On the Evolution of Multicellularity and Eusociality

Larry Bull
Faculty of Computer Studies
  and Mathematics
University of
the West of England
Bristol, BS16 1QY
U.K.
larry@ics.uwe.ac.uk

Abstract   In this article versions of the abstract NKC model are used to examine the conditions under which two significant evolutionary phenomena—multicellularity and eusociality—are likely to occur and why. First, comparisons in evolutionary performance are made between simulations of unicellular organisms and very simple multicellular-like organisms, under varying conditions. The results show that such multicellularity without differentiation appears selectively neutral, but that differentiation to soma (nonreproductives) proves beneficial as the amount of epistasis in the fitness landscape increases. This is explained by considering mutations in the generation of daughter cells and their subsequent effect on the propagule's fitness. This is interpreted as a simple example of the Baldwin effect. Second, the correspondences between multicellularity and eusociality are highlighted, particularly that both contain individuals who do not reproduce. The same process is then used to explain the emergence of eusocial colonies.

1   Introduction

Approximately 550 million years ago the Cambrian explosion brought forth all the major phyla of multicellular animals. However multicellularity is thought to have evolved up to 200 million years before that and has occurred at least three times—in fungi, plants, and animals. The correspondences between multicellular organisms and eusocial colonies have long been noted [29]. For example, the somatic cells of eukaryotes are genetically all very similar due to the way in which they are produced, as are the analogous nonreproducing individuals in most eusocial colonies. Indeed, it is this similarity that has been used to explain why selection at the level of the individual does not disrupt integration at the higher level in both cases ([19] and [11], respectively). The emergence of multicellularity and eusociality meant that individuals surrendered their ability to reproduce in favor of a close social existence and it is this phenomenon that is investigated here using an abstract model of (coupled) fitness landscapes.

Kauffman's [16] genetics-based NKC model, which allows the systematic alteration of various aspects of an evolving environment, is used to show that nonreproduction can prove beneficial when the complexity of the fitness landscape is increased. It is shown that multicellular-like individuals without differentiation appear to be selectively neutral in comparison to equivalent unicellular individuals, but that simple differentiation to nonreproduction (i.e., soma) can prove beneficial in terms of mean performance. One explanation for these results may be found by considering the Baldwin effect [2]. If the multicellular-like individuals are each considered as a selective whole, reproducing cells (gametes) can have their "true" genetic fitness altered by producing daughter cells that

are slightly different to them, via the background mutation. In this way natural selection can be guided toward better genetic combinations than those that already exist in the reproducing population—the Baldwin effect. That is, less fit mothers connected to fitter daughters can stand a higher chance of selection than equivalent mothers on their own when the differences between mother and daughter are produced by unavoidable mutations. Further, it is shown that the reasoning used to explain the emergence of nonreproduction in such multicellular individuals can be used to explain the emergence of eusocial-like colonies, since it is found that a simplified form of eusociality proves beneficial over sociality with reproduction as the (inter) organism complexity increases; the Baldwin effect can be used to suggest why individuals surrendered their ability to reproduce in early eusocial colonies.

The article is arranged as follows: The next section describes the natural phenomena of multicellularity and eusociality. Section 3 describes the NKC model of coevolution used throughout and in Section 4 the model is used to examine the emergence of multicellular nonreproduction, including results from a version of the model that considers aggregates of unicellular organisms. In Section 5 the emergence of eusocial nonreproduction is examined. Finally, all findings are discussed in Section 6.

2   Multicellularity and Eusociality: Nonreproduction

Most multicellular organisms have bodies consisting of differentiated cells, each containing roughly the same genetic material. The phenomenon exists in plants, animals, and fungi. This article is concerned with the conditions under which multicellularity could have emerged in a unicellular environment *before* functional differentiation evolved. Most traditional models of the emergence of such organisms assume functional differentiation in the daughter (e.g., improved feeding abilities [31]). Similarly, the other closely related computational model of multicellular origins, the latest version of Tierra (e.g., [23]), starts with differentiated multicellular organisms having already emerged and then allows evolution to adapt them to their environment. Recently these designed multithreaded/multicellular organisms have been sustained in the system [24]. The focus of the Tierra work is not to investigate how such organisms may emerge but the dynamics of what happens afterward, as was the case in the earlier work on self-reproduction.

Multicellularity is not strictly a eukaryotic phenomenon—it is also seen in prokaryotes through the aggregation of (unicellular) organisms. For example, the rod-shaped Myxobacteria usually live together in loose colonies pooling their digestive enzymes. When food supplies are exhausted or scarce they aggregate into a fruiting body, within which they differentiate, to produce spores that can survive hostile conditions; some individuals relinquish their ability to reproduce. Unicellular eukaryotes, such as *Myxomycota*, have similar life cycles. Eukaryotic green algae range from single-celled organisms (e.g., *Chlamydomonas*) to aggregates of a few cells (e.g., of the genus *Gonium*) to fully multicellular organisms with differentiation (e.g., *Volvox*), prompting the suggestion that multicellularity may have evolved from unicellular aggregates (e.g., [30]). The comparative efficiency of this form of multicellularity is also examined in this article.

Eusocial species are defined as those that form social groups containing sterile individuals from overlapping generations and that exhibit cooperative brood care. Examples are found amongst species of termites, ants, wasps, bees, aphids, mole-rats and spiders. Wheeler [29] was first in highlighting the close analogy between a eusocial colony and a multicellular organism in that both have differentiation into reproductives (gametes/queens) and nonreproductive specialists (soma/workers). The individual elements also share a number of basic properties such as self-nutrition, self-protection,

regeneration, and so on. The similarities between the aggregation form of multicellularity and eusociality have also been highlighted [9].

