Normalized Mutual Entropy in Biology: Quantifying Division of Labor

Root Gorelick,[1,*] Susan M. Bertram,[1,†] Peter R. Killeen,[2,‡] and Jennifer H. Fewell[1,§]

1. School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501;
2. Department of Psychology, Arizona State University, Tempe, Arizona 85287-1104

Submitted November 25, 2003; Accepted June 25, 2004; Electronically published September 29, 2004

abstract: Division of labor is one of the primary adaptations of sociality and the focus of much theoretical work on self-organization. This work has been hampered by the lack of a quantitative measure of division of labor that can be applied across systems. We divide Shannon's mutual entropy by marginal entropy to quantify division of labor, rendering it robust over changes in number of individuals or tasks. Reinterpreting individuals and tasks makes this methodology applicable to a wide range of other contexts, such as breeding systems and predator-prey interactions.

Keywords: information theory, entropy, task specialization, division of labor.

> The greatest improvement in the productive powers of labor, and the greater part of the skill, dexterity, and judgment with which it is anywhere directed, or applied, seem to have been the effects of the division of labor. (Adam Smith, *Wealth of Nations*, 1776)

One of the primary attributes of sociality is division of labor. Division of labor is ubiquitous across social groups, from insects to humans, and is of paramount importance in explaining the success of eusocial groups such as the social insects (Wilson 1971). Division of labor has received increased theoretical and empirical attention because of the recent development of self-organizational models, in which global patterns of social organization arise from local rules (Hemelrijk 2002). Unfortunately, the qualitative definition of division of labor—that different individuals perform different tasks—does not allow systematic comparisons of division of labor across empirical or theoretical systems (Beshers and Fewell 2001).

A first step in quantifying division of labor is to ask whether, knowing the task, we can predict the individual or the class of individual performing it. To the extent we can, there is high individual specialization. But we may also ask whether, knowing the individual or the class of the individual, we can predict the task. To the extent we can, there is high task specialization. This distinction has created some confusion in the literature. Although these measures co-vary, they are not the same, as later examples will show. It is generally believed that either type of specialization can increase efficiency at the individual and colony level. Together, individual and task specialization measure the distribution of different individuals across different tasks. In this note we present a quantitative description of division of labor that represents these considerations in a statistic that allows comparison across diverse systems. This measure uses Shannon's information theory, in particular his indices of mutual and marginal entropy, including Shannon's diversity index (Shannon 1948). Shannon described how information gets transmitted from source to receiver across a noisy channel (Shannon 1948). For our applications, individuals comprise the source, and tasks comprise the receiver.

A division of labor statistic should have the following three properties. First, it should permit comparison of division of labor statistics across data sets with different numbers of monitored individuals or tasks. We operationalize this by normalization. Second, it should reflect the extent to which each individual is a specialist. We operationalize this by quantifying the degree to which knowing the identity of a particular individual identifies the task being performed (i.e., whether individuals are specialists). Third, it should reflect the extent to which each task requires specialized performance. We operationalize this by quantifying the degree to which knowing the identity of a particular task identifies the individuals performing it (i.e., whether tasks are boutique). The difference between the second and third criteria is portrayed in figure 1 and by a homely example. In touch typing, each key is

* Corresponding author; e-mail: cycad@asu.edu.

† E-mail: sbertram@asu.edu.

‡ E-mail: killeen@asu.edu.

§ E-mail: j.fewell@asu.edu.

Am. Nat. 2004. Vol. 164, pp. 677–682.

struck by a unique finger, a complete division of tasks. Given knowledge of the task (the key), you know the individual (the finger). Conversely, each index finger is tasked with up to six keys. Knowledge of the finger used still leaves uncertainty about which key was struck.

