Hierarchic Social Entropy: An Information Theoretic Measure of Robot Group Diversity

Tucker Balch
The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213-3891

Abstract

As research expands in multiagent intelligent systems, investigators need new tools for evaluating the artificial societies they study. It is impossible, for example, to correlate heterogeneity with performance in multiagent robotics without a quantitative metric of diversity. Currently diversity is evaluated on a bipolar scale with systems classified as either *heterogeneous* or *homogeneous*, depending on whether any of the agents differ. Unfortunately, this labeling doesn't tell us much about the *extent* of diversity in heterogeneous teams. How can it be determined if one system is more or less diverse than another? Heterogeneity must be evaluated on a continuous scale to enable substantive comparisons between systems. To enable these types of comparisons, we introduce: *(1)* a continuous measure of robot behavioral difference, and *(2)* hierarchic social entropy, an application of Shannon's information entropy metric to robotic groups that provides a continuous, quantitative measure of robot team diversity. The metric captures important components of the meaning of diversity, including the number and size of behavioral groups in a society and the extent to which agents differ. The utility of the metrics is demonstrated in the experimental evaluation of multirobot soccer and multirobot foraging teams.

To appear, Autonomous Robots, vol 8, no 3, July, 2000.

1 Introduction

Heterogeneous systems are a growing focus of robotics research [FM97, GM97, Par94, Bal99]. Presently, diversity in these systems is evaluated on a bipolar scale; systems are classified as either *heterogeneous* or *homogeneous* depending on whether any of the agents differ. This view is limiting because it does not permit a quantitative comparison of heterogeneous systems. A principled study of diversity requires a quantitative metric. Such a metric would enable the investigation of issues like the impact of diversity on performance, and conversely, the impact of other task factors on diversity. To address this, we propose *social entropy* (computed using Shannon's information entropy formulation [Sha49]) as an appropriate measure of diversity in robot teams.

We do not claim diversity is always desirable. In fact, experiments described in this article show that for some tasks homogeneous teams perform better than diverse teams. The goal of this work, rather, is to enable the investigation of the origins of diversity in learning teams and to help developers build appropriately diverse robot teams for the tasks and environments in which they operate. A quantitative metric is crucial to the investigation of these issues.

While the results of this work are applicable in other areas, we focus specifically on evaluating diversity in teams of mechanically similar agents that use reinforcement learning to develop behavioral policies. Behavior is an especially interesting dimension of diversity in learning systems since as they learn, agents choose the degree of diversity in their society on their own.

Contributions of this work include *(1) simple social entropy,* a novel application of Shannon's information entropy to the measurement of diversity in robot groups, *(2) behavioral difference,* a continuous, quantitative measure of difference between individual robots, *(3) hierarchic social entropy,* a metric combining simple social entropy with behavioral difference to provide a continuous scale of diversity, capturing even minute differences between societies, and *(4)* example applications of these metrics in the evaluation of actual robot systems.

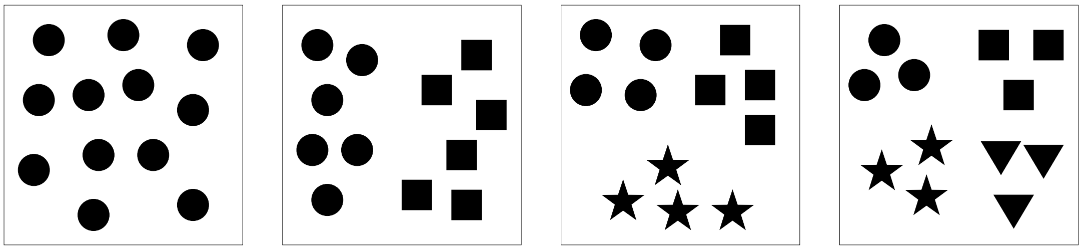
This article begins with an examination of the meaning of diversity for multiagent robotic systems. That discussion is followed by a presentation of information entropy, an explanation of how it can be applied to evaluate robot team diversity and a justification for its application in this domain. Classification and agent difference are addressed after that. Next a formulation of hierarchic social entropy (that combines the concepts of agent difference, classification and entropy) is introduced. Finally, the utility of simple social entropy and hierarchic social entropy is demonstrated in example applications, including experiments in robot soccer and multirobot foraging.

2 The meaning of diversity

What does *diverse* mean? Webster [MW89] provides the following definition:

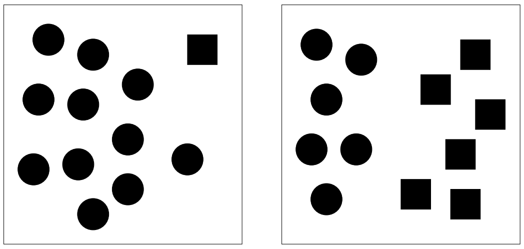
di.verse *adj* **1:** differing from one another: unlike **2:** composed of distinct or unlike elements or qualities

Clearly, difference plays a key role in the meaning of diversity. In fact, an important challenge in evaluating robot societal diversity is determining whether agents are alike or unlike. Assume for



a **b** **c** **d**

Figure 1: Several collections of shapes. The number of homogeneous subsets in each collection grows from one in **a** to four in **d**. Should measured diversity depend on the number of homogeneous subsets?



a **b**

Figure 2: In both of these groups there are the same number of shapes and the same number of homogeneous subsets, but the proportion of elements in each subset is different.

now that any two agents are either alike (in the same behavioral subset) or not. (The degree of difference *is* important but that issue is addressed later.)

Now consider what *diverse* means for societies composed of several distinct behavioral subsets. To make the discussion more concrete, suppose the "society" under examination is a collection of four different shapes: circles, squares, triangles and stars. Figures 1 and 2 illustrate several sets of shapes as examples of ways the groupings can differ. The goal is to develop a quantitative metric that captures the meaning of diversity illustrated in these examples.

First, how should the number of distinct subsets in a society impact the measured diversity? Consider Figure 1: four sets of 12 shapes. Each set has a different number of homogeneous subsets; from one homogeneous subset in Figure 1a (all circles) to four in Figure 1d. This example suggests that the number of homogeneous subsets in a society is an important component of measured diversity.

Now consider Figure 2. Which group of shapes is more diverse? In both cases there are exactly 12 shapes and exactly two different types. In Figure 2a however, there is a much higher proportion of circles than in 2b where there is an equal number of circles and squares. This example suggests that the relative proportion of elements in each subset is an important component of diversity.

These examples highlight the fact that the *distribution* of the agents between homogeneous subsets is at the core of the meaning of diversity. In light of this observation, we make the following commitment: the measured diversity of a multiagent society depends on the number of homogeneous subsets it contains and the proportion of agents in each subset.

3 Simple social entropy

How should diversity be quantified? Shannon faced a similar problem when he sought to quantify the uncertainty, or randomness, of an information source [Sha49]. The uncertainty of an information source has important implications for communications systems, particularly with regard to the minimum bandwidth required to transmit error-free messages.

Interestingly, the properties Shannon sought in a measure of information uncertainty are also useful in the measurement of societal diversity [Sha49]. In fact, researchers in a number of disciplines have adopted information theoretic concepts of diversity. As an example, consider this passage from Wilson's book *The Diversity of Life* [Wil92]:

> Suppose that we encounter a fauna of butterflies consisting of 1 million individuals divided into 100 species. Say one of the species is extremely abundant, represented by 990,000 individuals, and each of the other species therefore comprises an average of about 100 individuals. One hundred species are present but, as we walk along the forest paths and across the fields, we encounter the abundant butterfly most of the time and each of the other species only rarely ... In a nearby locality we encounter a second butterfly fauna, comprising the same 100 species, but this time all are equally abundant, represented by 10,000 individuals each. This is a fauna of high equitability, in fact the highest possible. Intuitively we feel that the high-equitability fauna is the more diverse of the two, since each butterfly encountered in turn is less predictable and therefore gives us more information on average.

Wilson's view embraces the idea that societies with members equally distributed among subsets are the most diverse. It also suggests that diversity and information are closely related concepts. Information entropy is used in a number of related fields as well. It is used by ecologists as a means of evaluating species' diversity [LVW83, LW80, Mag88], by sociologists as a model of societal evolution [Bai90], and by taxonomists as a tool for evaluating classification methodologies [SS73, JS71].

Shannon's measure, *information entropy*, is easily adapted to suit the needs of a societal diversity metric [Sha49]. Bailey popularized the application of information entropy to the study of social groups in his book *Social Entropy Theory* [Bai90]. Although the formulation of simple social entropy for robot groups is somewhat different than Bailey's formulation for human societies, we adopt his term here. This section provides a mathematical basis for the calculation of simple social entropy in robot groups and illustrates why it is an appropriate measure of multiagent system diversity.

3.1 Mathematical formulation of simple social entropy

Before proceeding we must introduce some additional notation:

- $\mathcal{R}$ is a society of N agents with $\mathcal{R} = \{r_1, r_2, r_3 ... r_N\}$
- $\mathcal{C}$ is a classification of $\mathcal{R}$ into M possibly overlapping subsets.
- c_i is an individual subset of $\mathcal{C}$ with $\mathcal{C} = \{c_1, c_2, c_3 ... c_M\}$
- $p_i = \frac{|c_i|}{\sum_{j=1}^{M} |c_j|}$ is the proportion of agents in the ith subset; $\sum p_i = 1$.

In the last section we argued that the measured diversity of a system should reflect the number of groups in the system and the distribution of elements into those groups; diversity should therefore be a function of M and the p_is as defined above. Assume that a diversity metric exists and call it H. The diversity of a society partitioned into M homogeneous subsets is written $H(p_1, p_2, p_3, ..., p_M)$. So, for instance, the diversity of the group of blocks depicted in Figure 2a is $H\left(\frac{1}{12}, \frac{11}{12}\right)$, while the

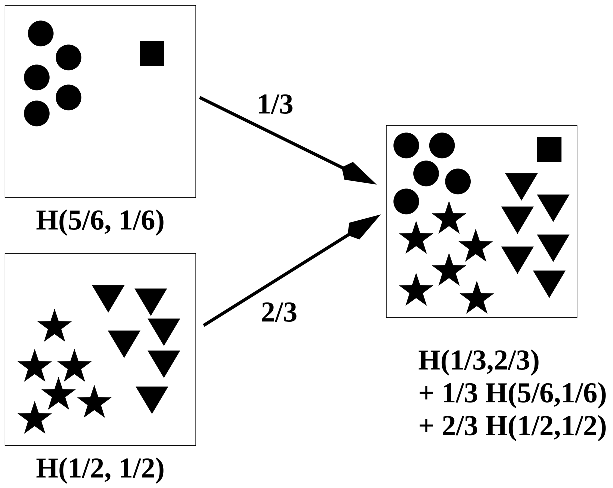


Figure 3: A new society (right) is generated by combining two others (left). The diversity of the new society is a weighted sum of the individual values of H for the subsets.

diversity for the group of blocks in Figure 2b is $H\left(\frac{1}{2},\frac{1}{2}\right)$. The diversity of a particular robot society $\mathcal{R}_a$ can also be expressed $H(\mathcal{R}_a)$.

Shannon prescribed three properties for a measure of information uncertainty [Sha49]. With slight changes in notation, they are equally appropriate for a measure of societal diversity:

Property 1 continuous: H should be continuous in the p_i.

Property 2 monotonic: If all the p_i are equal, (i.e. $p_i = \frac{1}{M}$), then H should be a monotonically increasing function of M. In other words, if there are an equal number of agents in each subset, more subsets implies greater diversity.

Property 3 recursive: If a multiagent society is defined as the combination of several disjoint sub-societies, H for the new society should be the weighted sum of the individual values of H for the subsets. This property is important for the analysis of recursively composed societies (e.g. [MAC97]).

The meaning of the requirement that H be recursive is illustrated in Figure 3. The two groups on the left are combined into a new society on the right. The groups on the left have diversities $H\left(\frac{5}{6},\frac{1}{6}\right)$ (top) and $H\left(\frac{1}{2},\frac{1}{2}\right)$ (bottom). The diversity of the new 18 element society is $H\left(\frac{5}{18},\frac{1}{18},\frac{6}{18},\frac{6}{18}\right)$. Because the sub-groups contribute $\frac{1}{3}$ and $\frac{2}{3}$ of the agents to new society, the recursive criteria requires:

$$H\left(\frac{5}{18},\frac{1}{18},\frac{6}{18},\frac{6}{18}\right) = \underbrace{H\left(\frac{1}{3},\frac{2}{3}\right)}_{\text{both groups}} + \underbrace{\frac{1}{3}H\left(\frac{5}{6},\frac{1}{6}\right)}_{\text{contribution of first group}} + \underbrace{\frac{2}{3}H\left(\frac{1}{2},\frac{1}{2}\right)}_{\text{contribution of second group}}$$

In general, for a society $\mathcal{R}_c$ composed of two societies, $\mathcal{R}_a$ and $\mathcal{R}_b$, the recursive criteria ensures that:

$$H(\mathcal{R}_c) = H(\alpha,\beta) + \alpha H(\mathcal{R}_a) + \beta H(\mathcal{R}_b)$$

where α is the proportion of agents in $\mathcal{R}_a$, β is the proportion of agents in $\mathcal{R}_b$ and $\alpha + \beta = 1$.