Hamilton's [11] kin selection is used to explain why eusocial offspring give up their right to reproduce. Briefly, a daughter has as many genes in common with its own offspring as its mother's and hence there is no selective difference between it raising its children or siblings. As noted in the introduction, Maynard-Smith and Szathmary [19, p. 8] have used Hamilton's theory to explain why reproductive differentiation occurs in multicellular organisms, following Wheeler's insight. That is, kin selection can also be attributed to somatic cells. This contrasts with Buss [6], who postulates propagule control of soma, and has been criticized by Michod [20], who suggests other factors, specifically cell-cell policing (similar to Buss) and germ-line segregation (reducing the potential for soma conflict), as being equally as important. Policing is also thought to have been important in the emergence of eusocial colonies (e.g., [8]).

Two possible evolutionary pathways to eusociality have been suggested: the subsocial route, in which the offspring of a single reproductive do not leave the nest and help rear siblings, and so forth; and the semisocial route, in which cooperative groups form and share duties, with some members not reproducing. It is suggested that the former could have occurred in ants (e.g., [15]) and the latter in wasps (e.g., [28]).

In this article the model used to examine the emergence of nonreproduction in multicellular-like individuals is extended to consider a number of scenarios (semisocial and subsocial) for the emergence of eusocial-like colonies. Previous computational models of evolving colonies (e.g., [7]) have used a cloning strategy to generate the workers and have not considered the emergence of eusocial, or eusocial-like, strategies. The abstract model of coupled fitness landscapes used to examine generic examples of these two phenomena is now introduced.

3 The NKC Model

Kauffman [16] (see Figure 1) introduced the NKC model to allow the systematic study of various aspects of multiorganism evolution. In the model an individual is represented by a haploid genome of N (binary) genes, each of whose fitness contribution depends upon K other genes in its genome (epistasis). Increasing K, with respect to N, increases the epistatic linkage, increasing the ruggedness of the fitness landscapes by increasing the number of fitness peaks, which increases the steepness of their sides and decreases their typical heights. Each gene is also said to depend upon C traits in the other individuals with which it interacts. The adaptive moves by one individual may deform the fitness landscape(s) of its partner(s). Altering C, with respect to N, changes the extent to which adaptive moves by each individual deforms the landscape(s) of its partner(s). As C increases, mean performance drops and the time taken to reach an equilibrium point increases, where the fitness level of the equilibrium decreases.

The model assumes all intergenome (C) and intragenome (K) interactions are so complex that it is only appropriate to assign random values (Gaussian distribution) to their effects on fitness. Therefore for each of the possible $K + C$ interactions, a table of $2^{(K+C+1)}$ fitnesses is created, with all entries in the range 0.0 to 1.0, such that there is one fitness for each combination of traits. The fitness contribution of each gene of a given genome is found from its individual table. These fitnesses are then summed and normalized by N to give the selective fitness of the total genome (the reader is referred to [16] for full details of the model).

Kauffman considered populations of one individual (said to represent a converged species) and mutation-based hill climbing to evolve each species in turn. That is, each species uses the current context of the others to determine progress. In this
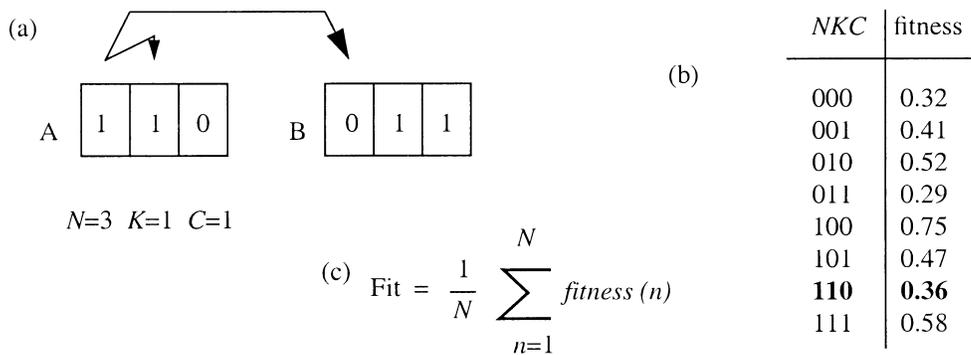
(a)

A | 1 | 1 | 0 | B | 0 | 1 | 1 |

$N=3$ $K=1$ $C=1$

(b)

NKC	fitness
000	0.32
001	0.41
010	0.52
011	0.29
100	0.75
101	0.47
110	**0.36**
111	0.58

(c) $\displaystyle \text{Fit} = \frac{1}{N} \sum_{n=1}^{N} \textit{fitness (n)}$

Figure 1. An example of the NKC model. (a) Each gene depends on one gene locally and on one in the other genome. Therefore there are eight possible allele configurations, each of which is assigned a random fitness as shown in (b). Each gene of each genome has such a table created for it. Total fitness of a given genome is the averaged sum of these values (c).

article a generational genetic algorithm (GA) [14] is applied to create a population-based synchronous version of the model.

4   Multicellularity

In this article the evolutionary performance of three types of simulated organism are compared: unicellular organisms, multicellular organisms without differentiation, and multicellular organisms with simple differentiation. Each type of simulated organism exists in a separate population such that there is no interbreeding between organism types. Within each population, each organism that survives the selection process is assumed able to divide once. In this way multicellular organisms consisting of two cells are compared to unicellular organisms.

Multicellular organisms are formed by a number of binding mechanisms. In higher plants the cells are connected via cytoplasmic bridges and exist within a rigid honeycomb of cellulose chambers. The cells of most animals are bound together by a relatively loose meshwork of large extracellular organic molecules (the extracellular matrix) and by adhesion between their plasma membranes. In all cases the cells exist as a larger whole during their lifetime. To take this into account for both types of multicellular organisms modeled here, an *average* (arithmetic mean) of the two cells' fitnesses is assigned to the reproductive cell(s). That is, it is assumed that the daughter and propagule affect each other's fitnesses and hence the average of their combined fitness is used for selection (this point will be returned to later because it is important). For example, the case of the bad mother/good daughter example mentioned in the introduction would represent a moderately fit single organism for selection.

