Kin-Selection: The Rise and Fall of Kin-Cheaters

Sherri Goings*, Jeff Clune*†, Charles Ofria and Robert T. Pennock

Michigan State University, East Lansing, USA

*Both authors contributed equally to this paper
†Article Correspondence: jclune@msu.edu

Abstract

We demonstrate the existence of altruism via kin selection in artificial life and explore its nuances. We do so in the Avida system through a setup that is based on the behavior of colicinogenic bacteria: Organisms can kill unrelated organisms in a given radius but must kill themselves to do so. Initially, we confirm results found in the bacterial world: Digital organisms do sacrifice themselves for their kin—an extreme example of altruism—and do so more often in structured environments, where kin are always nearby, than in well-mixed environments, where the location of kin is stochastically determined. Having shown that helping one's kin is advantageous, we turn our attention to investigating the efficacy and implications of the strategies of kin-cheaters, those who receive help from kin but do not return it. Contrary to the expectations of current theory, we find that kin-cheaters outcompete kin-altruists. Our results cause us to question the stability of strategies that involve altruism between kin. Knowing that kin-altruism persists in biological systems, however, we search for, and find, conditions that allow kin-based altruism to persist in evolving systems despite the presence of kin-cheaters.

Introduction

At first glance, the persistence of altruistic behavior (defined as helping other organisms at a net cost to the acting organism) is puzzling from an evolutionary perspective. However, in nature we frequently see acts that seem altruistic. Many theories exist that help explain such apparent acts of altruism. They usually demonstrate that the acting entity, when properly identified, actually receives a net gain by the apparently altruistic behavior. A prominent theory is kin selection, where an individual suffers or risks a net cost to help its kin (Hamilton 1963). The theory of kin selection recognizes the gene as the acting agent and argues that a certain altruistic gene (e.g. gene A) is actually helping out copies of itself by causing an individual that possesses it to act altruistically toward that individual's kin (since the individual's kin is likely to have gene A) (Dawkins 1976). In this paper we demonstrate the existence of kin selection in the Avida digital evolution system and investigate an aspect of the theory that is often unexamined; whether kin that cheat on their relatives are favored by natural selection. We accomplish this by comparing organisms that are altruistic

towards close relatives (kin-altruists) and those that are only altruistic towards identical copies of themselves (clone-altruists). We refer to these clone-altruists as 'kin-cheaters' to emphasize that they are not altruistic towards non-identical kin even though such non-identical kin are altruistic towards them. Most theorists have overlooked the potential success of this type of cheater. We investigate whether the assumption generally made by theorists that kin-cheaters should not be selected for is a valid one.

It is often helpful to look at extreme cases when investigating a theory. The extreme act of altruism is giving one's life for another. For this reason we chose to study the evolution of this behavior as our means of studying kin selection in artificial life. A further reason is because kin-selection driven by an organism sacrificing its life has been well studied in the bacterial world (Chao and Levin 1981). Checking our results against these findings enables us to confirm that the computational system we are using to investigate evolutionary phenomena is behaving similarly to the biological systems to which we wish to extrapolate our findings.

A small percentage of colicinogenic bacteria will produce a toxin until they explode, releasing the toxin into the surrounding area. This toxin is harmful to those that are not immune to it. Since colicinogenic bacteria are immune to this toxin, those harmed by it are typically non-kin. This strategy has been likened to that of 'suicide bombers' (Lenski and Velicer 2000). Chao and Levin found that this trait is more likely to be beneficial in structured environments, where one's kin are next to one, versus well-mixed environments, such as a liquid culture, where resources and spatial location are randomized. They attribute the greater benefit in structured environments to the increased likelihood that the extra resources provided by killing non-kin will be received by kin due to their frequent proximity, thus differentially (and "selfishly") helping those likely to share one's genes.