Meyer and McIntosh have developed an index of ethnic diversity used by *US News* and *USA Today* [MM92]. Their index measures the probability that two people chosen at random (with replacement) will differ along at least one ethnic dimension. The index value ranges from 0 to 1; greater diversity is indicated by a larger value. A value of zero applies to a population in which everyone is the same. If every person is different from every other person on at least one dimension, the value is maximized. In practice the index can never reach unity because an infinite number of p_is would be required. The metric has intuitive appeal, and may be of interest as a measure of multiagent social diversity. Using the notation introduced above, Meyer's metric can be written:

$$H_m(X) = 1 - \sum_{i=1}^{M} p_i^2 \tag{1}$$

H_m provides mathematical properties 1 and 2 but it does not provide for recursively defined societies (Property 3). Shannon's *information entropy*, however, meets all three criteria [Sha49]. The information entropy of a system X is given as[1]:

$$H(X) = -K \sum_{i=1}^{M} p_i \log_2(p_i) \tag{2}$$

where K is a positive constant. Because K merely amounts to the choice of a unit of measure, Shannon sets $K = 1$ [Sha49].

Equation 2 (with $K = 1$) is adopted for the measurement of multiagent societal diversity. $H(\mathcal{R}_a)$ is the *simple social entropy* of agent society $\mathcal{R}_a$.

In addition to Properties 1, 2 and 3, H has a number of additional properties that further substantiate it as an appropriate measure of diversity. First, as we would expect, H is minimized for homogeneous societies; these groups are the least diverse. Also, for heterogeneous groups H is maximized when there are an equal number of agents in each subset. More precisely:

Property 4: $H = 0$ if and only if all the p_i but one are zero. In other words H is minimized when the system is homogeneous. Otherwise H is positive.

Property 5: For a given M (number of homogeneous subsets), H is maximized when all the p_i are equal, i.e. $p_i = \frac{1}{M}$. This is the case when there are an equal number of agents in each subset.

Property 6: Any change toward equalization of the proportions $p_1, p_2, \ldots, p_M$ increases H. Thus if $p_1 < p_2$ and we increase p_1, decreasing p_2 an equal amount so that they are more nearly equal, H increases. An important implication is that there are no locally isolated maxima.

Even if these properties are desirable in a diversity metric, why choose information entropy over another function possessing the same properties? Because, as it turns out, information entropy

[1] $H(X)$ is used in coding theory as a lower-bound on the average number of bits required per symbol to send multi-symbol messages. The random variable X assumes discrete values in the set $\{x_1, x_2, x_3 \ldots x_M\}$ (the alphabet to be encoded) and p_i represents the probability that $\{X = x_i\}$.

(Equation 2) is the *only* function satisfying Properties 1, 2 and 3. Shannon proved this result using the mathematically equivalent properties he required of an information uncertainty metric [Sha49].

3.2 Example evaluations

Consider the simple social entropy of a heterogeneous group composed of one square and three star shapes. The society consists of four elements, $\mathcal{R} = \{r_1, r_2, r_3, r_4\}$. One element, r_4 (the square) is not equivalent to the others so there are two homogeneous subsets, $C = \{c_1, c_2\}$, with $c_1 = \{r_1, r_2, r_3\}$ (the star class) and $c_2 = \{r_4\}$ (the square class). Then,

$$\begin{aligned}
p_1 &= \frac{3}{4} = .75 \\
p_2 &= \frac{1}{4} = .25 \\
H(\mathcal{R}) &= -\sum_{i=1}^{2} p_i \log_2(p_i) \\
&= -((p_1 \log_2(p_1)) + (p_2 \log_2(p_2))) \\
&= -((.75 \log_2(.75)) + (.25 \log_2(.25))) \\
&= .811
\end{aligned}$$

Next the simple social entropy of a homogeneous group is evaluated. The group consists of elements $\mathcal{R} = \{r_1, r_2, r_3, r_4\}$. Homogeneity implies there is only one class, so $\mathcal{C} = \{c_1\}$, and $c_1 = \{r_1, r_2, r_3, r_4\}$. Then:

$$\begin{aligned}
p_1 &= 1 \\
H(\mathcal{R}) &= -\sum_{i=1}^{1} p_i \log_2(p_i) \\
&= -(p_1 \log_2(p_1)) \\
&= -(1 \log_2(1)) \\
&= 0
\end{aligned}$$

The entropy of a number of other example systems is given in Figure 4.

4 Limitations of simple social entropy

A potential limitation of simple social entropy as a diversity metric is the loss of information incurred when diversity is summarized in a single number. There are perhaps an infinite number of societies matching any particular value of diversity. Figure 5 for example, illustrates two very different societies whose entropy differs by less than 0.01.

A single number does not tell us how many classes of agents there are or how many agents in each class. This loss of information occurs whenever any characteristic of a multi-dimensional system is described as a single value. Such measurements are useful, however, because they enable

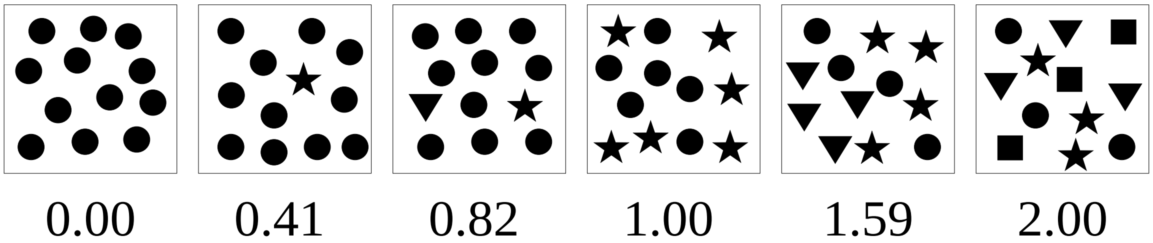


Figure 4: A spectrum of diversity. In the diagram above, each of the six squares encloses a multiagent system, from least diverse (homogeneous) on the left, to most diverse (most heterogeneous) on the right. The *simple social entropy*, a qualitative measure of diversity, is listed underneath each system.

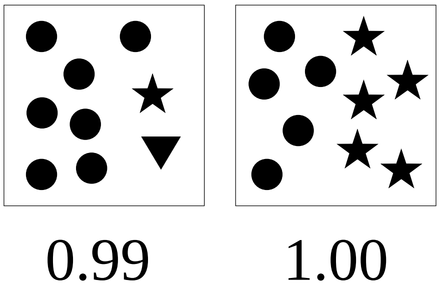


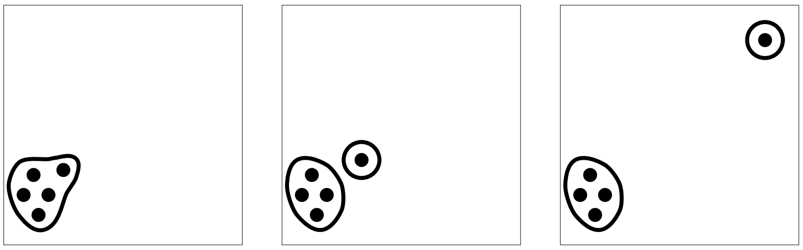
Figure 5: Two very different systems have similar entropy.

generalization and comparison. A thermometer, for example, does not reveal the position and velocity of every molecule in the environment but it does enable us to select our wardrobe for the day.

A more serious limitation of simple social entropy concerns its lack of sensitivity to the degree of difference between agents. Suppose, for example, we are evaluating the diversity of a number of agents distributed in a two-dimensional space (the dimensions may represent aspects of behavior or perhaps morphological axes). Agents that are close to one another are grouped in the same class. Figure 6 illustrates.

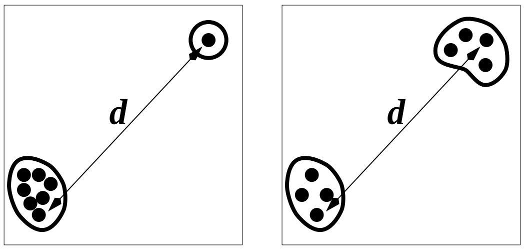
The figure shows three systems. In each system, the four elements in the lower left remain unchanged, but from 6a to 6c a fifth agent appears in several locations progressively more distant from the others. In Figure 6a it is just close enough to be classified with the others, while in 6b it is just far enough away to be placed in a separate category. The simple social entropy metric cannot differentiate between the distribution of agents in 6b and 6c because there is no difference in the number and size of the subsets. Also, the entropy measure finds a greater difference between the systems in 6a and 6b than between those in 6b and 6c.

One potential solution is to consider the maximum difference between agents as an additional component of diversity; e.g. the distance d in Figure 7. In the biological taxonomy literature d is referred to as *maximum taxonomic distance*. Taxonomic distance is useful, but as Figure 7 illustrates, it cannot serve as the only measure of diversity. This example shows two societies: one society with most of the agents classified together, but one "outlier" at distance d (7a); and another society with two equally sized subsets separated by the same distance (7b). Both of these systems have the same maximum difference but quite different distributions of agents into subsets. Taxonomic difference



a b c

Figure 6: One difficulty in the analysis of diversity. Dots representing agents are plotted in a two-dimensional space. Lines enclose agents grouped in the same class. The entropy metric cannot distinguish between the systems illustrated in **b** and **c**.



a b

Figure 7: Maximum taxonomic distance is a useful metric, but it does not account for the distribution of elements in the space.

captures the greatest difference between agents in the society but ignores the distribution of agents in the space. In the next two sections we show how a more comprehensive metric can be developed that reflects the extent of difference between the agents in a society.

5 Classification and clustering

The discussion of diversity left open the question of how agents are classified into subsets. It was assumed that any two agents are either alike (in the same subset) or unlike. In actuality, the robotic agents to be classified are distributed in a multi-dimensional space where the dimensions correspond to components of behavior and difference corresponds to the distance between agents in the space. Difference between agents is likely to vary along a continuous spectrum instead of in the binary manner assumed previously.

The limitations of simple social entropy discussed in Section 4 suggest that the diversity calculation would be improved if consideration were given to the spatial structure of the system. Here "spatial structure" refers to the distribution of elements in the classification space. In other words, the "clumpiness" of the system and the distribution of the clumps in the space are important.

The challenge of finding and characterizing clumps or clusters of elements distributed in a continuous multi-dimensional space is exactly the problem faced by biologists in building and using taxonomic systems. In the case of biology the dimensions of the space represent aspects of morphology or behavior that distinguish one organism from another. In this research the dimensions are the

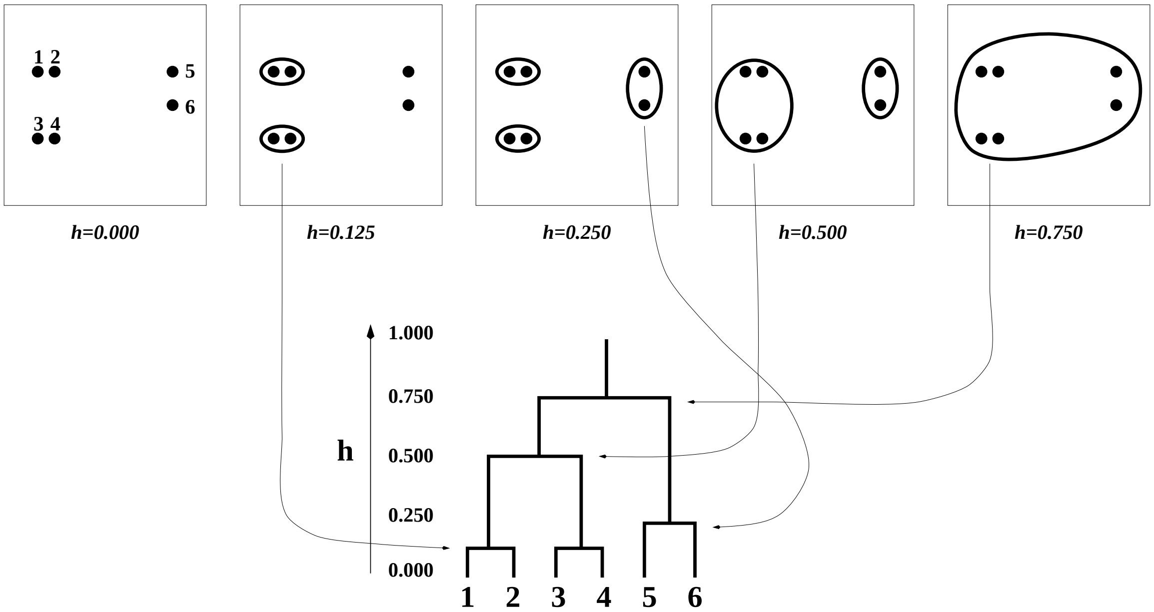


Figure 8: Example classification using numerical techniques. The top row shows how the system is clustered at several levels, parameterized by taxonomic level h (h is distinct from information entropy H). The classification is summarized in a taxonomic tree, or dendrogram (bottom). Strong similarities between elements are indicated by grouping near the bottom of the dendrogram; weaker similarities between groups are reflected in converging branches at higher levels.

components of behavior that distinguish one robot from another.