Unicellular organisms: A population (size P) of unicellular organisms consists of $P/2$ individuals that are the offspring of the previous evolutionary generation (via selection and mutation) and $P/2$ individuals that are their offspring (via mutation); selection works over P individuals to produce $P/2$ offspring, each of which divides to create a population of P separate individuals. Each individual is evaluated on the given NKC function.

Nondifferentiated multicellular organisms: A population of nondifferentiated multicellular organisms consists of $P/2$ offspring from the last generation (via selection and mutation), each of which produces a connected daughter cell (via mutation). That is, P genomes exist in total. At the end of a generation *all* cells, both propagules and

daughters, are able to reproduce; again selection works over P genomes to produce $P/2$ offspring for the next generation. Both the mother and daughter are evaluated on the given NKC function and an averaged fitness is assigned.

Differentiated multicellular organisms: A population of differentiated multicellular organisms also consists of $P/2$ offspring from the previous generation, each of which produces a connected daughter cell. At the end of each generation selection works only on the initial $P/2$ propagules and *not* the daughter cells; the daughter cells are said to have differentiated to soma. Therefore selection works over $P/2$ individuals to produce $P/2$ offspring, but again P evaluations occur at each generation because the daughters are also evaluated on the given NKC function and an averaged fitness is again assigned.

4.1   Genetic Algorithm Model

A standard generational GA is applied to Kauffman's NKC model, using fitness proportionate selection ("roulette wheel") and mutation (recombination is not used). The organisms' environment is represented by an extra evolving population of unicellular organisms so that the effects of increased environmental pressure (C_e) can be examined; an individual from the environmental population (size $3P$) is picked to be evaluated with the current organism, where the organism is dependent upon the state of the environment. Various values of N have been tried with no significant difference in results being found. $N = 12$ and/or 24 are used throughout this section. All species have the same K value.

All experiments consist of running a generational GA over 2,000 generations, for 100 trials (10 runs on each of 10 NKC functions), for various parameters.

4.2   Results

Table 1 and Figure 2 show the general result for $P = 100$, mutation rate set at 0.01 per bit and $C_e = 1$, for various K. It can be seen that as the amount of intragenome epistasis (K) increases, the differentiated multicellular (d.multi) organisms do better in terms of mean fitness than both the unicellular (uni) and nondifferentiated multicellular (nd.multi) organisms. It can also be seen that there is no significant difference between the best and mean performances of the latter two types of organism. That is, the evolution of what is perhaps the first step in the process of becoming a differentiated multicellular organism, that of the propagule staying joined to the daughter cell, appears to be selectively neutral (this is returned to later).

Table 1 also shows the general result for the same model when the amount of dependence with the environment is increased ($C_e = 3$). It can be seen that the amount of intragenome epistasis must increase ($K > 4$) before the differentiated multicellular organisms again do better than the others. Kauffman [16, p. 249] also notes that increasing K improves performance under higher C conditions; equilibria are more readily encountered due to the increase in the number of peaks in the landscape, because individuals become stuck on optima, effectively stopping them from (much) further adaptation. The same effect can be achieved by altering the mutation rate (e.g., [4]).

The effects of altering the population size and mutation rate have also been examined. Increasing (e.g., $P = 200$) and decreasing (e.g., $P = 50$) the population size, with all other parameters as before, does not appear to have any effect on the general result reported above (results not shown). However, the unicellular and nondifferentiated organisms often do better when the mutation rate is decreased (e.g., 0.001 per bit), doing as well as the differentiated organisms in terms of mean performance and often better in terms of optima found (Figure 3). Results (not shown) are the same as those above for larger rates of mutation (e.g., 0.02 per bit). A second aspect of the evolution of multicellularity is now considered.
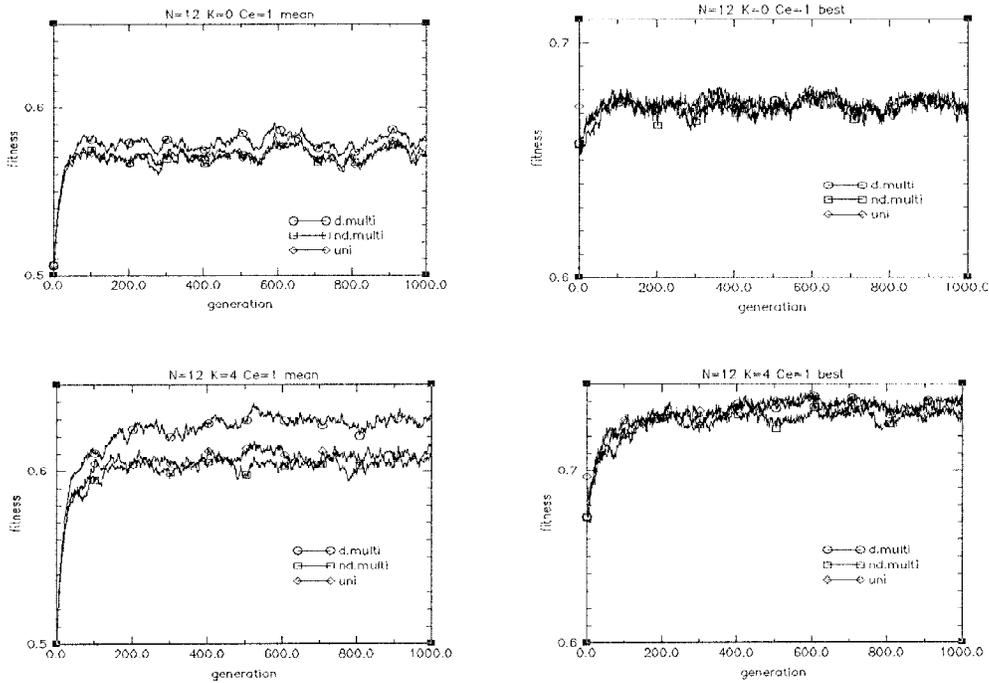
Figure 2. The (simply) differentiated multicellular organisms do better in terms of mean performance for increasing intragenome epistasis *K*.