We modified the Avida digital evolution system to facilitate the study of kin-altruism. To do this we added to Avida the element that makes colicinogenic bacteria ideal for studying kin-altruism: the ability of organisms to kill a number of non-kin in their surroundings by sacrificing their lives. Initially, we find that this strategy is overwhelmingly selected for. We then replicate some of the findings of Chao and Levin: sacrificing one's life to aid one's kin is more effective in structured environments than in well-

mixed (randomized) environments. We further investigate whether the presence of kin-cheaters can prevent kin-based altruism from being a stable evolutionary strategy, and show that

1) Kin-altruists thrive in the absence of kin-cheaters.
2) Once kin-cheaters arrive on the scene, they outcompete kin-altruists, raising the question of how the altruism predicted by kin-selection theory persists in nature.
3) Limiting factors on the destructive power of kin-cheaters exist, allowing kin-based altruism to persist.

Methods

All of the following experiments use the Avida digital evolution system, a virtual environment in which digital 'organisms' evolve through random mutation and natural selection (Ofria and Wilke 2004). At the start of each 'run' of Avida (one execution of the software program), a virtual world is seeded with a digital organism that can self-replicate. Each digital organism has a sequence of instructions considered to be its genome. Self-replication involves copying this genome and then dividing into two child organisms. The copy process is imperfect, however, so each instruction has a chance of mutating to any random instruction when copied. These organisms quickly fill up the virtual environment and compete for a limiting resource: SIPS (Single Instruction Processing units). These are the basic unit of energy available to an organism. Since organisms need this energy to execute their genomes, and thereby replicate, those that earn more SIPS (by performing tasks) or use less (via efficiency) will tend to be selected for naturally. As Daniel Dennett says, "evolution will occur whenever and wherever three conditions are met: replication, variation (mutation), and differential fitness (competition)" (Dennett 2002). The Avida system includes these conditions and is thus a tractable system that we can use to investigate the general properties of any evolving system (Lenski et al. 2003).

We modified the system by adding an instruction called `explode` to the list of instructions an organism is able to execute. An organism that executes this instruction will probabilistically "explode", killing itself and emitting virtual toxins. The percent probability that executing the `explode` instruction will cause an organism to explode varies from organism to organism depending on its genotype (see detailed methods). For example, one organism may explode 90% of the time it executes `explode,` whereas another may do so only 2% of the time. A probabilistic approach is necessary for otherwise all or none of a lineage would explode. If the instruction does not cause the organism to explode, it has no effect except to use one of the organism's allotted SIPS. If the organism does explode, it kills itself and all non-kin organisms within a given explosion radius. The kin/non-kin distinctions are made based on the genetic (Hamming)

distance between the exploding organism and those inside this radius. Whether an exploding organism considers another its kin depends on whether the Hamming distance (number of genomic differences) between the organisms is less than or equal to the exploding organism's Hamming distance threshold (HDT). This Hamming distance threshold can be different for different organisms and is set before the run, unless otherwise specified. The radius parameter gives the distance an exploding organism can propel its fatal toxins when exploding. The radius was set to two for all experiments except for baseline runs, meaning an exploding organism affects the 24 organisms surrounding it. This parameter is also set at the start of each run and does not change during the course of a run. A population size of 3,600 organisms and a genomic mutation rate of .2 were used for all runs unless otherwise noted.

Experiments and Results

Evolution of Kin-Altruism

Once we had modified the system to allow the `explode` instruction to mutate in, we were interested in seeing if it would be selected for. We set a global HDT of 0, so that an exploding organism kills all organisms with non-identical genomes in the blast radius of 2. We performed twenty runs seeded with the default organism that does not have the `explode` instruction in its genome and thus never explodes. In every run (20/20) the number of explosions went up substantially (fig. 1). In these structured runs, when organisms replicate, their children are placed into neighboring cells (killing the organism that previously existed in that cell). The killing of non-kin by an exploding organism thus eliminates organisms that had the potential to kill the exploding organism's kin (if any such kin were nearby). This benefit conferred on an exploding organism's kin was large enough to outweigh the cost of death, as evidenced by the strong selection for this behavior in our initial twenty runs. In order to assure that this selection was due to the benefit provided to kin, we performed two different baseline tests that eliminated this benefit. In the first set of baseline runs we set the radius to zero so that an exploding organism killed only itself. In the second set of baseline runs the exploding organism killed every organism in a radius of two, kin and non-kin alike. In both of these cases, the frequency of explosions did not rise above the minimal level that is to be expected by the high rate at which the `explode` instruction mutates in. The clear difference between the experiment and the control runs can be seen in figure 1.