The aims of taxonomic classification are distinct from other types of classification in that one goal is to arrange the elements in a hierarchy reflecting their distribution in the classification space. Conversely, many classification tasks only require a simple partitioning of the space (e.g. categorizing e-mail into folders). Taxonomic trees (the end result of the taxonomic classification process, e.g. Figure 8) are potentially more useful in the analysis of diversity than simple partitionings because they provide more information about the society's spatial structure.

Biology offers a rich literature addressing this problem. In fact, an entire field — *numerical taxonomy* — is devoted to ordering organisms hierarchically using principled numerical techniques [SS73, JS71]. Many of the approaches in numerical taxonomy are directly applicable to the problem of robot classification. They include mechanisms for building and analyzing classification structures (e.g. taxonomic trees) and for identifying organisms on the basis of these structures.

5.1 Tools from numerical taxonomy

Figure 8 provides an example of the numerical taxonomic approach. Six elements (they could be organisms, species or robots) are distributed about a two-dimensional space. The location of each element in the space is determined by the value of each trait (e.g. tail length, weight, etc.) used in the classification. Each trait corresponds to a dimension in the classification space. The goal is to build a taxonomic tree that reflects the spatial distribution of elements in the system: closely related elements should be classified together at the bottom; similarities between groups are indicated as

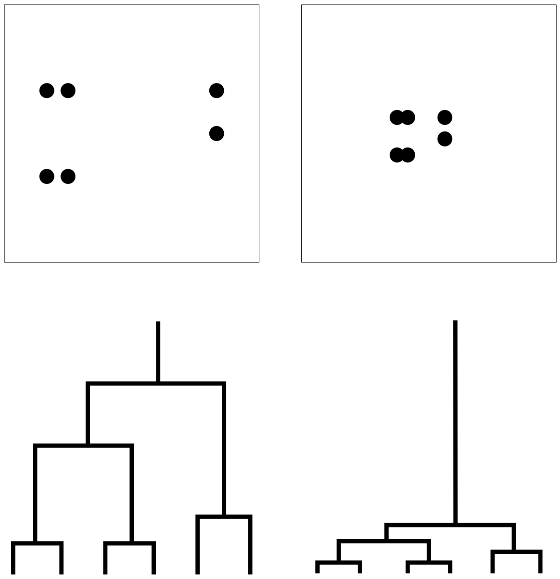


Figure 9: The branching structure of the dendrograms for these two societies is the same. However, the more compact distribution of elements in the system on the upper right is reflected in the branches being compressed towards the bottom of the corresponding dendrogram (lower right).

the branches converge at higher levels. These relations are expressed graphically in a *dendrogram* (Figure 8, bottom).

Techniques from numerical taxonomy address the problem of how to classify organisms, or groups of organisms, at various levels. At the lowest level in biological classification for instance, humans and gorillas are more likely to be classified together than, say, humans and dogs. But at a higher level, primates are in fact grouped with canines in the class *mammalia*. Dendrograms provide an orderly hierarchic view of the these classifications. While dendrograms *per se* are not necessary for the evaluation of diversity, they are useful visualization tools and their construction provides clues for the evaluation of overall societal diversity.

Dendrograms are constructed using a clustering algorithm parameterized by h, the maximum difference allowed between elements in the same subset. The notation $D(a, b)$ is used to refer to the difference between the elements a and b. In most applications the difference metric is normalized so that taxonomic distance between any two elements varies between 0 and 1. When $h = 1$ all elements are grouped together in one cluster (see the cluster at the top right in Figure 8 for example). As h is reduced from 1 down to 0 cluster boundaries change; the number of subsets increases as they split into smaller clusters. The splits are reflected as branches in the dendrogram. Finally, when $h = 0$ each element is a separate cluster; a "leaf" at the bottom of the dendrogram "tree."

Dendrograms can reveal subtle differences in societal structure. Figure 9 for example, shows two societies with the same relative arrangement of elements, but one grouping is compact while the other is spread out over a larger area. The difference in scale is reflected in a compressed dendrogram for the spatially compact society (Figure 9 right). Can these differences be accounted for in the evaluation of diversity?

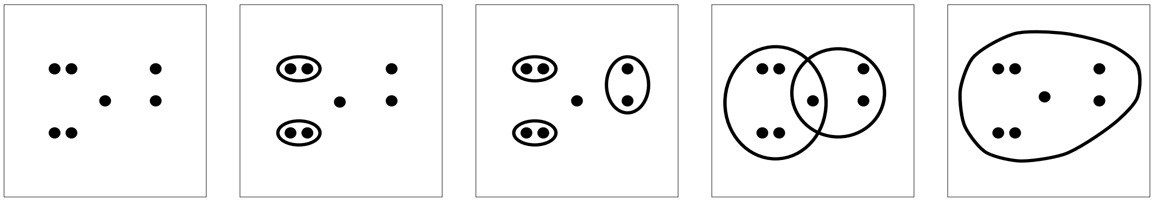


Figure 10: Example of hierarchic overlapping clustering. Clusters are enclosed in black lines.

Before addressing this, it is necessary to examine some of the details of clustering algorithms used to build a taxonomic tree. After that, the discussion returns to how these techniques can be used in the evaluation of diversity.

5.2 Clustering

Literally hundreds of clustering algorithms have been developed by researchers in a wide range of fields (Sneath and Sokal present a comprehensive taxonomy of clustering methods in [SS73]). One reason for the proliferation of techniques is the lack of generally agreed upon optimality criteria for evaluating the various methods. Jardine, for instance, suggests information-based metrics for biological clustering applications, but this may not be appropriate in all domains [JS71]. Because we are interested in the advantages of taxonomic representations of societal structure, the field of numerical taxonomy is an appropriate source of techniques for this research.

Most clustering methods used in numerical taxonomy are *hierarchic*. In *hierarchic* classifications any member of a lower ranking taxon is also a member of a higher ranked taxon. *Nonhierarchic* classifications do not exhibit ranks in which subsidiary taxa become members of larger more inclusive taxa. For traditional biological taxonomy, hierarchic classifications are required [SS73].

Another important distinction between clustering algorithms is whether or not overlap is allowed between clusters. In a *nonoverlapping* method, taxa at any one rank are mutually exclusive; a member of one taxon cannot also be a member of a second taxon at the same rank. Nonoverlapping classifications must sometimes arbitrarily assign elements to one or another equally distant subset. By relaxing this constraint, *overlapping* methods allow membership in more than one taxa.

Overlapping clustering methods are characterized by the degree of overlap allowed. Overlap can be quantified as the diameter of overlap or as the number of elements in the overlapping region.

The C_u overlapping clustering method is used in this research [JS71]. C_u or *u-diametric clustering* methods permit the diameter of overlap between clusters at level h to be at most uh (in this work $u = 1$). A cluster at level h is a maximally linked set such that for all elements r_i and r_j in the cluster $D(r_i, r_j) \leq h$.

Reviewing the notation presented earlier, the society of N elements to be clustered is $\mathcal{R} = \{r_1, r_2, r_3, \ldots r_N\}$. The society will be divided into M possibly overlapping clusters $\mathcal{C} = \{c_1, c_2, c_3, \ldots c_M\}$.

The C_u algorithm for clustering at level h proceeds in the following sequence:

1. Initialize N clusters with $c_i = \{r_i\}$.
2. For each cluster c_i:
   (a) For each r_j (except r_i) in $\mathcal{R}$:
      i. If $(D(r_j, r_k) \leq h)$ for every r_k already in c_i) add element r_j to cluster c_i.
3. Discard redundant clusters.

An example society classified using C_u clustering is presented in Figure 10. The clusterings for several values of h are illustrated with h increasing from left to right. Notice in the fourth diagram that the element in the middle of the space is claimed by two clusters. This clustering technique is *hierarchic* because elements classified at one level, or value of h, are also members of higher level taxa. In addition, the taxa (clusters) become larger and more inclusive at higher levels.

The spatial extent of elements in a taxonomic space is a reflection of the degree of difference between agents. It has already been pointed out (in Section 4) that such differences are important in the evaluation of diversity, especially for distinguishing between societies with similar structure and numbers of elements but with differing spatial size.

Note that sensitivity to the degree of difference between elements in hierarchic clustering depends on h. Because h is a *parameter* of the clustering algorithm, it can be varied to examine clusterings at any scale. Hierarchic algorithms are, in effect, variable power clustering microscopes. For values of h near zero the tiniest difference between elements will cause them to be classified separately, while the clusterings at large values of h reveal societal structure at a macroscopic level. This feature is exploited in the development of a diversity measure sensitive to differences in the spatial size of societies.

6 Hierarchic social entropy

Now consider how tools from numerical taxonomy can be applied to the measurement of diversity. The discussion of hierarchic clustering algorithms above described how the number and size of clusters depend on h. But how is simple social entropy impacted by changes in h? Since the partitioning of a society is based on h the entropy also depends on it. An example of the relationship is illustrated in Figure 11. Entropy changes in discrete steps as h increases. Note that points where change occurs correspond to branch points in the dendrogram.

Compare the dendrograms and entropy plots of the two societies in Figure 12. As in the earlier example, the two groups have the same relative structure, but the society represented on the right is more compact, resulting in branching compressed towards the bottom of the tree. The difference in scale is also readily apparent in the plots of entropy. Entropy drops to zero much more quickly in the plot corresponding to the compact society. Because the value of simple entropy depends significantly on h when hierarchic clustering is used, we augment the notation to account for this:

$$H(\mathcal{R}, h) \quad = \quad H(\mathcal{R}) \text{ for the clustering of } \mathcal{R} \text{ at taxonomic level } h \tag{3}$$

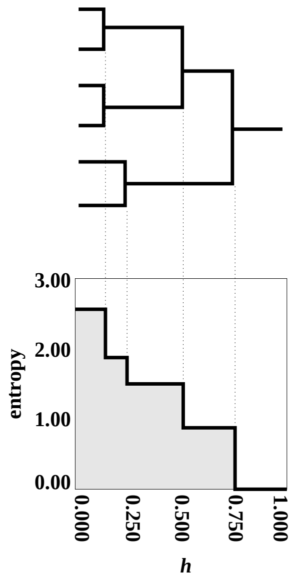


Figure 11: Entropy depends on h. Note that changes in entropy correspond to the branch points in the dendrogram. (For easier reference, the dendrogram is rotated 90 degrees.)

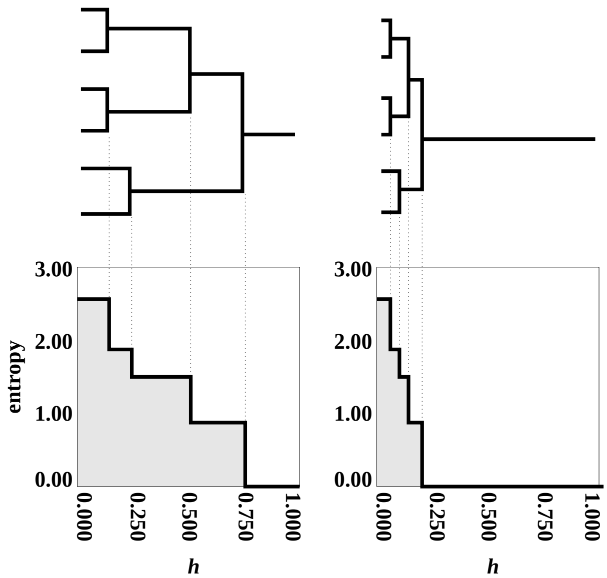


Figure 12: A comparison of entropy versus h for for two societies.