A few measures of division of labor have been developed, but in each instance they only quantify how specialized individuals are on tasks (e.g., Kolmes 1985; Gautrais et al. 2002) or how specialized tasks are with respect to the individuals performing them (e.g., Seeley 1982; Beshers and Traniello 1996). Kolmes quantified task specialization of an individual by associating a probability vector to each individual in a group, describing the percentage of an individual's time spent on each task (Kolmes 1985). He computed Shannon's diversity index for each individual and averaged that across all individuals. This measures specialization of individuals, but not the degree to which tasks require specialized performance. In fact, antithetically, his method yields maximum division of labor when all individuals are specialists on the same task. Using an alternative formulation, Gautrais et al. (2002) also capture the notion of individual task specialization, but not the degree to which tasks require specialized performance. Both Seeley (1982) and Beshers and Traniello (1996) take the opposite tack and quantify the degree to which tasks require specialized performance in the context of quantifying age polyethism. Their approaches, however, do not quantify how specialized individuals are on tasks. We offer a general and symmetric treatment of these divisions.

Quantifying Division of Labor

We characterize division of labor by a set of individuals (labeled X), a set of tasks (labeled Y), and a bivariate probability distribution over the combination of these two sets ($p(x, y)$, $x \in X$, $y \in Y$). Perform the following steps to compute division of labor. First, create a data matrix in which each entry represents the time an individual spends on a task (below we also show how this algorithm can be modified to use presence/absence data or data on the percentage of each individual's time spent on each task). Second, normalize the data matrix by dividing each entry by the total time spent by all individuals on all tasks so that the sum of all entries of the resulting normalized matrix equals 1. Third, calculate a division of labor statistic that is a function of the normalized matrix in the following manner. Let $H(Z) = -\sum_{z\in Z} p(z)\log[p(z)]$ be Shannon's index (entropy) over Z, where Z can be the set of individuals (X) or tasks (Y). We can also let Z be a conditional distribution, thereby defining $H(X|Y = y)$ and $H(Y|X = x)$, where X and Y are the sets of individuals and tasks, respectively. Define conditional entropy of X given Y and Y given X as $H(X|Y) = \sum_{x\in X} p(x)H(Y|X = x)$ and $H(Y|X) = \sum_{y\in Y} p(y)H(X|Y = y)$, respectively (Cover and Thomas 1991). Define mutual entropy (mutual information, transinformation) over the joint distribution of X and Y as

$$I(X, Y) = \sum_{x\in X, y\in Y} p(x, y)\log\left[\frac{p(x, y)}{p(x)p(y)}\right]$$
$$= H(X) - H(X|Y) = H(Y) - H(Y|X),$$

which is computed over all cells in the matrix (Shannon 1948; Cover and Thomas 1991). Then we define division of labor as

$$\text{division of individuals } (X) \text{ into tasks } (Y)\text{: } D_{Y|X} = \frac{I(X, Y)}{H(X)},$$

$$\text{division of tasks } (Y) \text{ into individuals } (X)\text{: } D_{X|Y} = \frac{I(X, Y)}{H(Y)},$$

$$\text{symmetric division of labor: } D_{X,Y} = \sqrt{D_{Y|X}D_{X|Y}} = \frac{I(X, Y)}{\sqrt{H(X)H(Y)}}.$$

The statistic $D_{Y|X}$ tells us the extent to which knowledge of X specifies Y. For the above touch typing example, consider the letters A through Z on a standard keyboard ($X = \{A, B, C, \ldots, Z\}$ and Y = set of fingers on both hands) and consider a hypothetical language in which all keys are typed with equal frequency. Then $D_{X|Y} = 0.60$ and $D_{Y|X} = 1.00$, the latter value being the maximum value it can attain, indicating that we have complete knowledge of the finger that typed the key.

When there are more individuals (X) than tasks (Y), as is usually the case in biological systems, $D_{Y|X}$ is invariant over the number of individuals in the sample, as shown in figure 2. Sample size is usually arbitrary, so this invariance is important. If the proportion of individuals performing each task remains invariant as population size changes, so also will $D_{Y|X}$. As populations grow, however, they may change their pattern of allocating individuals across tasks as a biological consequence of scale. When this happens, $D_{Y|X}$ will appropriately reflect these changes. Because of this invariance over sample size and proportionate change, $D_{Y|X}$ (division of individuals into tasks) will undoubtedly be the most commonly used of our three statistics.

