard, A. & Hochberg, F. G. 1977 Skin patterning in *Octopus* and other genera. *Symp. Zool. Soc. Lond.* **38**, 191–231.
- Ripley, B. D. 1996 *Pattern recognition and neural networks*. Cambridge University Press.
- Roper, C. F. E. & Hochberg, F. G. 1988 Behavior and systematics of cephalopods from Lizard Island, Australia based on color and body patterns. *Malacologia* **29**, 153–193.
- Roweis, S. & Ghahramani, Z. 1999 A unifying review of linear Gaussian models. *Neur. Comput.* **1**, 305–345.
- Stutz, J. & Cheeseman, P. 1995 AUTOCLASS—a Bayesian approach to classification. In *Maximum entropy and Bayesian methods* (ed. J. Skilling & S. Sibisi). Dordrecht: Kluwer.
- Wallace, C. S. & Dowe, D. L. 2000 MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. *Stat. Comput.* **10**, 73–83.
- Zipf, G. K. 1935 *Psycho-biology of languages*. Boston, MA: Houghton-Mifflin.