Table 1. The effects of varying the amount of epistasis with regard to which cell configuration performs best in terms of finding optima (**b**) and mean population performance (**m**).

K	1	4	7	10
b	all=	all=	all=	all=
m	all=	d.multi	d.multi	d.multi
		$N = 12,\ C_e = 1$		

K	1	4	7	10	13	16
b	all=	all=	all=	all=	all=	all=
m	all=	d.multi	d.multi	d.multi	d.multi	d.multi
			$N = 24,\ C_e = 1$			

K	1	4	7	10
b	all=	all=	all=	all=
m	all=	all=	d.multi	d.multi
		$N = 12,\ C_e = 3$		

K	1	4	7	10	13	16
b	all=	all=	all=	all=	all=	X
m	all=	all=	d.multi	d.multi	d.multi	X
			$N = 24,\ C_e = 3$			

4.3 Multicellularity From the Aggregation of Unicellular Organisms

In the previous model the effects on fitness of two cells being joined together during their lifetime was considered by giving the propagule(s) an average of the two "true" fitnesses. However, the NKC model allows intergenome epistasis/dependence to be modeled explicitly. It is therefore possible to add a second parameter to the model (C_d) that considers the unavoidable interactions between the two cells, caused by cytoplasmic bridges, for example. Once this is done it is no longer possible to compare the two forms of multicellularity with their equivalent unicellular ancestors because the
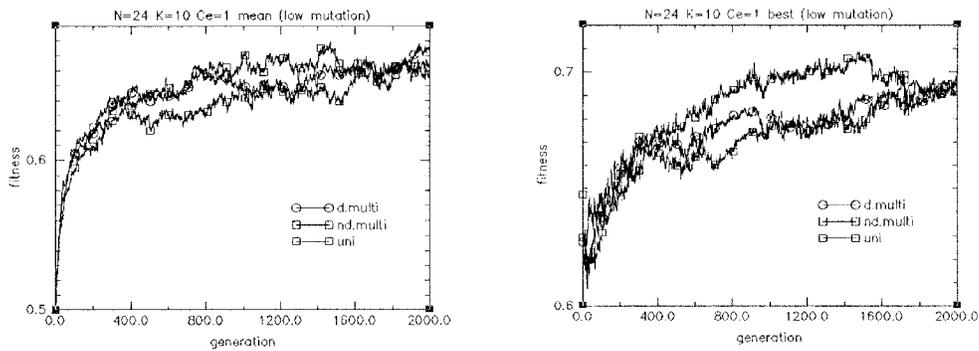
Figure 3. The effects of decreasing the rate of mutation on the different organisms.

fitness functions will be different—unicellular organisms would have tables of $2^{(K+C_e+1)}$ entries whereas multicellular organisms would have tables of size $2^{(K+C_e+C_d+1)}$. Comparing multicellular organisms at various levels of intercellular dependence to equivalent unicellular organisms is possible, however, if the latter are assumed to exist in a colonial form; the interdependence between the individuals in the aggregate, caused by cooperative feeding, for example, is modeled as the interdependence between the cells of the multicellular organism (C_d). In this section a unicellular aggregation consists of the initial individual and its offspring (other possible combinations are considered in Section 5).

4.3.1 Genetic Algorithm Model

As stated above, the parameter C reflects the effects of others on a given individual's fitness. However, for individuals that spend a significant amount of their life cycle in very close proximity, the effects of the given individual on the others' fitnesses should also still be considered in evaluating *its* fitness. That is, such a collective of individuals, whether multicellular or an aggregate, should be viewed as a functional whole for selection; an averaged fitness measure of all parts should be assigned to the reproductive individual(s). This reasoning is similar to that used in the discussion of close interspecies symbioses [1, 17], and is related to group selection [32] (see also Wheeler's original arguments [29]).

Here, as above, the reproductive cells of multicellular organisms are given an averaged fitness measure. For the equivalent aggregates of unicellular organisms both possible cases are considered; aggregate models using individual fitnesses are examined as well as averaged fitnesses. Note that unicellular aggregates with an averaged fitness measure are here equivalent to multicellular organisms without differentiation because mothers are partnered with their daughters.

4.3.2 Results

Table 2 and Figure 4 show the general result for $P = 100$, mutation rate set at 0.01 per bit and $C_e = 1$, for various K and C_d, in both types of aggregate. It can be seen that the differentiated multicellular organisms do better in terms of mean performance under most conditions with significant intergenome and intragenome dependence; the results correspond to those in the previous section. Again the difference between the unicellular organisms and nondifferentiated organisms is not significant. Increasing C_e also causes the advantages of multicellular differentiation to be lost, as before (not shown).

Table 2. The effects of varying the amount of epistasis with regard to which cell configuration performs best in terms of finding optima (**b**) and mean population performance (**m**) using aggregates of unicells.