There will be occasions for the other two division of labor statistics. It is unusual for there to be more tasks than individuals, for example, investigating division of labor for a pair of ordinarily solitary individuals placed together in a forced association (Fewell and Page 1999). But, when there are more tasks than individuals, $D_{X|Y}$ is invariant over the number of tasks in the sample when individual effort is reallocated proportionately to the new tasks. All three division of labor statistics are insensitive to changes in numbers of individuals and tasks if the ratio of individuals to task remains constant. The statistics $D_{Y|X}$ and $D_{X|Y}$ are useful in the cases where data are asymmetrical (e.g., brood care by worker bees). If the data are largely symmetrical (e.g., concurrent female and male mate choice), then $D_{X,Y}$ provides a symmetric measure.

The basic statistics of information theory have well-defined sampling distributions: asymptotically, $I(X, Y)$, $H(X)$, and $H(Y)$ are each χ^2 distributed (Miller 1955; Hutcheson 1970). Therefore, our three division of labor statistics asymptotically have an F distribution. Alternatively, to avoid invoking asymptotic assumptions, bootstrap and permutation techniques permit estimating the extent to which different data sets have the same division of labor.

All three of our division of labor statistics simultaneously capture the following two facets: individuals specialize on one task instead of distributing their effort evenly across tasks, and different individuals specialize on different tasks so that task performance varies across individuals in the group. Capturing both facets into this one index allows us to statistically compare across systems. The previously cited works (Seeley 1982; Kolmes 1985; Beshers and Traniello 1996; Gautrais et al. 2002) captured only one of these two facets. Although their authors never proposed this, they could capture the other facet simply by