In studying colicinogenic bacteria, Chao and Levin found that sacrificing one's life to benefit one's kin is a more effective strategy in structured environments, where kin remain near one another, than in well-mixed environments, where the locations of kin and resources are continuously randomized (Chao and Levin 1981). They

reason that the benefits of sacrificing one's life—more available resources—are more likely to be shared by one's kin in structured environments than in well-mixed environments. The same reasoning should apply to our setup: The benefit of sacrificing one's life—eliminating those that can kill one's kin—will only be of benefit if one's kin are nearby when one explodes. Chao and Levin's experiments focused on whether small colonies of colicinogenic bacteria were more likely to *invade* non-colicinogenic bacteria in these two different environments. We took a slightly different approach, looking at whether strategies involving sacrificing one's life were more likely to *evolve* in these disparate environments. Having already shown that such strategies evolve in structured environments (see above), we did a set of 20 runs in a well-mixed environment that were otherwise identical to the first set. Our findings support Chao and Levin's results and explanation: the strategy of sacrificing oneself for one's kin did not evolve in any (0/20) of our well-mixed runs (fig. 1).
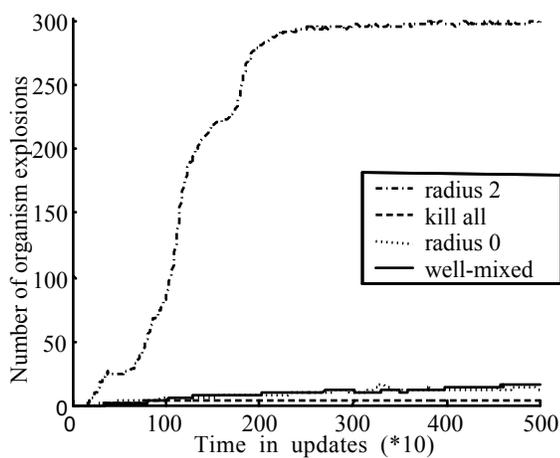

Figure 1: Number of organism explosions per 100 updates averaged over all 20 runs. The "…." line represents one base-line experiment in which an exploding organism kills only itself. The "----" line signifies the other base-line run in which an exploding organism kills itself and all surrounding organisms, regardless of kinship, within a radius of 2. The ".-.-" line represents the first experiment in which an exploding organism kills itself and all surrounding non-kin within a radius of 2. These three treatments all occurred in a structured environment. The "—" line represents the second experiment: It is identical to the first except that it occurs in a well-mixed environment.

While this does not show that such behavior cannot evolve in well-mixed environments (for it would surely evolve if the blast radius were drastically increased) these results do show that it is much less likely for the strategy to evolve in

well-mixed environments, all else being equal[1]. Note that this also reinforces our conclusion above that the reason this strategy is evolving is due to the benefits it confers on kin. It should be noted that, as far as we know, organisms in natural systems are unable to determine the exact genetic distance between themselves and other organisms. Instead, they tend to use whatever reliable indicators of kinship are available (e.g. spatial proximity, smell, physical similarity, etc.), and rely on the fact that close kinship and genetic similarity are typically correlated. While this difference in information may mean that natural organisms will be less able to implement beneficial strategies effectively, we do not believe this difference will substantively change the underlying strategies themselves. Therefore it should not preclude us from extrapolating our findings from the digital organisms in our setup to organisms in nature.