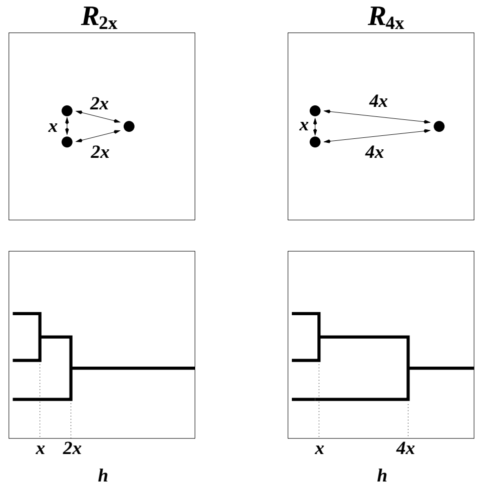


Figure 13: These two example systems are used to demonstrate how hierarchic social entropy can distinguish differences between societies regardless of scale. Spacing between the elements is parameterized by x (top). Because hierarchic entropy is scale invariant it can distinguish between the two societies regardless of the value of x. Dendrograms (bottom) illustrate the values of h where clusterings change.

H is a function of $\mathcal{R}$ *and* h because the classification of agents into subsets, and therefore the entropy, depends on them both. This highlights the fact that the entropy of a particular clustering is only a snapshot of the society's diversity. A comprehensive evaluation of diversity should account for clustering at all taxonomic levels. This is easily accomplished using the area under the entropy plot as a measure of diversity. This augmented metric, called *hierarchic social entropy*, is defined as:

$$S(\mathcal{R}) = \int_0^\infty H(\mathcal{R}, h)dh \tag{4}$$

where $\mathcal{R}$ is the robot society under evaluation, h is a parameter of the clustering algorithm indicating the maximum difference between any two agents in the same group and $H(\mathcal{R}, h)$ is the simple entropy of the society for the clustering at level h. Note that as $h \to \infty$ a point is reached where all elements are clustered in the same subset (the maximum taxonomic distance). $H(\mathcal{R}, h)$ drops to 0 at this point. In the behavioral difference measure used in this work, the maximum possible difference between elements is fixed at 1.0, so the upper limit of the integration is 1 rather than ∞ as in the general case.

6.1 Why hierarchic social entropy is a useful metric

Hierarchic social entropy is a continuous ratio measure; it has an absolute zero (when all elements are identical) and equal units. This enables a total ordering of societies on the basis of diversity. It also provides for quantitative results of the form "$\mathcal{R}_b$ is *twice* as diverse as $\mathcal{R}_a$." This is a significant advantage over the categorization of systems as simply "homogeneous" or "heterogeneous." Several other useful properties of hierarchic entropy are examined below.

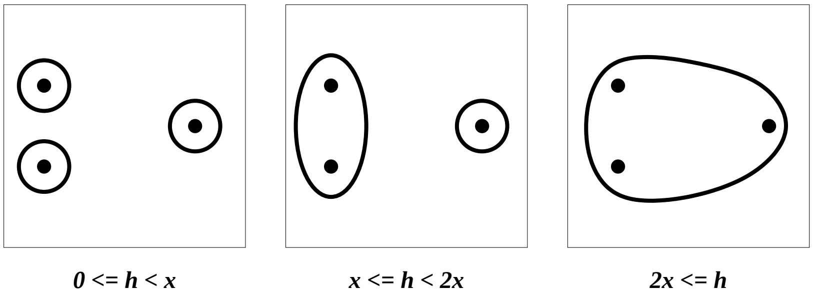


Figure 14: C_u clusterings of $\mathcal{R}_{2x}$ for different values of h.

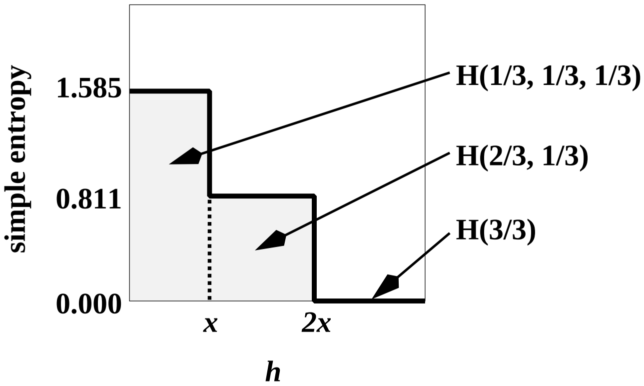


Figure 15: Simple entropy of $\mathcal{R}_{2x}$ as a function of h. There are three distinct regions with different values.

Hierarchic social entropy can distinguish differences between societies regardless of scale. Societies with infinitesimally small differences are compared as easily and precisely as systems spanning millions of units. This property is demonstrated with an example. Figure 13 illustrates two societies of three elements arranged in triangles. In both cases the two elements on the left are spaced a distance x apart. A third element is placed either $2x$ or $4x$ from the other elements in societies $\mathcal{R}_{2x}$ and $\mathcal{R}_{4x}$ respectively. Because hierarchic social entropy is scale invariant, it is able to distinguish between these two systems for all values of x. This will be demonstrated for $x = \frac{1}{1000000}$, for $x = 1000000$ and proven for all x.

First, observe that due to the spacing of the elements, there are three distinct C_u clusterings for each system (depending on h). For society $\mathcal{R}_{2x}$, the three elements are placed in three separate clusters when $0 \leq h < x$. Two clusters are present when $x \leq h < 2x$. Finally, all three elements are grouped together in one cluster when $2x \leq h$.

The groupings over all three ranges are illustrated in Figure 14 (groupings are similar for $\mathcal{R}_{4x}$ except the final clustering does not occur until $h \geq 4x$). The simple entropy for each clustering of $\mathcal{R}_{2x}$ is

$$\begin{aligned} H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) &= 1.585 \text{ for } 0 \leq h < x \\ H\left(\frac{2}{3}, \frac{1}{3}\right) &= 0.811 \text{ for } x \leq h < 2x \end{aligned}$$

$$H\left(\frac{3}{3}\right) \quad = \quad 0.000 \text{ for } 2x \leq h$$

These values and the regions over which they apply are illustrated in Figure 15. Similarly, the simple entropy for each clustering of $\mathcal{R}_{4x}$ is

$$\begin{aligned}
H\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) &= 1.585 \text{ for } 0 \leq h < x \\
H\left(\frac{2}{3},\frac{1}{3}\right) &= 0.811 \text{ for } x \leq h < 4x \\
H\left(\frac{3}{3}\right) &= 0.000 \text{ for } 4x \leq h
\end{aligned}$$

Now, suppose $x = \frac{1}{1000000}$. Can hierarchic entropy distinguish between these two systems? First we calculate the hierarchic entropy of society $\mathcal{R}_{2x}$. Recall the definition of hierarchic social entropy $S(\mathcal{R}, h)$ (Equation 4):

$$S(\mathcal{R}) \quad = \quad \int_0^\infty H(\mathcal{R}, h)dh$$

As was pointed out above, $H(\mathcal{R}, h)$ takes on distinct values over three regions depending on h. Therefore, the integral can be broken into parts corresponding to these regions:

$$\int_0^\infty H(\mathcal{R}_{2x}, h)dh \quad = \quad \int_0^x H(\mathcal{R}_{2x}, h)dh + \int_x^{2x} H(\mathcal{R}_{2x}, h)dh + \int_{2x}^\infty H(\mathcal{R}_{2x}, h)dh$$

Substituting 10^{-6} for x and the simple entropy values above for $H(\mathcal{R}_{2x}, h)$, we have

$$\begin{aligned}
\int_0^\infty H(\mathcal{R}_{2x}, h)dh &= \int_0^{10^{-6}} H\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) + \int_{10^{-6}}^{2\times10^{-6}} H\left(\frac{2}{3},\frac{1}{3}\right) + \int_{2\times10^{-6}}^\infty H\left(\frac{3}{3}\right) \\
&= \int_0^{10^{-6}} 1.585 + \int_{10^{-6}}^{2\times10^{-6}} 0.811 + \int_{2\times10^{-6}}^\infty 0.000 \\
&= 2.396 \times 10^{-6}
\end{aligned}$$

The hierarchic social entropy of $\mathcal{R}_{2x}$ is 2.396×10^{-6}. The calculation for $\mathcal{R}_{4x}$ is similar:

$$\begin{aligned}
\int_0^\infty H(\mathcal{R}_{4x}, h)dh &= \int_0^x H(\mathcal{R}_{4x}, h)dh + \int_x^{4x} H(\mathcal{R}_{4x}, h)dh + \int_{4x}^\infty H(\mathcal{R}_{4x}, h)dh \\
&= \int_0^{10^{-6}} H\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) + \int_{10^{-6}}^{4\times10^{-6}} H\left(\frac{2}{3},\frac{1}{3}\right) + \int_{4\times10^{-6}}^\infty H\left(\frac{3}{3}\right) \\
&= \int_0^{10^{-6}} 1.585 + \int_{10^{-6}}^{4\times10^{-6}} 0.811 + \int_{4\times10^{-6}}^\infty 0.000 \\
&= 4.018 \times 10^{-6}
\end{aligned}$$

For system $\mathcal{R}_{4x}$ we have $S(\mathcal{R}_{4x}) = 4.018 \times 10^{-6}$. Therefore when $x = 10^{-6}$ $S(\mathcal{R}_{2x}) < S(\mathcal{R}_{4x})$ and $\mathcal{R}_{4x}$ is 1.68 times more diverse than $\mathcal{R}_{2x}$.

What if $x = 1000000$? For $\mathcal{R}_{2x}$ the computation proceeds as follows:

$$\begin{aligned}
\int_0^\infty H(\mathcal{R}_{2x},h)dh &= \int_0^x H(\mathcal{R}_{2x},h)dh + \int_x^{2x} H(\mathcal{R}_{2x},h)dh + \int_{2x}^\infty H(\mathcal{R}_{2x},h)dh \\
&= \int_0^{10^6} H(\frac{1}{3},\frac{1}{3},\frac{1}{3}) + \int_{10^6}^{2\times10^6} H(\frac{2}{3},\frac{1}{3}) + \int_{2\times10^6}^\infty H(\frac{3}{3}) \\
&= \int_0^{10^6} 1.585 + \int_{10^6}^{2\times10^6} 0.811 + \int_{2\times10^6}^\infty 0.000 \\
&= 2.396 \times 10^6
\end{aligned}$$

Similarly, the hierarchic entropy for system $\mathcal{R}_{4x}$ is 4.018×10^6. So when $x = 1000000$ we again have $S(\mathcal{R}_{2x}) < S(\mathcal{R}_{4x})$; society $\mathcal{R}_{4x}$ is again 1.68 times more diverse than $\mathcal{R}_{2x}$

In fact, $S(\mathcal{R}_{2x}) < S(\mathcal{R}_{4x})$ holds for all values of $x > 0$:

$$\begin{aligned}
\int_0^\infty H(\mathcal{R}_{2x},h)dh &\stackrel{?}{<} \int_0^\infty H(\mathcal{R}_{4x},h)dh \\
\int_0^x H(\frac{1}{3},\frac{1}{3},\frac{1}{3}) + \int_x^{2x} H(\frac{2}{3},\frac{1}{3}) + \int_{2x}^\infty H(\frac{3}{3}) &\stackrel{?}{<} \int_0^x H(\frac{1}{3},\frac{1}{3},\frac{1}{3}) + \int_x^{4x} H(\frac{2}{3},\frac{1}{3}) + \int_{4x}^\infty H(\frac{3}{3}) \\
\int_0^x 1.585 + \int_x^{2x} 0.811 + \int_{2x}^\infty 0.000 &\stackrel{?}{<} \int_0^x 1.585 + \int_x^{4x} 0.811 + \int_{4x}^\infty 0.000 \\
1.585x + 0.811x + 0.000 &\stackrel{?}{<} 1.585x + 0.811 \times 3x + 0.000 \\
2.396x &< 4.018x
\end{aligned}$$

In addition to scale invariance, hierarchic entropy benefits from several other advantages. Hierarchic entropy addresses a key weakness of simple social entropy by accounting for continuous differences between elements in the society. Figure 16 illustrates the kind of difference in societal structure hierarchic entropy can distinguish. In an earlier example, simple social entropy could not resolve differences between these systems (Figure 6). However, when hierarchic social entropy is employed, the measured diversity of the three systems increases linearly as the one agent is positioned further and further away. As one would expect, the difference in diversity between systems 16a and 16b is much smaller than that between 16b and 16c. This is not necessarily the case when simple entropy is used (as the earlier example illustrates).