C	K	1	4	7	10	C	K	1	4	7	10
1	**b**	all=	all=	all=	all=	1	**b**	all=	all=	all=	all=
	m	all=	d.multi	d.multi	d.multi		**m**	d	d.multi	d.multi	d.multi
3	**b**	all=	all=	all=	all=	3	**b**	all=	all=	all=	all=
	m	d	d.multi	d.multi	d.multi		**m**	d	d.multi	d.multi	d.multi
5	**b**	all=	uni=d	uni=d	all=	5	**b**	all=	all=	all=	all=
	m	d	d.multi	d.multi	d.multi		**m**	d	d.multi	d.multi	d.multi

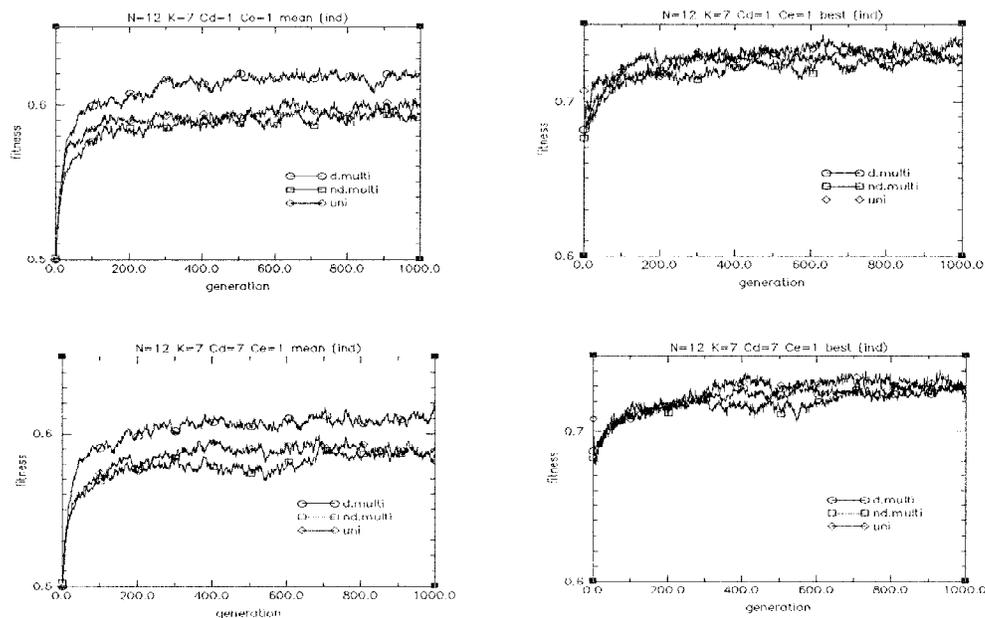$N = 12$, $C_e = 1$ Aggregate: individual $N = 12$, $C_e = 1$ Aggregate: averaged

Figure 4. The (simply) differentiated multicellular organisms do better in terms of mean performance for increasing intergenome epistasis in comparison to nondifferentiated organisms and unicellular aggregates (using individual fitnesses here).

The effects of altering the population size and mutation rate have also been examined for these two models. Increasing (e.g., $P = 200$) and decreasing (e.g., $P = 50$) the population size, with all other parameters as before, does not appear to have any effect on the general result reported above, except that for higher K ($K > 10$) both forms of unicellular aggregate sometimes do better in terms of optima found for bigger P (results not shown). The unicellular and nondifferentiated organisms, particularly the unicellular aggregates, again often do better than before (results not shown) when the mutation rate is decreased (e.g., 0.001). Results (not shown) are the same as those above for larger rates of mutation (e.g., 0.02).

4.4   Discussion

In the above models it has been found that multicellularity with a simple form of differentiation (to soma) proves beneficial when the fitness landscape is rugged. With the division error rate set low, a few selection-generated genomes are evaluated in an environment in which their own connected daughter is slightly different from them. The original genome receives the average of their combined fitnesses; the generated genome gets the average of its "true" fitness and a slightly different one. The effect of this is to alter the shape of the underlying fitness landscape for selection since a genome can do better, or worse, than it would on its own. That is, if a given genome can be mutated into a better daughter genome, it must be near a good combination of traits for the function/niche, and so the search is "guided" toward that region since the original genome operated on by selection gets a higher than expected fitness.

Previously Hinton and Nowlan [13] (and many others since) have shown how lifetime learning can achieve the same effect. They showed that a "needle in a haystack" problem can be solved more easily when individuals are given the ability to learn, since the learning can guide evolution to good gene combinations by altering the shape of the fitness landscape; evolution-generated genomes receive different (better) fitnesses than they would without learning because they are able to do local (random) search. They describe how learning can turn a difficult problem into a smooth unimodal function because genomes closer to the needle-like optimum are, on average, increasingly able to find it via random learning. This phenomenon is known as the Baldwin effect [2]. More specifically, the Baldwin effect is the two-stage process by which gene values not currently within a population can be learned and then assimilated [27] into the population in subsequent generations (via genetic mechanisms).

Mayley [18] has recently applied Hinton and Nowlan's model to a version of the model used here and found that learning becomes increasingly beneficial in terms of mean performance (only) with increasing K. As K increases, the number of optima in the fitness landscape increases. The ability to do local search (learning) becomes increasingly useful because it allows low fitness regions between optima to be "smoothed" over and hence crossed. Therefore the simple form of cell differentiation in multicellular organisms can be seen as an example of the Baldwin effect. The fact that considerably decreasing the mutation rate led to a loss of advantage supports this because less learning is possible in an effectively smaller gene pool; the differentiated organisms often did worse in terms of optima found.

The same fitness-altering effect also occurs in both the nondifferentiated multicellular case and in the aggregate unicellular case where a combined fitness is used. However, here if a low fitness individual is paired with a good one both receive an average fitness. This means that low fitness genomes look better for selection and will stand a higher chance of producing offspring than appropriate. The fitness-landscape-altering effect hinders the search process. Conversely, under multicellular differentiation, even when a good genome creates a less fit daughter, the genome that goes back into the population for selection is better than its fitness implies (although the search may be hindered/delayed somewhat by this phenomenon).

Therefore results here indicate that the evolution of differentiated multicellularity appears no more likely to have occurred via unicellular organisms living in aggregates (Section 4.3) than via those living alone (Section 4.1); the Baldwin effect is caused by the same phenomenon under both conditions and gives the same advantage. However, nondifferentiated multicellularity appears selectively neutral and therefore any collective advantage from the new aggregation, such as feeding, in a previously nonaggregating species may have given a selective advantage and started the progress to full multicellularity. The inclusion of costs due to joining with daughters also supports this scenario for the emergence of multicellularity because, if collective advantages through

aggregation already exist, the cost of joining would greatly reduce the chances of it giving any net benefit; the difference between unicellularity and nondifferentiated multicellularity is no longer neutral when costs are considered. Certainly, multicellularity from aggregation appears to be an evolutionary dead end [26].