A

25	25	25
25	25	25
25	25	25
25	25	25
25	25	25
25	25	25

a

25	25	25	25	25	25
25	25	25	25	25	25
25	25	25	25	25	25

B

100	0	0
0	100	0
25	25	25
25	25	25
25	25	25
25	25	25

b

100	0	25	25	25	25
0	100	25	25	25	25
0	0	25	25	25	25

C

100	0	0
100	0	0
0	100	0
0	100	0
0	0	100
25	25	25

c

100	100	0	0	0	25
0	0	100	100	0	25
0	0	0	0	100	25

D

100	0	0
100	0	0
0	100	0
0	100	0
0	0	100
0	0	100

d

100	100	0	0	0	0
0	0	100	100	0	0
0	0	0	0	100	100

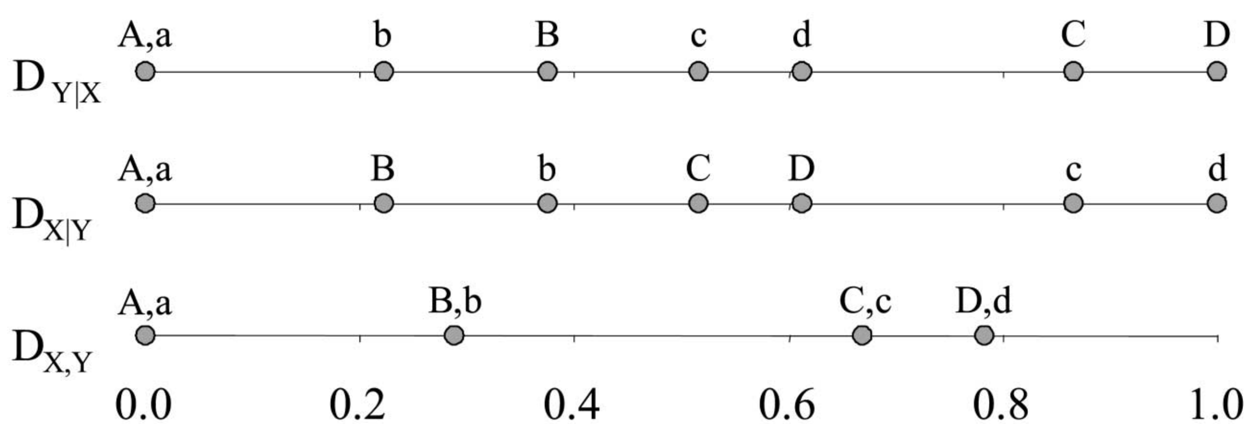

Figure 1: Mapping data matrices to values of division of labor statistics. In each data matrix, rows represent individuals; columns represent tasks. Individual entries represent amount of time an individual spends on a task. Data matrices designated by uppercase letters show six individuals with three tasks; lowercase letters show three individuals with six tasks. The symmetry between uppercase- and lowercase-lettered data matrices is evident in their resulting values of $D_{X|Y}$, $D_{Y|X}$, and $D_{X,Y}$.

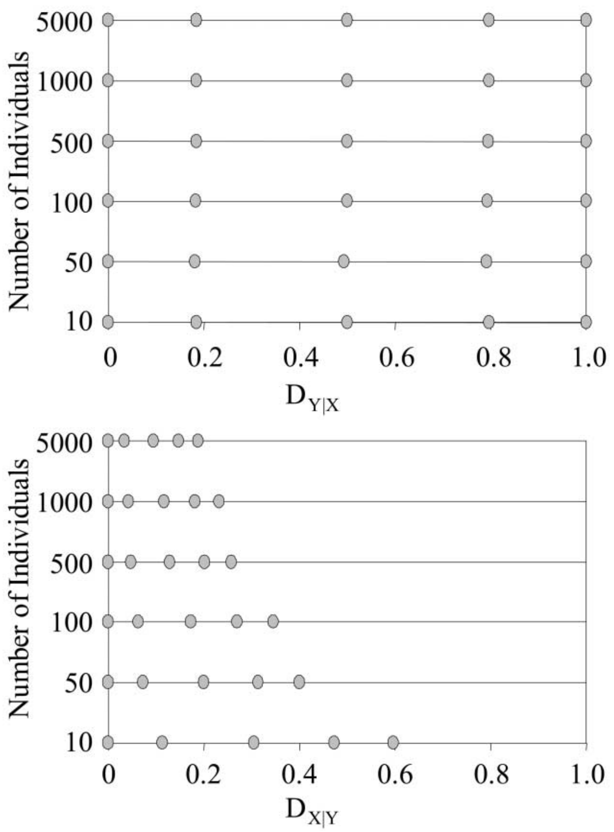

Figure 2: How size influences the division of labor statistics and insensitivity of $D_{Y|X}$ to number of individuals. Shown along the horizontal axis of each plot are the division of labor statistics for six social groups, varying only in the propensity of individuals to perform different tasks. In each social group, individuals could perform five tasks. Along each vertical line, we increased the number of individuals but retained the same proportion of time each class of individual spent on each task. All individuals in the social groups on the leftmost side of both plots were task generalists; they divided their time equally among five tasks. On the rightmost side of the plots, all individuals were task specialists, only performing one task. These individuals divided up task labor by one-fifth of them performing each task. The social groups in the middle of each plot varied in the propensity of individuals to be task specialists or generalists. From left to right, one-fifth, one-half, or four-fifths of individuals were specialists, while the remainder were task generalists. Note that the proportion of individuals performing each task does not change. With more individuals than tasks, $D_{Y|X}$ is insensitive to changing number of individuals; $D_{X|Y}$ is not.

transposing individuals and tasks, that is, transposing the data matrix $p(x, y)$. We could then quantify division of labor with a pair of statistics. The problem is that using a pair of scalar statistics does not allow for easy statistical comparison across systems. In the two-dimensional Cartesian plane, it is extremely difficult to order values into smaller and larger. Normalized mutual entropy obviates that problem by providing a one-dimensional scalar statistic.