Hierarchic entropy preserves the basic properties of simple social entropy when agent differences are binary. Hierarchic entropy is a more general metric than simple entropy, subsuming the properties of H at each taxonomic level h. In the case where difference between agents is binary, (either alike or unlike), Equation 4 degenerates to $H(\mathcal{R})$ (simple entropy) because the clustering does not depend on h. However, when continuous differences are important, hierarchic entropy can resolve structural difference in societies that simple social entropy cannot.

Figure 17 shows how the basic properties of simple social entropy are preserved with hierarchic entropy. In this example two subsets are located a fixed distance apart in the classification space.

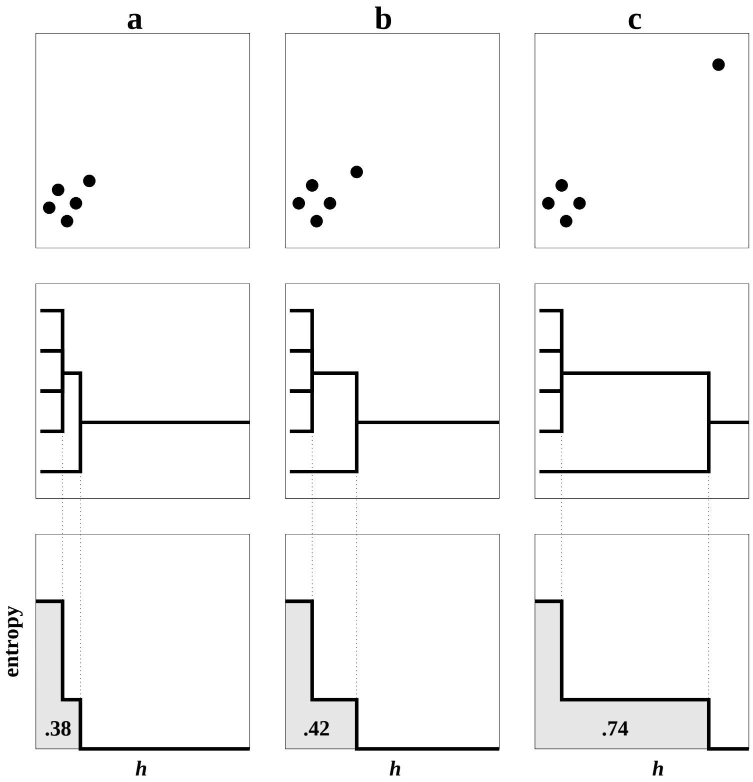


Figure 16: Hierarchic social entropy (bottom) is computed for three societies (top). The values are 0.715 for the system on the left and 1.00 for the system on the right. The calculated value increases as the element on the upper right is positioned further away from the group. Dendrograms for the groups are also displayed (middle row).

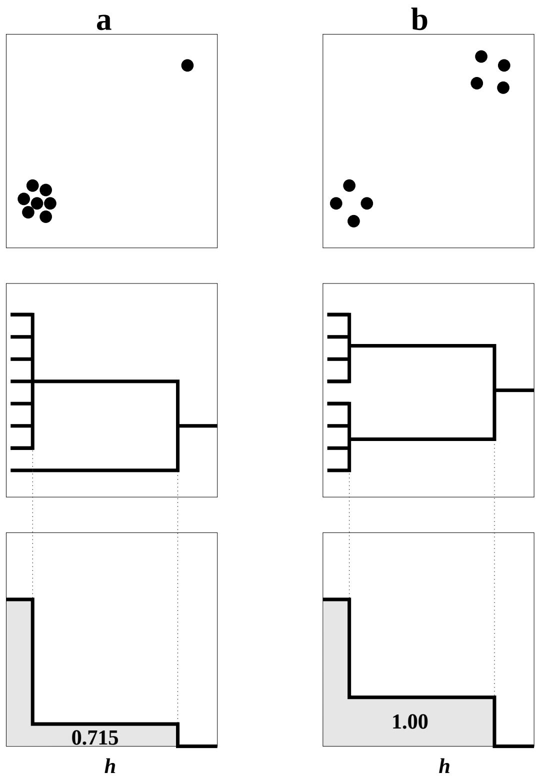


Figure 17: Hierarchic social entropy retains the basic properties of simple entropy. The computed value (bottom) depends on the distribution of elements in the subsets. Dendrograms for the two groups are also displayed (middle row).

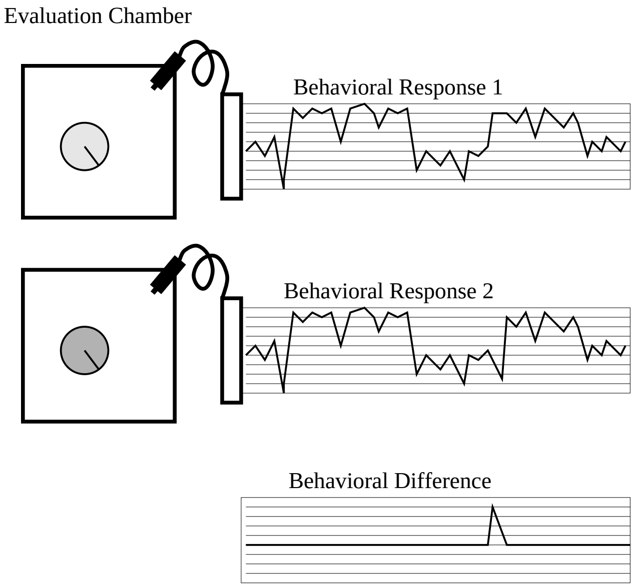


Figure 18: Evaluating behavioral difference using an idealized "evaluation chamber." Robots are evaluated in the chamber (left), where their response to every situation is recorded as a trace (readout, right). The behavioral difference between two agents is the difference between their traces (bottom). A single quantitative value is given by integrating the difference.

The two societies pictured differ only in the distribution of elements between the subsets. Hierarchic entropy properly captures the increased diversity of the system with agents distributed equally between the subsets.

It is important to note however, that the properties of simple entropy were formulated on the assumption of a fixed partitioning of the group being studied. This assumption does not necessarily hold when continuous (rather than binary) differences are used to cluster the elements. The number of subgroups and the proportion of agents in each both depend on h, the taxonomic level.

7 Behavioral difference

To summarize the paper so far, hierarchic clustering is a means of dividing a society into subsets of behaviorally equivalent agents at a particular taxonomic level. Diversity is evaluated at each taxonomic level based on the number of subsets and the number of robots in each subset at that level. Integrating the diversity across all taxonomic levels produces an overall measure of diversity for the system. Previous sections have described the overall diversity metric and algorithms for clustering the agents into subsets. This section focuses on the difference metric used for clustering.

How should the behavior of two agents be compared? One possibility would be to evaluate their difference in an "evaluation chamber" in which the robots are exposed to all situations and their their responses recorded. Even though it is unlikely such a chamber could be built, the analogy is useful. Figure 18 illustrates the procedure. As the agents are exposed to various situations, responses are recorded as a trace. After the experiment is concluded, the traces are compared to evaluate the difference between agents. In the figure, the horizontal axis of the traces represents all distinct perceptual situations a robot might experience, while the vertical component encodes the agent's response.

Since a real evaluation chamber would be practically impossible to build, an alternative method for evaluating behavioral difference is proposed. The technique advocated here is to look for differences in the agents' behavioral coding. In many cases (e.g. [BBC+95, Mat92, GM97]) robot behavior is coded statically ahead of time, thus individuals may be directly compared by evaluating their behavioral configuration. Learning multirobot systems (e.g. [Bal97, Mat94]) pose a challenge because their behavior evolves over time. To avoid that problem in this research, the policies of learning agents are evaluated after agents converge to stable behavior.

This approach depends on three key assumptions:

Assumption 1: At the time of comparison, the robots' policies are fixed and deterministic.

Assumption 2: The robots under evaluation are substantially mechanically similar: differences in overt behavior are influenced more significantly by differences in policy than by differences in hardware.

Assumption 3: Differences in policy are correlated with differences in overt behavior.

If these conditions are not met in a particular multirobot system, the approach may not be appropriate. But the assumptions are reasonable for the conditions of this research, namely: experiments conducted on mechanically similar robots built on the same assembly line. Control systems running on the robots differ only in the data specifying each agent's policy. The comparison of these policies is the crux of the approach.

7.1 Example: multirobot foraging

The objective is to show how behavioral difference can be evaluated by examining differences in robots' behavioral coding. Before proceeding, an example encoding is presented. Although this example describes a particular robot architecture, this is only for illustration, the method is applicable to other architectures as well.

Consider how behaviors could be designed for a team of foraging robots. The task is to collect colored objects (red) and place them into colored bins (red and blue). For this example, one agent will be programmed to place the objects in the red bin, while the other will deposit them in the blue bin.[2]

In this approach to behavioral configuration, the agent is provided several behavioral assemblages (skills) that correspond to steps in achieving the task (e.g. *wander*, *acquire*, *deliver*, and so on) [ABN93]. Binary perceptual cues are used to sequence the robot through the steps in achieving the task.

The agents are provided with the perceptual features enumerated in Table 1. At the behavior selection level, the robot's perception can be represented by four bits (one bit per perceptual feature).

[2]This task is a simplified version of the task for robots in the AAAI-97 contest. The simplification is necessary in order to allow a complete enumeration of the robots' policies. The complete system is explained in more detail in Section 9.

Table 1: Perceptual features available to the foraging robots. Each feature is equivalent to one bit of information; the entire perceptual state is a four-bit value.

perceptual feature	meaning
`red_visible`	a red attractor is visible.
`red_in_gripper`	a red attractor is in the gripper.
`close_to_red_bin`	close enough to the red delivery area to drop an attractor in it.
`close_to_blue_bin`	close enough to the blue delivery area to drop an attractor in it.

Table 2: Behaviors the robots select from in accomplishing the foraging task.

behavior	robot activity
wander	Search the environment for attractors.
acquire_red	Proceed to the closest red object and grasp it.
deliver_blue	Go the the blue delivery area.
deliver_red	Go to the red delivery area.

Given the perceptual state, the robot selects from one of the four behaviors listed in Table 2. Decomposing the task into a state/action space enables a robot's policy to be enumerated by pairing perceptual states with actions. Some of the 16 states are never actually encountered since it is impossible for an agent to be simultaneously in the red and blue delivery zones.

Using this approach, policies for the two robots are now described. One of the robots collects red objects and places then in the red bin, while the other places them in the blue bin. The policy for robot r_{red} is to search for red attractors using the *wander* behavior. When it sees an attractor, it activates the *acquire_red* behavior. Once it has grasped the object, it uses the *deliver_red* behavior to go to the red bin. Robot r_{blue} is similar, except it delivers to the blue bin instead. Policies for the two agents are enumerated in Table 3. The behaviors in the center of the table are activated when the corresponding perceptual situations on the left are encountered.[3]

The actions selected by the agents described above differ in six of the states. In the case where the robots have a red object in their gripper but aren't close to a bin, they choose different actions (to either go to the red or blue bin). When they are close to the correct bin, they both drop the attractor and resume the *wander* behavior. Next, we explain how a numerical value can be assigned to this behavioral difference.

[3]Note that in the *wander* behavior, the robot's gripper opens automatically. A transition to the *wander* behavior causes the robot to drop the attractor and begin a new search.

Table 3: The policies of two foraging robots. Robot r_{red} collects red objects and places them in the red bin, Robot r_{blue} collects red objects and places then in the blue bin. Differences between the actions are listed on the right. The state bits represent, from left to right, `red_visible`, `red_in_gripper`, `close_to_red_bin`, and `close_to_blue_bin`. Impossible states are indicated with an asterisk (*).

state	robot r_{red} action	robot r_{blue} action	response difference
0000	*wander*	*wander*	0.0
0001	*wander*	*wander*	0.0
0010	*wander*	*wander*	0.0
0011*	*wander*	*wander*	0.0
0100	*deliver_red*	*deliver_blue*	1.0
0101	*deliver_red*	*wander*	1.0
0110	*wander*	*deliver_blue*	1.0
0111*	*wander*	*wander*	0.0
1000	*acquire_red*	*acquire_red*	0.0
1001	*acquire_red*	*acquire_red*	0.0
1010	*acquire_red*	*acquire_red*	0.0
1011*	*acquire_red*	*acquire_red*	0.0
1100	*deliver_red*	*deliver_blue*	1.0
1101	*deliver_red*	*wander*	1.0
1110	*wander*	*deliver_blue*	1.0
1111*	*wander*	*wander*	0.0

7.2 Definition of behavioral difference

To facilitate the discussion, the following additional symbols and terms are defined:

- i is a robot's perceptual state.
- a is the action (behavioral assemblage) selected by a robot's control system based on the input i.
- π_j is r_j's policy; $a = \pi_j(i)$.
- p_j^i is the number of times r_j has encountered perceptual state i divided by the total number of times all states have been encountered. Experimentally, p_j^i is computed *post facto*.