The Problem of Kin-Cheaters

Having shown that it is sometimes advantageous to give one's life to aid one's kin, we turned our attention to the problem of kin-cheaters (organisms that receive benefits if a relative sacrifices its life but will kill this relative if they explode). To analyze the effect of these cheaters we started by assessing what occurs in their absence. We wanted to test if, in the absence of kin-cheaters, it is advantageous to be less than maximally discriminating about who one considers to be kin. To translate the question into the language of sexual species, is it more beneficial to be maximally discriminatory, and thus only help identical twins, or should one be less discriminatory, and thus help brothers, nieces, cousins, second cousins, etc.?

Efficacy of Kin-Altruism: Organisms in the original runs (described above) can be considered maximally discriminatory because they killed anyone with a non-identical genotype when they exploded. To test whether less discriminatory strategies would invade, we seeded 50 runs with two groups of organisms; one maximally discriminatory and the other non-maximally discriminatory. Logistically, organisms from these two groups differed only in two respects. The first difference was the Hamming distance threshold (HDT) setting in their `explode` instruction (see methods). The maximally discriminatory organisms had HDTs set to 0 (HDT.0) and thus killed all non-identical genomes. The less discriminatory organisms were set to HDT.5, and thus killed all creatures with a genetic distance of 6 or greater (i.e. more than 5 of the 100 instructions in an affected organism's genotype differed from the exploding organism). The second difference was the changing of some of the neutral nop-x instructions to effectively neutral nop-a instructions, to create a genetic distance of 12 between organisms of the two groups, ensuring that each

---

[1] Chao and Levin found that this strategy is less likely to invade in well-mixed environments, where it only invades if it begins above a certain frequency, versus in structured environments, where it can invade no matter how small its initial frequency (Chao and Levin 1981).

type considered the other to be non-kin at the outset. 95% of the initial population consisted of maximally discriminatory organisms and 5% were less discriminary. In all 50 runs the more altruistic lineage, which began as a small minority, invaded and went to fixation. From this we conclude that less discriminatory (more altruistic) strategies are more beneficial in the absence of kin-cheaters than maximally discriminatory (selfish) strategies.

Efficacy of Kin-Cheating: Where one finds altruism, one typically finds cheating. With regard to altruism amongst kin, however, evolutionary theorists largely ignore the existence of kin-cheaters. While the strategy of faking that one is indeed related to a kin-altruist is often discussed, as in the case of the cuckoo, rarely does the literature on kin selection probe the idea that a true relative may not be reciprocating the generosity of its brethren. The main reason for this seems to be the persuasive argument that a gene is better off helping copies of itself in another organism and thus would be worse off by cheating (see, for example, Dawkins 1982). It has been argued, however, that the cheating phenomenon should be no less prevalent for genes than individuals: any gene that receives help from its kin but does not reciprocate should do better off compared to its altruist relatives (Sober and Wilson 1999). We used our setup to evaluate these competing theories by testing the efficacy of a kin-cheating strategy.

In our setup, kin-cheaters are those who, if they explode, will kill less discriminating relatives but will not be killed by such relatives (because the less discriminating relatives consider them kin). It should be noted that we are stretching the concept of 'cheater' a bit. Cheaters in our setup are altruistic towards some organisms (namely identical copies of themselves), but can be considered cheaters because, when interacting with more altruistic relatives, they receive a benefit but do not return it. To test whether such kin-cheaters would invade kin-altruists, we again ran 50 runs seeded with two groups of organisms; one maximally discriminatory (HDT.0) and one less discriminatory (HDT.5). This time, however, the initial genetic (Hamming) distance between the organisms of the two groups was one. This means that organisms in the less discriminatory group are altruists that consider cheaters kin—and thus do not kill them when exploding—but the maximally discriminatory organisms are kin-cheaters that kill altruists when exploding. The population was seeded with 95% kin-altruists and 5% kin-cheaters. In 45 out of the 50 runs the kin-cheater lineage went to fixation. Whether the two are related, then, makes a significant difference (p-value < .001 using Fisher's exact test). Figure 2 represents two of these runs; one where the kin-cheater went to fixation and one where it did not.