Applications

Division of labor can be used to measure evolution of social structure across phylogenetically diverse groups. Our key hypothesis for the evolution of social structure is that the evolution of eusociality allows the evolution of more structured and extreme systems of division of labor. In such groups, individuals appear to be locked into specific task repertoires. This is expected to be more adaptive in eusocial systems than in communal groups where individual group members retain the potential for reproduction. Testing hypotheses on the evolution of sociality has been difficult because there has been no way to quantify division of labor across diverse groups with different task repertoires. The statistic $D_{Y|X}$ allows phylogenetic comparison of division of labor across diverse systems and contexts, even ranging across colonies of dozens to millions of individuals or across taxa whose last common ancestor lived tens or hundreds of millions of years ago.

The statistic $D_{Y|X}$ can also be used to test hypothesized increases in division of labor over ontogeny of a social group. A majority of social insect colonies, including ant colonies with thousands of workers, begin as a single or a few reproductive individuals (Hölldobler and Wilson 1990). Self-organizational theory predicts that division of labor should increase with group size (Theraulaz et al. 1998). An ontogenetic time history of $D_{Y|X}$ also provides a measure of when the colony has reached steady state, if a steady state exists.

A practical issue arises in analyzing division of labor in social groups: What happens if one researcher more finely subdivides tasks than another? As figure 2 shows, $D_{Y|X}$ is invariant only to number of individuals and not invariant to number of tasks. The opposite is true for $D_{X|Y}$, which is invariant to number of tasks, but not number of individuals. The statistic $D_{Y|X}$ decreases as tasks are more finely subdivided. This problem is unavoidable with a mutual entropy framework. But, as is evident from figure 2 (if we switch the labels X and Y), it requires almost an order of magnitude change in number of tasks to make an appreciable difference in $D_{Y|X}$.

This note was predicated on quantifying division of labor in social groups, but the model has applications in diverse fields unrelated to sociobiology. We may redefine tasks and individuals (agents) to a broader array of interactions, in which agents may specialize among tasks and/or tasks may specialize among agents. By reconceptualizing tasks as another form of agents, division of labor statistics apply to any context where two sets of agents interact, such as mating and predator-prey systems. For example, diet specialization within a population can be quantified using $D_{Y|X}$ by redefining tasks to be the different components of the diet.

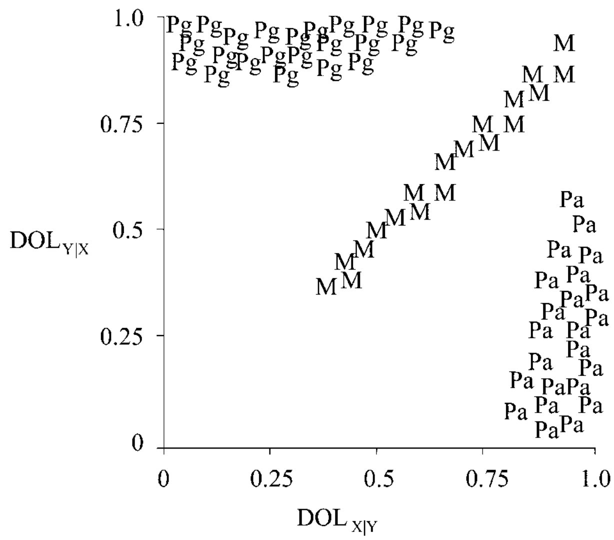

Figure 3: Division of labor statistics and mating systems and schematic of mating systems. *Pg* = polygyny; *Pa* = polyandry; *M* = monogamy. $X = \{x|x \text{ is male}\}$; $Y = \{y|y \text{ is female}\}$. Each point represents an entire hypothetical population.