The approach is to evaluate behavioral difference by comparing the robots' policies. The two foraging robots introduced earlier, for example, exhibit behavioral differences that are reflected in and caused by their differing policies. In the terminology introduced above, i represents the perceptual features an agent uses to selectively activate behaviors. In the case of the foraging robots, assign a bit to each perceptual feature, so, for example, $i =$ `0001` indicates that only the last perception (`close_to_blue_bin`) is activated. For the foraging robots $\pi(i)$ is the activated behavior (e.g. *wander*, *deliver*).

Definition 1: r_a and r_b, are **absolutely behaviorally equivalent** iff they select the same behavior in every perceptual state.

In complex systems with perhaps thousands of states and hundreds of actions it may also be useful to provide a scale of equivalence. This would allow substantially similar agents to be grouped in the same cluster even though they differ by a small amount. The approach is to compare two robots, r_a and r_b, by integrating the differences between their responses, $\mid \pi_a(i) - \pi_b(i) \mid$ over all perceptual states i. If the action is a single-dimension scalar, as in a motor current for instance, the difference can be taken directly. However, complex actions like *wander* and *acquire* are treated as nominal values with response difference defined as 0 when $\pi_a(i) = \pi_b(i)$ and 1 otherwise. This approach is often used in classification applications to quantify difference between nominal variables (e.g. eye color, presence or absence of a tail, etc.). Using this notation, a simple behavioral difference metric can be defined as:

$$D'(r_a, r_b) = \frac{1}{n} \int \mid \pi_a(i) - \pi_b(i) \mid di \tag{5}$$

or for discrete state/action spaces:

$$D'(r_a, r_b) = \frac{1}{n} \sum_i \mid \pi_a(i) - \pi_b(i) \mid \tag{6}$$

where $\frac{1}{n}$ is a normalization factor to ensure the difference ranges from 0 to 1. In the case of the discrete sum, n corresponds to the number of possible states. If r_a and r_b select identical outputs ($\pi_a(i) = \pi_b(i)$) in all perceptual states (i), then $D'(r_a, r_b) = 0$. When r_a and r_b select different outputs in all cases $D'(r_a, r_b) = 1$. In the numerical taxonomy literature, this difference is called the *mean character difference* [SS73]. The calculation parallels the idealized evaluation chamber procedure introduced earlier (Figure 18).

Equations 5 and 6 weigh differences equally across all perceptual states. This may be problematic for agents that spend large portions of their time in a small portion of the states. Consider two foraging robots that differ only in their reaction to blue attractors. If, in their environment, no blue attractors are present the agents would appear to an observer to have identical policies.

There may be other important reasons certain states are never visited. In learning a policy, for instance, the robots might discover in early trials that certain portions of the state space should be avoided due to large negative rewards. Because these portions of the space are avoided, the agents will not refine their policies there, but avoid them entirely. It is entirely possible for the agents to differ significantly in these portions of the space even though they may appear externally to behave identically.

To address this, the response differences in states most frequently visited should be emphasized while those that are infrequently experienced should be de-emphasized. This is accomplished by multiplying the response difference in each situation by the proportion of times that state was visited by each agent ($p_a^i + p_b^i$). Formally, **behavioral difference** between two robots r_a and r_b is defined as:

$$D(r_a, r_b) = \int \frac{(p_a^i + p_b^i)}{2} \mid \pi_a(i) - \pi_b(i) \mid di \tag{7}$$

or in discrete spaces

$$D(r_a, r_b) = \sum_i \frac{(p_a^i + p_b^i)}{2} \mid \pi_a(i) - \pi_b(i) \mid \tag{8}$$

Table 4: Sample evaluation of the behavioral difference between the two agents whose policies are listed in Table 4.3. The number of times each state was visited by each agent is listed and used to compute p^i for each state for each robot. In turn, the proportion of visits to each state is used to normalize the response difference between the agents. Note: these values were not gathered from experiment, they are presented for example only.

state	times r_{red} visited	p^i_{red}	times r_{blue} visited	p^i_{blue}	difference (from Table 4)	normalized difference
0000	100	0.1	100	0.1	0.0	0.00
0001	-	-	100	0.1	0.0	0.00
0010	100	0.1	-	-	0.0	0.00
0011	-	-	-	-	0.0	0.00
0100	200	0.2	200	0.2	1.0	0.20
0101	-	-	200	0.2	1.0	0.10
0110	200	0.2	-	-	1.0	0.10
0111	-	-	-	-	0.0	0.00
1000	100	0.1	100	0.1	0.0	0.00
1001	-	-	100	0.1	0.0	0.00
1010	100	0.1	-	-	0.0	0.00
1011	-	-	-	-	0.0	0.00
1100	100	0.1	100	0.1	1.0	0.10
1101	-	-	100	0.1	1.0	0.05
1110	100	0.1	-	-	1.0	0.00
1111	-	-	-	-	0.0	0.00
totals	1000	1.0	1000	1.0	6.0	0.55

When r_a and r_b select differing outputs in a given situation, the difference is normalized by the joint proportion of times they have experienced that situation.

As an example of how behavioral difference is calculated, suppose the robots introduced earlier are evaluated in an experimental run.[4] During the experiment, the number of times each agent visits each state is recorded. This log, along with the response differences listed in Table 3 can be used to compute the behavioral difference between the two agents. The calculation is illustrated in Table 4. The number of times each agent visited each state is enumerated, then used to compute p^i for each robot for each state. The normalized behavioral difference at each state is listed in the right column, and summed at the lower right. The value in the lower right-hand corner, 0.55, is the behavioral difference between robots r_{red} and r_{blue}.

The measure of behavioral difference provides for the following definitions:

Definition 2: r_a and r_b, are **ϵ-equivalent** iff $D(r_a, r_b) < \epsilon$.

Definition 3: $\equiv_\epsilon$ indicates ϵ-equivalence, $r_a \equiv_\epsilon r_b$ means r_a and r_b are ϵ-equivalent.

Definition 4: A robot society, $\mathcal{R}$, is **ϵ-homogeneous** iff for all $r_a, r_b \in \mathcal{R}$, $r_a \equiv_\epsilon r_b$

ϵ in these definitions is closely tied to the parameter h used in C_u clustering. A classification at taxonomic level h will consist of h-homogeneous clusters.

Currently this approach is limited to evaluating behavioral difference between policies represented as a deterministic function, e.g. $\pi(i)$. This is reasonable for the analysis of policies developed using reinforcement learning techniques since once learning is complete, the policies are fixed. It does not address robots utilizing FSAs for behavioral sequencing. An FSA might generate a different

[4]This example experiment is for illustrative purposes only.

output (action) in the same perceptual state, depending on the sequence of inputs up to that point. To address the problem a quantitative technique for comparing FSAs is required. One avenue of approach would be a comparative analysis of the "languages" (actually sequences of perceptual state) accepted by two agents under evaluation. This is beyond the scope of the present investigation however.

8 Application of simple social entropy in simulated soccer experiments

We now describe how the simple and hierarchic social entropy measures can be employed experimentally. In this first set of experiments, specialization in learning simulated robot soccer teams is examined using simple social entropy. In the following section, hierarchic social entropy is used to evaluate foraging robot teams in simulation and on laboratory platforms.

In both sets of experiments, the agents are provided a common set of skills (motor schema-based behavioral assemblages) from which they build a task-achieving strategy using reinforcement learning. The agents learn individually to activate particular behavioral assemblages given their current situation and a reward signal.

8.1 The simulated soccer task

Robot soccer is an increasingly popular domain for robotics research [KAK+97]. It is an attractive domain for multiagent investigations because a robot team's success against a strong opponent usually requires some form of cooperation. Also, it is familiar to many audiences and it provides opportunities for diversity among the team members. We provide a brief review of the system and experiments here. For more detail, the reader is referred to [Bal98].

In this soccer task, each team is composed of four simulated robot players. Once play begins the teams attempt to push and/or kick the ball into the opponent's goal (in a difference from RoboCup soccer, the goal spans the width of the field's boundary). Play is continuous; after a scoring event, the ball is immediately replaced to the center of the field without repositioning the agents. In each game, play continues until a total of 10 points are scored. The experiments were conducted by engaging an *experimental* learning team against a fixed opponent *control* team in soccer contests. Performance is evaluated as the total number of points scored by the learning team. We begin by describing the control team's behavioral configuration.

8.2 The control team

The control team will always follow a fixed policy against the teams under evaluation. The control team's design is based on the following observations. First, points are scored by bumping the ball across the opponent's goal. Second, robots must avoid bumping the ball in the wrong direction, lest they score against their own team. A reasonable approach then, is for the robot to first ensure it is behind the ball, then move towards it to bump it towards the opponent's goal. Alternately, a defensive robot may opt to remain in the backfield to block an opponent's scoring attempt.

Table 5: The control soccer team's policy summarized as policy tables. The 1 in each row indicates the behavioral assemblage selected by the robot for the perceived situation indicated on the left. The abbreviations for the assemblages are introduced in the text.

Control Team Forward

perceptual feature	**assemblage** *mtb*	*gbb*	*mtbf*
not behind_ball	0	1	0
behind_ball	1	0	0

Control Team Goalie

perceptual feature	**assemblage** *mtb*	*gbb*	*mtbf*
not behind_ball	0	1	0
behind_ball	0	0	1

Each robot selects from a set of behavioral assemblages to complete the task. The behaviors are sequenced to form a complete strategy. The behavioral assemblages developed for these experiments are:

- *move_to_ball (mtb)*: The robot moves directly to the ball. A collision with the ball will propel it away from the robot.
- *get_behind_ball (gbb)*: The robot moves to a position between the ball and the defended goal while dodging the ball to avoid bouncing it in the wrong direction.
- *move_to_back_field (mtbf)*: The robot moves to the back third of the field while being simultaneously attracted to the ball. The robot will kick/bump the ball if it comes within range.

Each of these behavioral assemblages is composed through the cooperative combination of several motor schemas. More detail on these behaviors is provided in [Bal98].

The overall system is completed by sequencing the assemblages with a selector that activates an appropriate skill depending on the robot's situation. This is accomplished by combining a boolean perceptual feature, *behind_ball* with a selection operator. The selector picks one of the three assemblages for activation, depending on the current value of *behind_ball*.

The control team includes three "forwards" and one "goalie." The forwards and goalie are distinguished by the assemblage they activate when they find themselves behind the ball: the forwards move to the ball while the goalie remains in the backfield. Both types of player will try to get behind the ball when they find themselves in front of it.

8.3 Learning soccer teams

To isolate the impact of learning on performance, the learning teams were developed using the same behavioral assemblages and perceptual features as the control team. This approach ensures that the performance of a learning team versus the control team is due only to differences in policy.

The control team's configuration uses a fixed selector for coordination. Learning is introduced by replacing the fixed mechanism with a selector that uses Q-learning instead. The Q-learner automatically tracks previous perceptions and rewards to refine its policy [WD92]. At each step, the learning module is provided the current reward and perceptual state. It learns over time to select the best assemblage given the situation.

The policy an agent learns depends on the reward function used to train it. One objective of this research is to discover how *local* versus *global* reinforcement impacts the diversity and performance of learning teams. Global reinforcement refers to the case where a single reinforcement signal is simultaneously delivered to all agents, while with local reinforcement each agent is rewarded individually. To that end, we consider two reinforcement functions for learning soccer robots:

- **Local performance-based reinforcement:** each agent is rewarded individually when it scores a goal, or is punished when it is nearest the ball when the team is scored against.
- **Global performance-based reinforcement:** all agents are rewarded when when the team scores, or punished when the team is scored against.

Complete details of the formulation of these reward functions are provided in [Bal98].

8.4 Soccer results: performance and diversity

Experimental data were gathered by simulating thousands of soccer games and monitoring robot performance. The learning robots are evaluated on two criteria: task performance (score) and diversity of behavior.

For each trial, the learning robots were initialized with all Q-values set to zero. A series of 100 10-point games were played. Information on performance was recorded after each game. The robots retain their learning set between games. An experiment is composed of 10 runs, or a total of 1000 10-point games. Each run uses the same initial parameters but different pseudo-random number seeds.

Performance is measured as the difference between the learning team's score and the opponent's score. A negative value indicates the team lost the game, while positive values indicate the team won the game. When rewarded using the global reinforcement signal R_{global}, the learning teams out-score the control team by an average of six points to four, yielding a performance of 2.0. This average includes the initial phase of training. When trained using the local reward R_{local}, the learning teams lose by an average of four points to six, or a performance of -2.0. In these soccer experiments, teams trained using global reinforcement perform best.