Conditions that Enable the Success of Kin-Altruists in the Presence of Kin-Cheaters: This result is interesting for two reasons. Initially, it shows that in most cases kin-cheaters eliminate kin-altruists. This result is consistent with Sober and Wilson's argument that kin-cheaters are at least a phenomenon that kin selection theory needs to account for. The widely held assumption that organisms

that cheat on their kin will be worse off, all else being equal, seems incorrect. Also interesting is the fact that kin-altruists win 10% of the time (5/50). How is the destructive power of kin-cheaters countered, here or in nature where we see kin altruism all the time? Sober and Wilson propose one theory, a group selectionist account, that requires kin-groups that do not have cheating members to outperform those that do. This cannot be the explanation for why kin-cheaters failed to fixate in these five runs, however, since our setup does not include the differential survival of multiple groups. Thus, whether or not one accepts Sober and Wilson's explanation of one force that may mitigate kin-cheaters, there appears to be at least one more force at work that manifests itself in our setup.

We watched a few of the runs in which altruists win and ran a few more in which the organisms were allowed to change their Hamming distance threshold via mutation
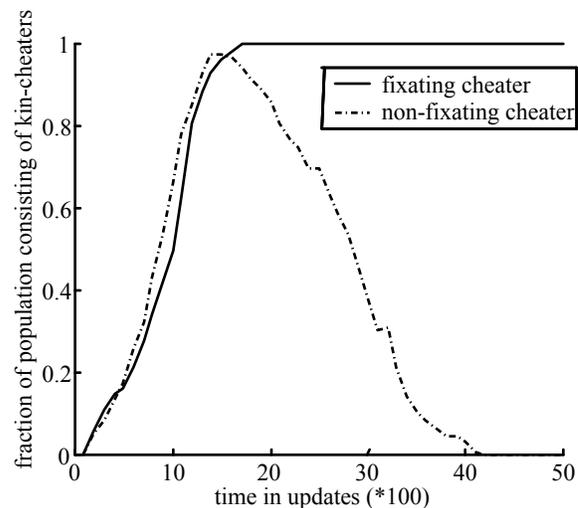

Figure 2: Two sample runs seeded with 95% kin-altruistic organisms (HDT.5) and 5% kin-cheaters (HDT.0) with a genetic (Hamming) distance of 1 between the two groups. In one of the displayed runs (solid line) the kin-cheater quickly fixates. This occurred in 45/50 runs. In the other run (dotted line) the kin-cheater quickly expands in the population, coming very close to fixation, but is stopped at the last moment by the kin-altruist, which has mutated away until it no longer considers the kin-cheater to be its relative, and thus is no longer altruistic towards it. This phenomenon occurred in 5/50 runs.

during the course of the run. We observed two independent, non-concurring phenomena that prevent cheaters from going to fixation in the population. The first cause occurs when a group unrelated to the cheaters evolves an altruistic strategy of self-sacrifice with a HDT greater than 0. This group outcompetes the cheaters because, as shown in our earlier experiment, a less discriminatory group of organisms will outcompete an unrelated group of maximally discriminatory organisms.