Mating systems are usually pigeonholed into monogamy, polygyny, or polyandry. However, it is often quantitatively difficult to compare breeding structure across taxa or environments. Let the set of individuals be the set of males; that is, $X = \{x|x \text{ is male}\}$. Let the set of tasks be the set of females; that is, $Y = \{y|y \text{ is female}\}$. Let nonzero matrix entries reflect mating between individual females and males. This could be a binary variable, or it could reflect mating frequency. For any species or population except those with highly skewed sex ratios or where only a small fraction of one sex mates, all three division of labor statistics are relatively insensitive to changes in population size. Monogamy occurs when both females and males are specialists, that is, large $D_{X|Y}$ and $D_{Y|X}$ (fig. 3). Polyandry is indicated by only females being specialists, $D_{X|Y} \gg D_{Y|X}$. Polygyny is indicated by only males being specialists, $D_{Y|X} \gg D_{X|Y}$. Together, $D_{X|Y}$, $D_{Y|X}$, and $D_{X,Y}$ give a quantitative measure of breeding structure applicable across taxa.

Division of labor statistics are also applicable to pollination biology. Let X be the set of pollinating species and Y the set of plant species. Let the nonzero entries in the data matrix indicate that the animal acts as a pollinator for the plant. Large $D_{Y|X}$ can occur when each pollinator focuses primarily on one species of plant. Extirpation of that plant species could therefore threaten the existence of the pollinator, but not conversely. Large $D_{X|Y}$ can occur when a plant is pollinated by primarily one animal species. Destruction of the pollinator could threaten existence of the plant, but not conversely. Conservation biologists can use $D_{X,Y}$ to reflect the joint specialization of both plants and their pollinators. Large $D_{X,Y}$ reflects greater conservation risk due to greater dependence of pollinators and plants on each other.

Precisely the same arguments apply to predator-prey interactions. However, a species can be both a predator and prey if it is in the middle of a food chain. Here $D_{X,Y}$ can be used to quantify how community structure of predator-prey interactions varies with patch size, latitude, altitude, or average annual precipitation. It can also be used to quantify such differences between marine, freshwater, and terrestrial ecosystems.

We can weaken the assumption that the input data for the division of labor statistics are given as time that each individual spends on each task. For example, data on predator-prey or pollinator-plant interactions are often given as a binary variable (not time) describing whether each predator species eats each prey species (not individuals) or as proportion of time (not actual time) that each predator species spends eating each prey species (not individuals). To circumvent this data-imposed constraint, in these instances, weight the rows and columns of the data matrix to account for the a priori relative rarity of predator and prey species. Without weighting, all predator species are assumed to be equally common, and all prey species are also assumed to be equally common. Lack of weighting can severely bias division of labor statistics if most species are rare (which is often the case; Preston 1948) or rare species have different levels of specialization when compared with common species. Let C be the set consisting of all the different classes or castes, let c be an element of C, and let $w(c)$ be the proportion of the group that belongs to class c. Then multiply each row of the data matrix by its $w(c)$ and renormalize so that the sum of all elements of the resulting matrix equals 1. Repeat this process for the columns, if necessary. Weighting is not needed if the data are amount of time (e.g., in hours) that each predator species spends pursuing or eating each prey species. With time data, both predator and prey species will be automatically weighted by their time recorded in the data matrix.