After the training phase, robot teams are evaluated for behavioral diversity by examining their policies. Altogether there are 9 possible policies for the learning agents since for each of the two perceptual states, they may select one of three assemblages. Based on these nine policies, there are a total of 6561 possible 4 robot teams.

The extent of diversity in these teams is quantified using the simple social entropy metric. Agents are classified as *behaviorally equivalent* only if they share exactly the same policy.

Two example teams, one homogeneous, the other heterogeneous, are illustrated in Figure 19. All members of the team on the left have converged to identical policies. In fact, *all* robots in the 10 locally-reinforced teams converged to the same "forward" policy used by the control team (Table 5). All 10 teams converged to fully homogeneous behavior. $H(\mathcal{R}) = 0$ for the homogeneous teams trained using local reinforcement.

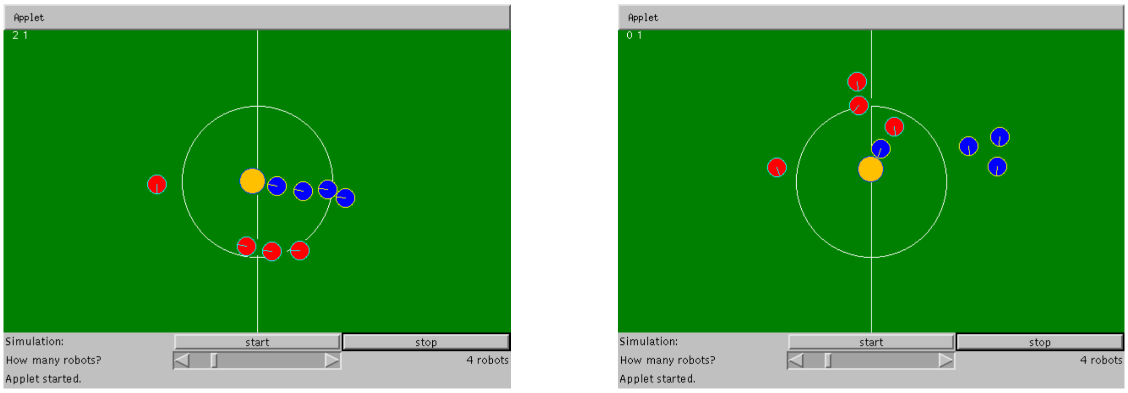


Figure 19: Examples of homo- and heterogeneous learning soccer teams. In both cases the learning team (dark) defends the goal on the right. The agents try to propel the ball across the opponent's goal by bumping it. A homogeneous team (left image) has converged to four identical behaviors which in this case causes them to group together as they move towards the ball. A heterogeneous team (right) has settled on diverse policies which spread them apart into the forward and middle of the field.

In contrast, all of the 10 globally-reinforced teams diversify to heterogeneous behavior. In all cases, the agents settle on one of three particular policies. All the teams include one robot that converges to the same "forward" policy used by the control team; they also include at least one agent that follows the same policy as the control team's "goalie." The other robots learn a policy of always selecting the *get_behind_ball* assemblage, no matter the situation (for convenience this policy is referred to as a "mid-back").

In cases where the team had not fully converged, investigation reveals that the changes are due to one agent alternating between the "goalie" and "mid-back" policies. In summary, the globally-reinforced teams always converged to one "forward," one or two "mid-backs" and one or two "goalies." $H(\mathcal{R}) = 1.5$ for the heterogeneous teams (the maximum entropy for a team of four soccer robots is 2.0).

8.5 Summary of results in the soccer task

Two reward functions, R_{local}, R_{global} were employed by learning soccer teams as they engaged a fixed opponent team in thousands of trials. The relative impact of the reward function on team performance and diversity was evaluated. Simple social entropy was employed to measure diversity in the teams. The primary results are

- in some cases individual learning robots automatically diversify to fill different roles on a team;
- global reinforcement leads to better performance and greater diversity in this simulated soccer task;
- local reinforcement leads to poorer performance and more homogeneous behavior.

The locally-reinforced teams converge to "greedy" behaviors that maximize their individual reward, but lead to poor team performance. This may suggest that defensive play is important in soccer but there is no incentive for a robot to fill a defensive role. With the local reward strategy a goalie would be "punished" every time the opponent scores and never receive a positive reinforcement.

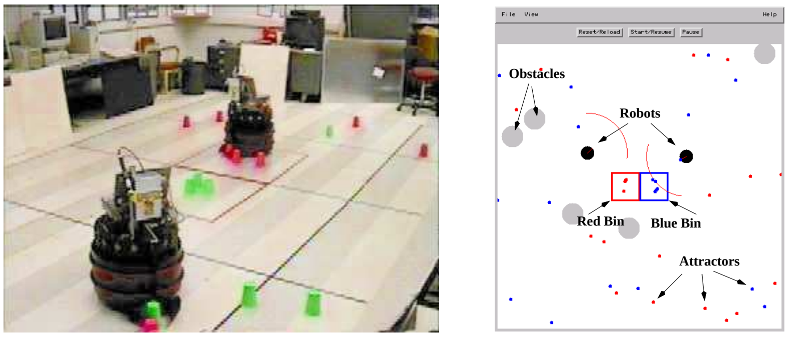


Figure 20: Real and simulated robot foraging. Left: two robots forage for colored attractors in the laboratory; after grasping an object, they deposit it in one of two delivery zones according to color. Right: in simulation, robots are represented as black circles, arcs indicate the robots' visual sensing range, obstacles are drawn as gray circles, the small discs are attractors. The robots deliver the attractors to the color-coded squares representing delivery areas.

The use of the simple social entropy metric enables the quantification of diversity in these learning teams and correlation of diversity with performance. The policies for these robots, however, are rather simple and lend themselves to the discrete classifications necessary for simple entropy. What about very large policy spaces?

Simple social entropy would probably not be as useful an analytic measure when many thousands of policies are possible. In this case it is likely for each robot to have slight differences in policy from every other robot (and therefore be in a different "class"). The resulting simple entropy would always be at a maximum for all robot teams — thus making it impossible to make relative comparisons of diversity.

Hierarchic social entropy can address this limitation. In the next section we examine a robot team with a more complex policy space and show how hierarchic social entropy may be used in its analysis.

9 Application of hierarchic social entropy in multirobot foraging experiments

The task in this set of experiments calls for robots to collect several different types of objects and deposit them in different bins according to type. As an example of why this task is useful, consider a janitorial robot responsible for collecting and sorting recyclable trash objects into glass, aluminum and paper bins. Many assembly and construction tasks also involve collecting parts or materials and placing them in a specific location. These tasks are referred to as *multi-foraging* tasks. *Multi* refers to the multiple types of object to deliver, not the number of robots engaged in the task. An example of robots executing a multi-foraging task is presented in Figure 20.

9.1 Behaviors for multi-foraging

As in the soccer example above, a schema-based reactive control system is used for robot programming. Each agent is provided several pre-programmed skills that correspond to steps in achieving

Table 6: Perceptual features available to the foraging robots. Each feature is one bit of environmental state; the entire perceptual state is a nine-bit value.

perceptual feature	meaning
`red_visible`	a red attractor is visible.
`blue_visible`	a blue attractor is visible.
`red_visible_outside_homezone`	a red attractor is visible outside the three meter radius home zone.
`blue_visible_outside_homezone`	a blue attractor is visible outside the home zone.
`red_in_gripper`	a red attractor is in the gripper.
`blue_in_gripper`	a blue attractor is in the gripper.
`close_to_homezone`	the agent is within 3 meters of the homebase.
`close_to_red_bin`	close enough to the red delivery area to drop an attractor in it.
`close_to_blue_bin`	close enough to the blue delivery area to drop an attractor in it.

the task (e.g. *wander*, *acquire*, *deliver*, and so on). Binary perceptual features are used to sequence the robot through steps in achieving the task. Selection of the appropriate behavior, given the situation, may be programmed by hand or discovered by the robot through reinforcement learning. (In addition to the learning strategies investigated here, these behaviors were also used to build successful hand-coded foraging strategies, including a winning entry in the AAAI-97 Robot Competition [Bal99].)

A range of skills were developed to support a number of foraging strategies and to avoid bias towards any particular approach. The repertoire is suitable for building behaviorally homogeneous foraging teams as well as various heterogeneous strategies. The behaviors are summarized below:

- *wander:* move randomly about the environment in search of attractors. Upon encountering an attractor, most agents learn to transition to an appropriate *acquire* behavior.
- *stay_near_home:* similar to the *wander* assemblage, but with an additional attractive force to keep the agent close to the homebase. This assemblage might be utilized in a territorial foraging strategy.
- *acquire_red:* move towards the closest visible red attractor. When close enough to grasp the attractor, most agents learn to close their gripper and transition to a *deliver* assemblage.
- *acquire_blue:* move towards the closest visible blue attractor.
- *deliver_red:* move towards the red delivery area. When close enough to deposit the attractor in the delivery area, most agents learn to open their gripper and transition to one of the *wander* assemblages.
- *deliver_blue:* move towards the blue delivery area.

All of the behaviors include a provision for obstacle and robot avoidance. More detail on the specifics of the behaviors (e.g. schema parameter values) is provided in [Bal98].

The perceptual state is a combination of nine *perceptual features*. Each feature is a single, abstracted bit of environmental or sensor state germane to the robot's task (e.g. whether or not the robot is holding an attractor in its gripper). The perceptual features used in this work are cataloged in Table 6. In addition to the features advising the robot whether an attractor is visible, there are also features indicating whether attractors are visible outside the delivery (or "home") zone. The visibility cues are used to allow hand-coded territorial agents (reported in [Bal99]) to search for attractors at a distance from the delivery zone (home zone) while ignoring the others (and *vice-versa*).

Overall, the policy space in this task is much larger than that explored in the soccer example. While only nine policies per robot were possible in the soccer example, there are more than 3000 possible foraging policies.

9.2 Learning strategies for foraging

As in the soccer experiments, the approach is to provide each agent a reward function that generates feedback at each movement step regarding the agent's progress, then to use that function over many trials to train the robot team. Again, Q-learning is used to associate actions with state. The learning agents are initialized with random Q-tables, thus random, poorly performing policies. Since each agent begins with a different policy, the teams are initially maximally diverse. They improve their policies using the reinforcement functions described below.

Three reward functions are investigated here:

- **Local performance-based reinforcement:** each agent is rewarded individually when it delivers an attractor.
- **Global performance-based reinforcement:** all agents are rewarded when any agent delivers an attractor.
- **Local shaped reinforcement:** each agent is rewarded progressively as it accomplishes portions of the task [Mat97].

Full details on the formulation of these reward functions are provided in [Bal98].

9.3 Foraging results: performance and diversity in simulation

Performance in the multi-foraging task is measured as the number of objects (attractors) collected and properly delivered by the robots in a 10 minute trial. Several environmental parameters affect the rate at which the agents collect and deliver the attractors including the number of attractors, obstacles in the environment, playing field size and the number of robots.

The following conditions were present in simulation experiments: 40 attractors (20 of each type, red and blue) and five 1 m^2 obstacles (5% coverage) randomly distributed about a 10 by 10 meter field with one to eight simulated robots. In laboratory runs there were 20 attractors and no obstacles (except arena boundaries) on a 5 by 10 meter playing field with one or two robots.

Statistical results were gathered in thousands of simulation trials. Each type of learning system under investigation was evaluated using one to eight simulated robots in five randomly generated environments. Performance is evaluated as the number of attractors collected in 10 minutes. 300 trials were run in each environment, or 12,000 runs overall. During the experiments, the perceptual state for each robot was logged at each timestep, thus enabling a *post facto* calculation of the proportion of time each robot spent in each state. Also, the policy for each robot was saved at the end of each trial for difference and diversity measurements.

Agents are able to learn the task using all three types of reinforcement. A plot of the average performance for each learning strategy versus the number of agents on the team is presented in Figure 21. (In separate research, the performance of three different hand-coded systems was also evaluated [Bal99]; performance of the best hand-coded system (a homogeneous strategy) is included in the graph for comparison).

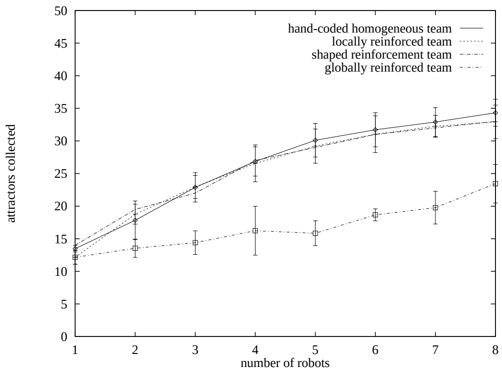


Figure 21: Performance of foraging teams versus the number of robots on a team. The error bars indicate 95% confidence intervals.