The second cause is if the diversity within the kin-altruist group reaches a point at which one section of the group is not willing to be altruistic toward the cheater because the genetic distance between the organisms in that section of the group and the cheater is greater than 5 (fig. 2). These two causes boil down to the same principle: The cheater will not go to fixation if it encounters a group before it fixates that is more altruistic than it but is not altruistic *towards* it (i.e. does not consider it to be kin). We hypothesize that there are a number of different environmental conditions that will increase the frequency of such an encounter occurring, including larger population sizes, decreased fitness of the cheaters (they will fixate less quickly), higher mutation rates, and the degree of altruism (how great a Hamming distance threshold) in kin-altruist groups (the higher the Hamming distance threshold the longer it takes to reach sufficient diversity)[2]. We tested one of the predictions of this hypothesis by varying the population size. Kin-cheaters had gone to fixation in 45 of the 50 runs with populations of 3,600 organisms. We performed 50 more runs with populations of 10,000 organisms. Lending strong support to the hypothesis, cheaters fixate in only 2 of the 50 runs (p-value < .001 using Fisher's exact test). Figure 3 summarizes these results.

	More Altruistic (Less Discriminatory)	More Selfish (More Discriminatory)
Unrelated	50/50	0/50
Related	5/50	45/50
Related Large Population	48/50	2/50

Figure 3: Summary of three experiments where two different groups of organisms competed. The data report the number of runs out of 50 in which that group went to fixation, eliminating the other group entirely. More selfish organisms kill all but their closest kin (HDT.0). More altruistic organisms are less likely to kill distant relations (HDT.5). When the two groups of organisms are unrelated, the more altruistic—or less discriminatory—group wins. When the two groups are related, the more selfish—or maximally discriminatory—variant is a 'kin-cheater.' Kin-cheaters will outcompete their more altruistic relatives under certain conditions but are less likely to do so in larger populations.

---

[2] This last force will rarely be significant in anything but the smallest of groups or at very low mutation rates, as group diversity will reach the Hamming distance threshold rather quickly otherwise.

The two types of events that check the ability of kin-cheaters to fixate are both interesting for different reasons. If the first type were the only mitigating force, where kin-cheaters are successful at outcompeting all of their relatives but lose when they encounter a non-related band of more altruistic organisms, kin-altruism would be individually unstable even though it could be a persistent feature within the population. In other words, every time kin-altruism evolves, its level of discrimination will continuously be reigned in by more discriminatory offspring until it ends up as maximally discriminatory. Under these harsh conditions, non-maximally discriminatory altruism could only be persistent in the population if such altruism evolved with a high enough frequency that there were always groups that had not yet transitioned to a maximally discriminatory strategy. While theoretically possible, such a setup in the natural world is unlikely and does not describe the more successful versions of kin-altruism we see in biological systems.

The second class of events relates more to the world we live in. In these situations, a cheater evolves within a group of related altruists and spreads until it encounters the border at which those that are related to it are no longer willing to give it preferential treatment. An analogy could be drawn to the human cultural practice of treating distant relations equal to strangers. One may find an open palm when seeking a loan from brothers, cousins, and even second cousins. At some genetic distance, however, the fact that relatives are treated on par with strangers means that one must secure a loan on merit instead of nepotism. This practice, whether conscious or not, limits the size of the area any given kin-cheater can exploit.

Conclusions

We evolved kin selection in the Avida digital evolution system using a setup similar in respects to colicinogenic bacteria. Organisms could sacrifice themselves and, in the process, kill non-kin greater than a certain genetic (Hamming) distance away. We further investigated the question as to whether kin-cheaters would outcompete their kin-altruist relatives. We show that without cheaters a more altruistic (less discriminatory) strategy outcompetes a more selfish (more discriminatory) strategy. With kin-cheaters, however, more discriminatory strategies are selected for. Unchecked, and assuming that cheaters will eventually arise, this tendency of more discriminatory strategies to outcompete less discriminatory strategies should prevent the ability of anything but maximally discriminatory types of altruism from persisting. These implications, the possibility of which have been overlooked by most theorists, make necessary an explanation for how altruism amongst kin persists in nature. We observe one candidate explanation: The possibility that kin-cheaters will run into altruists they cannot exploit serves as a check on their tendency to eliminate altruism towards distant kin. We further show that many conditions exist that make this possibility quite likely, providing one potential way to

explain how non-maximally discriminatory altruism amongst kin endures.