Discussion

We have normalized a standard tool from information theory, mutual entropy, to produce a suite of three statistics for quantifying division of labor, each of which is mutual entropy divided by marginal entropy. Earlier researchers have used mutual entropy in molecular biology (Yockey 1992), but never in ecology. Mutual information has previously been normalized in molecular biology, but in a way with which information is lost, especially compared with the approach that we have taken here (Atchley et al. 2000). Statistical comparison of estimates of any of our three division of labor statistics from one population with that of another or with theoretical models can be done using resampling statistics or the asymptotic property of normalized mutual entropy being an F distribution. All three division of labor statistics reflect the degree to which knowing the identity of an individual identifies the task it performs and the degree to which knowing the identity of task identifies the individual that performed it. The choice of which division of labor statistic to use depends on whether the biology suggests that you concentrate on individuals (e.g., social insects; $D_{Y|X}$), both individuals and tasks (e.g., polygyny and polyandry; $D_{Y|X}$ and $D_{X|Y}$), or the interactions between individuals and tasks (e.g., predator-prey interactions; $D_{X,Y}$). Large $D_{Y|X}$ requires that individuals be specialists, tasks be boutique, and that there be more individuals than tasks. Large $D_{X|Y}$ requires that individuals be specialists, tasks be boutique, and that there be more tasks than individuals. The mirror image interpretations of $D_{Y|X}$ and $D_{X|Y}$ can be most easily seen when individuals and tasks are reinterpreted as being similar entities, for example, females and males with polygyny and polyandry (fig. 3). Although classification of which events are best conceptualized as individuals and which can best be conceptualized as tasks can seem arbitrary, the utility of these statistics is invariant over that choice because of the symmetry of mutual entropy. Thus, these division of labor statistics provide a valuable and theoretically sound tool applicable to many areas of biology.

Acknowledgments

We thank R. Jeanson, T. J. Taylor, D. S. Wilson, the Santa Fe Institute Working Group on Social Insects, and an anonymous reviewer. This work was supported by U.S. National Science Foundation grants IBN 0131728 to S.M.B., IBN 0236821 to P.R.K., and IBN 0093410 to J.H.F.

Literature Cited

Atchley, W. R., K. R. Wollenberg, W. M. Fitch, W. Terhalle, and A. W. Dress. 2000. Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. Molecular Biology and Evolution 17:164–178.

Beshers, S. N., and J. H. Fewell. 2001. Models of division of labor in social insects. Annual Review of Entomology 46:413–440.

Beshers, S. N., and J. F. A. Traniello. 1996. Polyethism and the adaptiveness of worker size variation in the attine ant *Trachymyrmex septentrionalis*. Journal of Insect Behavior 9:61–83.

Cover, T. M., and J. A. Thomas. 1991. Elements of information theory. Wiley, New York.

Fewell, J. H., and R. E. Page. 1999. The emergence of division of labour in forced associations of normally solitary ant queens. Evolutionary Ecology Research 1: 537–548.

Gautrais, J., G. Theraulaz, J. L. Deneubourg, and C. Anderson. 2002. Emergent polyethism as a consequence of increased colony size in insect societies. Journal of Theoretical Biology 215:363–373.

Hemelrijk, C. K. 2002. Understanding social behaviour with the help of complexity science. Ethology 108:655–671.

Hölldobler, B., and E. O. Wilson. 1990. The ants. Harvard University Press, Cambridge, Mass.

Hutcheson, K. 1970. A test for comparing diversities based on the Shannon formula. Journal of Theoretical Biology 29:151–154.

Kolmes, S. A. 1985. An information-theory analysis of task specialization among worker honey bees performing hive duties. Animal Behaviour 33:181–187.

Miller, G. A. 1955. Note on the bias of information estimates. Pages 95–100 *in* H. Quastler, ed. Information theory in psychology: problems and methods. Free Press, Glencoe, Ill.

Preston, F. W. 1948. The commonness and rarity of species. Ecology 29:254–283.

Seeley, T. D. 1982. Adaptive significance of the age polyethism schedule in honeybee colonies. Behavioral Ecology and Sociobiology 11:287–293.

Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal 27:379–423, 623–656.

Smith, A. 1776. An inquiry into the nature and causes of the wealth of nations. Whitestone, Dublin.

Theraulaz, G., E. Bonabeau, and J.-L. Deneubourg. 1998. Response threshold reinforcement and division of labour in insect societies. Proceedings of the Royal Society of London B 265:327–332.

Wilson, E. O. 1971. The insect societies. Harvard University Press, Cambridge, Mass.

Yockey, H. P. 1992. Information theory and molecular biology. Cambridge University Press, Cambridge.

Associate Editor: Joel S. Brown