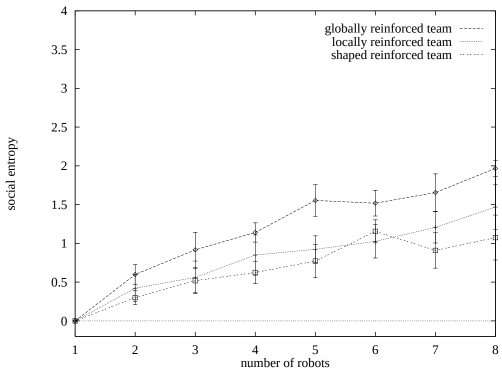


Figure 22: Hierarchic social entropy versus size of the team for learning teams; larger numbers indicate greater diversity, error bars indicate 95% confidence intervals.

The plot shows that, of the learning strategies, local performance-based and heuristic (shaped) reinforcement systems perform best. Performance in the globally reinforced system is worse than the other learning teams. Note that the performance plots for teams using local and shaped rewards are nearly identical and that one's confidence interval overlaps the other's mean value. Both also overlap the performance of the hand-coded homogeneous policy. In fact, there is no statistically significant difference between the homogeneous hand-coded systems and the best learning systems. Local and shaped reinforcement systems perform as well as the best hand-coded system.

Hierarchic social entropy is used to evaluate diversity in the learning teams. The average diversity is computed for robot teams trained with each type of reinforcement. Results are plotted versus the size of robot teams in Figure 22. In all cases with two or more agents, the globally reinforced teams are most diverse. In all but one case the teams using shaped reinforcement are the least diverse and locally reinforced teams lie between the two extremes.

Spearman's Rank-order Correlation Test is used to evaluate the relationship between diversity and performance in these systems [PTVF88]. The test measures the correlation between rankings in one dimension (e.g. performance) and another (e.g. diversity). Spearman's test indicates the rankings are strongly negatively correlated, with $r = -0.96$. The probability of the null hypothesis being true (that the rankings occur by chance) is 0.000028. Diversity and performance are negatively

Table 7: Summary of performance in learning foraging robot trials. Policies learned using local performance-based rewards were used in all trials.

configuration/trial	performance before training	after
1 robot trial 1	1.0	9.0
trial 2	0.0	10.0
trial 3	0.0	8.0
trial 4	0.0	7.0
trial 5	0.0	8.0
average	**0.2**	**8.4**
2 robots trial 1	0.0	15.0
trial 2	1.0	15.0
trial 3	0.0	16.0
trial 4	1.0	14.0
trial 5	0.0	13.0
average	**0.4**	**14.6**

correlated in these learning teams.

9.4 Foraging results: performance on mobile robots

To verify the simulation results, the learning systems were ported to Nomad 150 mobile robots. The Java-based behavioral configuration system used in this work enables the behaviors and features to be utilized on mobile robots and in simulation. Identical control software was employed in simulation and on the mobile robots.

Performance was evaluated before and after learning using local performance-based rewards on one and two robots. In each case, the robots were initialized with a random policy (the behavior for each situation is set randomly), then evaluated in a 10 minute trial. The Q-tables were transferred to the simulation system and trained for 300 trials. After training, the policies were transferred back to the robots for another evaluation. The process was repeated five times for each number of robots. Performance of the robots running learned policies is summarized in Table 7. A photograph one of the mobile robot trials is presented in Figure 20.

As in simulation the robots perform much better after the learning phase. However, they do not collect as many attractors as comparable simulated systems. This is due to the reduced number of attractors available for collection.

9.5 Summary of foraging experiments

As in the soccer example, experimental results show that the choice of reinforcement function significantly impacts the diversity and performance of learning foraging teams. Interestingly, the relationship between diversity and performance in soccer (positive correlation), is exactly opposite the relationship reported for foraging in this work (negative correlation). The reasons for this difference aren't known for certain, but we believe they are due to the differences in task. Soccer is unavoidably a *team* activity while foraging can be accomplished by an individual agent. We believe that when multiple agents are required, it is more likely that the team will benefit from diversity.

These experiments in foraging show that agents using local reinforcement strategies converge to more homogeneous societies and perform better than robots using a global reward structure. Greater homogeneity with local reinforcement is due to the fact that individuals are rewarded for their own actions, thus making reinforcement of the same state/action pair more likely in different agents than with global reinforcement. The relationship between diversity and performance is exactly opposite that found in robot soccer experiments (reported separately), but in both soccer and foraging, local rewards lead to greater homogeneity [Bal98].

The diversity of each system was evaluated using hierarchic social entropy. Globally-rewarded teams were found to be the most diverse, followed by the locally rewarded teams. Teams using shaped reinforcement were the least diverse. This is because agents using shaped reinforcement are provided more uniform "guidance" in finding a policy, and are thus less likely to settle on diverse solutions. In these learning systems, diversity and performance are negatively correlated with $r = -0.96$ and prob $= 0.000028$.

10 Conclusion

This work is motivated by the idea that behavioral diversity should be evaluated as a *result* rather than an initial condition of multirobot experiments. Previously, researchers configured robot teams as homogeneous or heterogeneous *a priori*, then compared performance of the resulting teams [FM97, GM97, Par94]. That approach does not support the study of behavioral diversity as an emergent property in multirobot teams.

Defining behavioral diversity as an independent rather than dependent variable enables the examination of heterogeneity from an ecological point of view. How and when does diversity arise in robot teams interacting with each other and their environment? This work provides the necessary quantitative measures for this new type of investigation.

Simple social entropy, an application of Shannon's information entropy [Sha49] to robot groups, is proposed as a measure of diversity in robot teams [Sha49]. It captures important components of the meaning of diversity, including the number and size of groups in a society. Researchers in many other fields also use information theoretic measures of diversity for the same reason [Wil92, LVW83, LW80, Mag88, Bai90, Dem92]. In order to evaluate the diversity of a team, however, a way to categorize or differentiate the behavior of individuals is also required. To address this, a measure of *behavioral difference* that provides for agent categorization is also developed. Difference refers to disparity between two specific agents, while diversity is a measure of the entire society.

It was shown that simple social entropy is subject to several limitations as a diversity metric; in particular, it does not capture the *extent* of difference between separate robot groups. To address this we introduce *hierarchic social entropy,* a metric combining simple social entropy with behavioral difference to provide a continuous scale of diversity. It captures even minute differences between societies, while preserving the basic properties of simple social entropy.

Diversity may not always be desirable. In fact, experimental results presented in this paper show that for at least one multirobot task (multi-foraging) homogeneous robot teams perform better than diverse teams. The aim of this work is to provide tools enabling the investigation of when diversity is important and which conditions give rise to it in learning teams. Social entropy provides the objective quantitative metric required for a principled investigation of these issues.

This research is focused specifically on diversity in teams of mechanically similar agents that use reinforcement learning to develop behavioral policies. Evaluation of diversity in teams of mechanically similar robots is challenging because when agents differ, they differ only in their behavior. Behavior is an especially interesting dimension of diversity in learning systems since as they learn, agents effectively choose between a hetero- or homogeneous society. The metrics developed in this work will help researchers investigate the origin and benefits of diversity in these learning systems.

Important future work includes the application of these tools in new and different multirobot task domains. The author hopes other researchers will adopt the measures of behavioral difference and robot team diversity introduced here in the evaluation of new multirobot systems. This will provide additional data points in the multiagent task/reward space and help us derive the relations between task, reward, diversity and performance more precisely.

Another important direction for future research is the extension of these tools to a broader range of robotic systems. The behavioral difference metric, for instance, is limited to the comparison of deterministic policies. Can we compare the behavior of agents coded in FSAs or more complex representations?

Finally, can the results of this research be applied in other fields? Researchers in behavior-based robotics often draw inspiration from biology and psychology; perhaps roboticists can provide tools for the sociobiologist. It is tempting, for instance, to draw parallels between robotic behavioral diversity and corresponding forms of specialization in human and animal societies. As the research and theory mature we may gain insights into the origins and benefits of diversity in natural as well as artificial social systems.

11 Acknowledgments

Thanks to Ron Arkin, Chris Atkeson, Maria Hybinette, Peter Stone, Michael Bowling, Gary Boone, M. Jeff Donahoo, Lynne Parker and anonymous reviewers for their suggestions on this work. The robots used in this work were provided by ONR/ARPA Grant #N00014-94-1-0215, 11/93-3/97, **Flexible Reactive Control for Multi-Agent Robotic Systems in Hostile Environments**.

References

[ABN93] R.C. Arkin, T. Balch, and E. Nitz. Communication of behavioral state in multi-agent retrieval tasks. In *Proceedings 1993 IEEE Conference on Robotics and Automation*, Atlanta, GA, 1993.

[Bai90] K. Bailey. *Social Entropy Theory.* State University of New York Press, Albany, 1990.

[Bal97] Tucker Balch. Learning roles: Behavioral diversity in robot teams. In *AAAI-97 Workshop on Multiagent Learning*, Providence, R.I., 1997. AAAI.

[Bal98] Tucker Balch. *Behavioral Diversity in Learning Robot Teams.* PhD thesis, College of Computing, Georgia Institute of Technology, 1998.

[Bal99] T. Balch. The impact of diversity on performance in multirobot foraging. In *Proc. Autonomous Agents 99*, Seattle, WA, 1999.

[BBC+95] T. Balch, G. Boone, T. Collins, H. Forbes, D. MacKenzie, and J. Santamaría. Io, Ganymede and Callisto - a multiagent robot trash-collecting team. *AI Magazine*, 16(2):39–51, 1995.

[Dem92] L. Demetrius. The thermodynamics of evolution. *Physica A*, 189(3-4):417–436, November 1992.

[FM97] M. Fontan and M. Mataric. A study of territoriality: The role of critical mass in adaptive task division. In *From Animals to Animats 4: Proceedings of the Fourth International Conference of Simulation of Adaptive Behavior*, pages 553–561. MIT Press, 1997.

[GM97] D. Goldberg and M. Mataric. Interference as a tool for designing and evaluating multi-robot controllers. In *Proceedings, AAAI-97*, pages 637–642, July 1997.

[JS71] N. Jardine and R. Sibson. *Mathematical Taxonomy.* John Wiley & Sons, 1971.

[KAK+97] M. Kitano, M. Asada, Y. Kuniyoushi, I. Noda, and E. Osawa. Robocup: The robot world cup initiative. In *Proc. Autonomous Agents 97*, Marina Del Rey, California, 1997.

[LVW83] D. Lurie, J. Valls, and J. Wagensberg. Thermodynamic approach to biomass distribution in ecological systems. *Bulletin of Mathematical Biology*, 45(5):869–872, 1983.

[LW80] D. Lurie and J. Wagensberg. Information theory and ecological diversity. In L. Garrido, editor, *Systems Far from Equilibrium*, pages 290–303, Berlin, West Germany, 1980. Sitges Conf. on Statistical Mechanics, Springer-Verlag.

[MAC97] D. MacKenzie, R. Arkin, and J. Cameron. Multiagent mission specification and execution. *Autonomous Robots*, 4(1):29–52, 1997.

[Mag88] A.E. Magurran. *Ecological Diversity and Its Measurement.* Princeton University Press, 1988.

[Mat92] M. Mataric. Designing emergent behaviors: From local interactions to collective intelligence. In *Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats 2*, pages 432–441, 1992.

[Mat94] M. Mataric. Learning to behave socially. In *Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats 3*, 1994.

[Mat97] Maja Matarić. Reinforcement learning in the multi-robot domain. *Autonomous Robots*, 4(1):73–83, January 1997.

[MM92] P. Meyer and S. McIntosh. The USA Today index of ethnic diversity. *International Journal of Public Opinion Research*, page 56, Spring 1992.

[MW89] Merriam-Webster. *Webster's ninth new collegiate dictionary.* Merriam-Webster, 1989.

[Par94] Lynne E. Parker. *Heterogeneous Multi-Robot Cooperation.* PhD thesis, M.I.T. Department of Electrical Engineering and Computer Science, 1994.

[PTVF88] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C.* Cambridge University Press, 1988.

[Sha49] C. E. Shannon. *The Mathematical Theory of Communication.* University of Illinois Press, 1949.

[SS73] P. Sneath and R. Sokal. *Numerical Taxonomy.* W. H. Freeman and Company, San Francisco, 1973.

[WD92] Christopher J. C. H. Watkins and Peter Dayan. Technical note: Q learning. *Machine Learning*, 8:279–292, 1992.

[Wil92] E.O. Wilson. *The Diversity of Life.* Norton, 1992.