The Baldwin effect through somatic mutation has also been suggested to occur in the immune system [12]. The use of recombination (not shown) in all of the above models does not alter the general result (after Szathmary [26], who suggests that meiosis was a necessary precursor to multicellularity). A version of the model used to determine the conditions under which multicellularity may have emerged through aggregation is now used to examine the origins of eusociality.

5 Eusociality

As stated in Section 2, a number of routes to full eusociality with sterile workers have been suggested. In this article the evolutionary performance of five types of colony formation are compared—two semisocial (sm) and three subsocial (sb). Colonies are said to consist of two individuals from overlapping generations.

- *Semisocial 1:* A population (size P) of reproducing individuals that come together randomly to form a colony. Each pair are evaluated on the given NKC function and then selection operates over all P individuals to produce P offspring for the next generation.

- *Semisocial 2:* A population of individuals as above, except that an *average* of the two partners' fitnesses is given to each because they are said to spend the significant part of their life cycle together (Sections 4.1 and 4.3.1).

- *Subsocial 1:* A population of reproducing individuals consisting of $P/2$ mothers from the previous generation (via selection and mutation), each of which produces one daughter (via mutation) to form a colony—that is, P genomes exist in total. At the end of each evaluation *all* individuals, mothers and daughters, are able to reproduce; selection works over P genomes to produce $P/2$ mothers for the next generation.

- *Subsocial 2:* A population of P individuals as above, except that an *average* of the two partners' fitnesses is given to each because, again, they are said to spend the significant part of their life cycle together.

- *Subsocial 3:* A population of $P/2$ individuals, each of which is generated from the previous generation (via selection and mutation) and produces an offspring (via mutation). At the end of each generation selection works only on the initial $P/2$ individuals and *not* the offspring; offspring are said the have become sterile caste workers. Again, an averaged fitness is given to the reproductive genome.

In comparison to the organisms modeled in Section 4.3, subsocial 1 is equivalent to unicellular aggregates with individual fitnesses, subsocial 2 is equivalent to nondifferentiated multicellularity (or unicellular aggregations with shared fitnesses), and subsocial 3 is equivalent to differentiated multicellularity. The two semisocial populations were not considered before because it is speculated that most individuals in a unicellular aggregation are from a common ancestor (e.g., [19, p. 214]). However, results here are applicable to the equivalent aggregation scenario. Unrelated pairs of female insects have been found in cooperative associations, for instance, in bees [25], although these have not been shown to persist over full life cycles [3, p. 83]; semisocial 1 is perhaps more common or realistic than semisocial 2.

Table 3. The effects of varying the amount of epistasis with regard to which cell configuration performs best in terms of finding optima (**b**) and mean population performance (**m**) using aggregates of unicells.

C \ K	1	4	7	10		C \ K	1	4	7	10
b 1	sm1	sm1	sm1	sm1		**b** 1	sm1	all=	sm1	sm1
m	sb3	sb3	sb3/sm1	sb3		**m**	all=	sb3	sb3	all sb=
b 3	all=	all=	all=	all=		**b** 3	all=	all=	all=	sm1
m	sb3	sb3	sb3	sb3		**m**	all sb=	all sb=	all sb=	all sb=
b 5	all sb=	all sb=	all sb=	all sb=		**b** 5	all sb=	all sb=	all sb=	X
m	sb3	sb3	sb3	sb3		**m**	sb3	sb3	all sb=	X

$$N = 12, \; C_e = 1 \qquad\qquad\qquad\qquad N = 12, \; C_e = 3$$

5.1 Genetic Algorithm Model

A standard GA is again applied to Kauffman's NKC model, using fitness-proportionate selection and mutation. The parameter C_e refers to the organisms' environment, represented by an extra evolving population of organisms (size $5P$), and C_d is the degree of dependence between the individuals; the aggregation model of Section 4.3 is used here. Various values of N have been tried with no significant difference in results being found. $N = 12$ and/or 24 are again used throughout this section. All experiments consist of running a generational GA over 2,000 generations, for 100 trials (10 runs on each of 10 NKC functions), for various parameters.

5.2 Results

Table 3 and Figure 5 show the general result for $P = 100$ mutation rate set at 0.01 per bit and $C_e = 1$, for various K. It can be seen that, for any intragenome epistasis (K), as the amount of intergenome dependence (C_d) increases, the subsocial 3 (fully eusocial) organisms do better in terms of mean performance, and as well as any other strategy in terms of optima found. There is no statistically significant difference between the other two subsocial strategies. These results match those of the equivalent experiments shown in Table 2, as expected.

5.3 Discussion

In the above model it has been found that eusociality with a simple form of caste differentiation (to worker) proves to be beneficial when the fitness landscape is rugged, due to the complexity of the colony. This result correlates with the findings of Section 4 in which simple differentiated multicellularity was found to be beneficial as complexity/epistasis increased. In both cases a founding individual is evaluated with an offspring that may differ slightly from itself and hence can have its genetic fitness altered in terms of selection, resulting in the potential to "smooth" complex fitness landscapes (Section 4.4); *the Baldwin effect can be seen to occur in both cases to the advantage of the organisms involved, even though the ability to reproduce is relinquished by offspring.*