Detailed Methods

All of the experiments were performed using version 2.0 beta7 of the Avida software. The default organism used in our experiments consisted of a 15-instruction long copy loop that performs self-replication and 85 nop-x instructions that perform no function when executed. The length of the organism is fixed at 100. An Avida organism consists of its "genome" (sequence of instructions), 2 stacks (only one of which is active at any given time), and 3 registers (A, B, and C). Unless otherwise noted, the default instruction set was used with the addition of the `explode` instruction. This set contains instructions for self-replicating, performing logic operations, and manipulating numbers in the stacks and registers. The added `explode` instruction, when executed, inputs the number in the organism's register A by default, although if it is followed by a nop-B or nop-C it will input the number in the specified register (B or C) instead. The instruction then mods this number by 100 to get the percent chance that the organism will explode. This percentage is then compared to a random number and, if the percentage is greater than or equal to the random number, the organism explodes (otherwise it continues executing its code).

The organism also inputs the number on top of the currently active stack and sets the Hamming distance threshold (HDT) parameter to this value. The Hamming distance between two organisms is determined by comparing their genomes site by site and totaling the number of differences found. The Hamming distance between two organisms is therefore the genetic difference between those organisms. An exploding organism with a HDT of 0 will kill all non-identical organisms; with a HDT of 1 it will kill any organism with more than one non-identical site, etc. All experiments use organisms with fixed lengths so no alignment is necessary before the determination of Hamming distances. In the first set of experiments (those researching the evolution of self-sacrifice) the HDT was fixed at 0 for all organisms. In the second set of experiments (the competitions), the HDT is set at 5 for one group of organisms and 0 for the other for the duration of the run. The units of the radius parameter are cells on the environment grid, so an organism exploding with a radius of 2 will affect the 24 organisms that are no more than 2 cells away. In the base run in which the radius is set to 0 the organism kills only itself.

The mutation rate used is .002 per instruction copied, which equals .2 mutations per genome, unless otherwise noted. The population consists of 3,600 organisms placed on a 60x60 grid, unless otherwise noted. Two birth methods were used: well-mixed (a.k.a. randomized, mass-action, mass-habitat, liquid, well-shaken, etc.) and neighborhood (structured). In well-mixed environments a new organism is placed in a cell randomly chosen from the entire population. In neighborhood environments a new organism is placed in a cell randomly chosen from the 8 cells in the 3x3 square surrounding the parent organism. In both methods the organism currently occupying the chosen cell is killed.

References

Adami, C. (1998). *Introduction to Artificial Life*. Santa Clara: TELOS Springer-Verlag.

Adami, C., Seki, R., and Yirdaw, R. (1998). Critical exponent of species-size distribution in evolution. In Adami, C., Belew, R., Kitano, H., and Taylor, C. editors, *Proceedings of "Artificial Life VI,"* p. 221-227. Cambridge, MA: MIT Press.

Chao, L. and Levin, B. R. (1981). Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proc. Natl. Acad. Sci. USA* 78:6324-6328.

Dawkins, R. (1982). *The Extended Phenotype*. Oxford and San Francisco: Freeman.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dennett, D. (2002). The new replicators. Pp. E83-92 in Pagel, M., ed. *Encyclopedia of Evolution*. New York: Oxford University Press.

Hamilton, W.D. (1963). The evolution of altruistic behavior. *American Naturalist* 97:354-356.

Lenski, R. E., Ofria, C., Pennock, R. T., and Adami, C. (2003). The evolutionary origin of complex features. *Nature* 423:139-144.

Lenski, R. E. and Velicer, G. J. (2000). Games microbes play. *Selection* 1:51-57.

Ofria, C. and Wilke, C. O. (2004). Avida: a software platform for research in computational evolutionary biology, *Artificial Life* 10:191-229.

Sober, E. and Wilson, D. S. (1999). *Unto Others*. Cambridge, MA: Harvard University Press.