When there is little dependence between members, temporal semisociality proves beneficial due, in part, to the larger amount of genetic mixing possible under the strategy. As colony member dependency increases, semisociality increasingly experiences
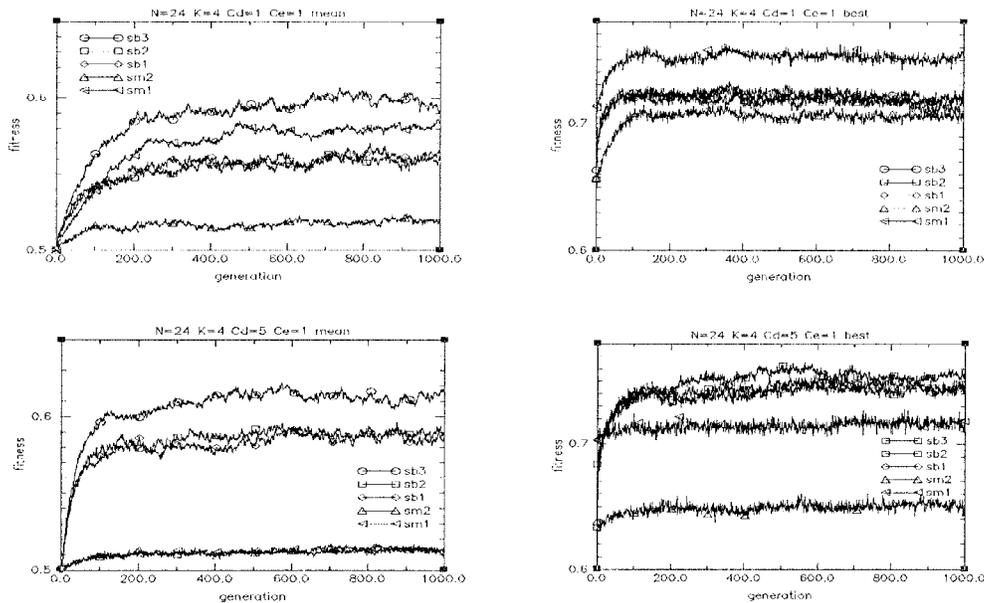
Figure 5. The eusocial (sb3) organisms do better in terms of mean performance and as well as any others in terms of optima found for increasing intergenome epistasis between the colony members, for any *K*.

the detrimental effects of partner variance, suffering the fitness landscape oscillations seen in between-species coevolution (Section 3). In subsociality offspring/partners are almost genetically identical to the foundress, allowing kin selection to operate and greatly reducing fitness landscape oscillations; intergenome dependence effectively becomes intragenome dependence, creating conditions that increasingly benefit from the Baldwin effect with increasing interdependence (also seen in Section 4.3). Permanent semisocialities perform least well due to the averaged fitness leaving individuals open to exploitation by less fit partners; as in sb2 and nondifferentiated multicellularity, evolution is hindered by less fit organisms looking better to selection than their genes indicate, only the effect is worse due to the lack of genetic similarity.

6   Conclusions

In this article the conditions under which simple multicellularity could have emerged have been examined. It has been found that multicellularity without differentiation to soma appears selectively neutral in comparison to equivalent unicellularity. However, for fitness landscapes of higher epistasis it has been found that the differentiation of daughter cells to soma proves beneficial in terms of mean performance. The Baldwin effect has been proposed as an explanation for this phenomenon. Further, the similarities between multicellularity and eusociality have been highlighted, leading to the same phenomenon being used to describe how eusociality could have emerged. Colony member dependence has been shown to be a critical factor.

   The two-celled organisms considered here are more like the early metazoa than plants, in that differentiation of the germ line is assumed to have occurred immediately. For early plantlike multicellular organisms it must have been possible for either the propagule or the daughter to become reproductive. A version of the model presented in Section 4 has been implemented, in which either cell randomly becomes the gamete

(results not shown). It was found that the advantage of the Baldwin effect is lost and that this form of multicellularity is also selectively neutral; plantlike differentiation did no better or worse than the unicellular and nondifferentiated multicellular organisms. This can be explained by considering the case where fitter mothers produce less fit daughters: If the less fit daughter becomes the propagule, it gets a moderate fitness for selection and so its less fit genome has a higher chance of reproducing than expected. Again, any advantage through increased numbers in a previously solitary species would tip the balance in favor of such multicellularity (Section 4.4).

The semisocial groups perform worst under higher interdependence C_d, particularly sm2, which always performs least well both in terms of optima found and population mean. This correlates with the above noted fact that semisocial nests have not been found to persist over time. The temporary semisocial nests (sm1) do well in terms of optima found under lower C_d but lose their advantage as partner dependence becomes significant ($C_d > 2$). These results support a scenario for the emergence of eusociality in wasps proposed in [9]. Table 3 also shows that increasing the effects of the environment decreases the benefits of eusociality (sb3) over the other two subsocial strategies, with the semisocial strategies performing roughly as before.

The effects of altering the population size and mutation rate have also been examined for this model. Increasing (e.g., $P = 200$) and decreasing (e.g., $P = 50$) the population size, with all other parameters as before, does not appear to have any effect on the general result reported above (results not shown). The subsocial strategies 1 and 2 often do better than before (results not shown) when the mutation rate is decreased (e.g., 0.001), however, matching eusociality (sb3). Results (not shown) are the same as those above for larger rates of mutation (e.g., 0.02).

The fact that only epistatic/complex organisms appear to benefit from multicellularity is potentially significant. Previously, it has been shown that the uptake of organelles by a cell reduces environmental instabilities and increases its epistasis [5]. This suggests that early (organelle carrying) unicellular eukaryotes were exactly the kind of organisms that would be able to take advantage of the type of differentiated multicellularity modeled here. Indeed, Maynard-Smith and Szathmary [19, p. 223] note that since the capacities for gene regulation and cell heredity were already present in (the genetically less complex) bacteria enabling, for example, the distinction between producing propagules and daughters, the evolution of multicellularity may have been limited by a lack of organelles. They suggest that key factors include the possibilities due to the cytoskeleton; the comparably more efficient photosynthesis/respiration of plasmids/mitochondria; and the apparent absence of any multicellular archaezoans.

Eusociality has emerged in the Hymenoptera (ants, bees, wasps, etc.) multiple times and it is suggested that their haplodiploidy (haploid males and diploid females) has been the major factor in this (e.g., see [10]). Briefly, it results in daughters being more related to each other than to their own offspring and so, it is said to predispose them to eusociality. However, the oldest eusocial organisms are the termites that are diploid. The results in this article do not conflict with the suggestion that haplodiploidy is beneficial to the evolution of eusociality but suggest that the Baldwin effect may have been key to its emergence; "the factors affecting the origin of eusociality could differ from those maintaining it" [3, p. 72]. This is also true for a number of other mechanisms suggested to have worked in synergy toward the establishment of eusocial colonies, for example, workers' advantages over solitary reproductives in raising broods [22]. The models presented in this article are now being extended to consider more than two participants.

References

1. Allee, W. C., Emerson, A. E., Schmidt, K. P., Park, T., & Park, O. (1949). (Eds.). *Principles in animal ecology*. London: Saunders.

2. Baldwin, J. M. (1896). A new factor in evolution. *American Naturalist*, *30*, 441–451.

3. Bourke, A. F. G., & Franks, N. R. (1995). (Eds.). *Social evolution in ants*. Princeton, NJ: Princeton University Press.

4. Bull, L. (1998). Evolutionary computing in multi-agent environments: Operators. In V. W. Porto, N. Saravanan, D. E. Waagen, & A. E. Eiben (Eds.), *Proceedings of the Seventh Annual Conference on Evolutionary Programming* (pp. 43–52). Berlin: Springer Verlag.

5. Bull, L. & Fogarty, T. C. (1996). Artificial symbiogenesis. *Artificial Life*, *2*, 269–292.

6. Buss, L. W. (1987). (Ed.). The evolution of individuality. Princeton, NJ: Princeton University Press.

7. Collins, R. J., & Jefferson, D. R. (1991). AntFarm: Towards simulated evolution. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial Life II* (pp. 579–602). Redwood City, CA: Addison-Wesley.

8. Frank, S. A. (1995). Mutual policing and repression of competition in the evolution of cooperative groups. *Nature*, *377*, 520–522.

9. Gadagkar, R. & Bonner, J. T. (1994). Social insects and social amoebae. *Journal of Biosciences*, *19*, 219–245.

10. Grafen, A. (1986), Split sex ratios and the evolutionary origins of eusociality. *Journal of Theoretical Biology*, *122*, 95–121.

11. Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, *7*, 1–52.

12. Hightower, R., Forrest, S., Perelson, A. S. (1996). The Baldwin effect in the immune system: Learning by somatic mutation. In R. K. Belew & M. Mitchell (Eds.), *Adaptive individuals in evolving populations* (pp. 159–167). Redwood City, CA: Addison-Wesley.

13. Hinton, G. E. & Nowlan, S. J. (1987). How learning can guide evolution. *Complex Systems*, *1*, 495–502.

14. Holland, J. H. (1975). (Ed.). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

15. Jaisson, P. (1985). Social behaviour. In G. Kerkut & L. Gilbert (Eds.), *Comprehensive insect physiology, biochemistry and pharmacology, Vol. 9, Behaviour* (pp. 673–694). Oxford: Pergamon Press.

16. Kauffman, S. A. (1993). (Ed.). *The origins of order: Self-organisation and selection in evolution*. New York: Oxford University Press.

17. Margulis, L. (1992). (Ed.). *Symbiosis in cell evolution*. New York: W. H. Freeman.

18. Mayley, G. (1996). The evolutionary cost of learning. In P. Maes, M. Mataric, J.-A. Meyer, J. Pollack, & S. W. Wilson (Eds.), *From Animals to Animats 4* (pp. 458–467). Cambridge, MA: MIT Press.

19. Maynard-Smith, J., & Szathmary, E. (1995). (Eds.). *The major transitions in evolution*. New York: W. H. Freeman.

20. Michod, R. E. (1996). Cooperation and conflict in the evolution of individuality. *Proceedings of the Royal Society of London B, Biological Sciences*, *263*, 813–822.

21. Nehaniv, C. (1998). (Ed.). *Mathematical and computational biology*. Providence, RI: American Mathematical Society.

22. Queller, D. C. (1989). The evolution of eusociality: Reproductive head start of workers. *Proceedings of the National Academy of Sciences USA*, *86*, 3224–3226.

23. Ray, T. S. (1997). Selecting naturally for differentiation. In J. R. Koza, K. Deb, M. Dorigo, D. B. Fogel, M. Garzon, H. Iba, & R. Riolo (Eds.), *Proceedings of the Second Annual Conference on Genetic Programming* (pp. 414–422). San Mateo, CA: Morgan Kaufmann.

24. Ray, T., & Hart, J. (1998). Evolution of differentiated multi-threaded digital organisms. In C. Adami, R. Belew, H. Kitano, & C. E. Taylor (Eds.), *Artificial Life VI* (pp. 295–304). Cambridge, MA: MIT Press.

25. Stark, R. E., (1992). Cooperative nesting in the multivoltine large carpenter bee *Xylocopa sulcatipes* maa: Do helpers gain or lose to solitary females? *Ethology*, *91*, 301–310.

26. Szathmary, E. (1994). Toy models for simple forms of multicellularity. *Journal of Theoretical Biology*, *169*, 125–132.

27. Waddington, C. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, 563–565.

28. West-Eberhard, M. J. (1978). Polygyny and the evolution of social behaviour in wasps. *Journal of Kansas Entomological Society*, *51*, 832–856.

29. Wheeler, W. M. (1911). The ant-colony as an organism. *Journal of Morphology*, *22*, 307–325.

30. Whittaker, R. H. (1969). New concepts of kingdoms of organisms. *Science*, *163*, 150–160.

31. Wolpert, L. (1990). The evolution of development. *Biological Journal of the Linnean Society*, *39*, 109–124.

32. Wynne-Edwards, V. C. (1962). (Ed.). Animal dispersion in relation to social behaviour. New York: Hafner.