On Coevolution: Asymmetry in the NKCS Model

Larry Bull

Department of Computer Science & Creative Technologies

University of the West of England, Bristol UK

Larry.Bull@uwe.ac.uk

Abstract

The NKCS model was introduced to explore coevolutionary systems, that is, systems in which multiple species are closely interconnected. The fitness landscapes of the species are coupled to a controllable amount, where the underlying properties of the individual landscapes are also controllable. Previous work has assumed symmetry with respect to the controlling parameters. This paper explores the effects of reducing that symmetry on the behaviour of the coevolutionary system, including varying genome complexity, the degree of landscape coupling, and the use of local learning. Significant changes in behaviour from the traditional model are seen across the parameter space. These findings are suggested as particularly pertinent to symbiotic relationships.

Keywords: Baldwin effect, evolution, fitness landscape, mutation, symbiosis.

1. Introduction

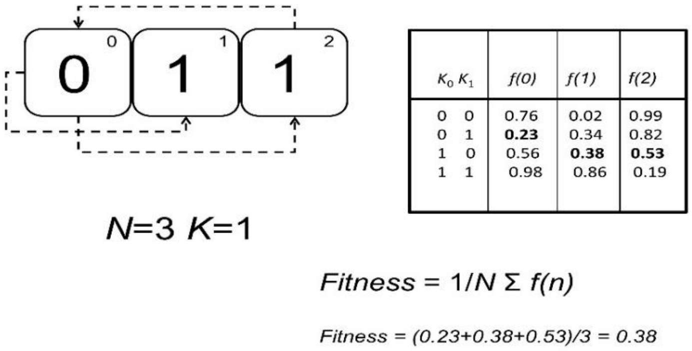
Coevolution is ubiquitous in natural systems and widely used in artificial systems. Almost thirty years ago, Kauffman and Johnsen [1992] presented a tuneable, abstract model through which to explore some of the basic properties of such systems – the NKCS model. Extending the well-known NK model [Kauffman and Levin, 1987] of rugged fitness landscapes, the NKCS model couples multiple NK landscapes to study the evolutionary dynamics of ecosystems containing multiple species. Versions of the model have also been applied to artificial systems, including management science (e.g., [Levitan et al., 1997]), receiver-based communication optimisation [Kauffman, 1995], and engineering design (e.g., [Preen et al., 2019]).

In almost all known instances, it has been assumed that the characteristics of the underlying NK fitness landscapes are the same, i.e., the model is symmetrical. This is clearly a simplification. For example, the NKCS model has previously been used to explore symbiosis, in particular the formation of hereditary (endo)symbioses where the relationship between species is maintained across generations [Bull & Fogarty, 1996]. However, these intimate symbioses are typically formed between species of very differing complexity, such as bacteria and multicellular eukaryotes. Further consideration of this difference, in the form of the relative speed of replication between the species involved, significantly altered the conditions under which such relationships may be expected to emerge [Bull, 2010] (see [Bull, 2020] for an overview). Heterogeneous complexity has been shown to be important within artificial systems as well, such as in competitive coevolutionary optimisation (e.g., [Stanley & Miikkulainen, 2004]).

This paper explores the effects of asymmetric complexity within the NKCS model, considering landscape ruggedness, the degree of coupling, partner size, mutation rate, reproduction speed, and the use of local learning. Variance in behaviour from the traditional symmetrical case is seen across the parameter space of the model.

2. Background: The NKCS Model

Kauffman and Levin [1987] introduced the NK model to allow the systematic study of various aspects of fitness landscapes. In the standard model, the features of the fitness landscapes are specified by two parameters: *N*, the length of the genome; and *K*, the number of genes that has an effect on the fitness contribution of each (binary) gene. Thus increasing *K* with respect to *N* increases the epistatic linkage, increasing the ruggedness of the fitness landscape. The increase in epistasis increases the number of optima, increases the steepness of their sides, and decreases their correlation [Kauffman, 1993]. The model assumes all intragenome interactions are so complex that it is only appropriate to assign random values to their effects on fitness. Therefore for each of the possible *K* interactions a table of $2^{(K+1)}$ fitnesses is created for each gene 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 is found from its table. These fitnesses are then summed and normalized by *N* to give the selective fitness of the total genome (Figure 1).



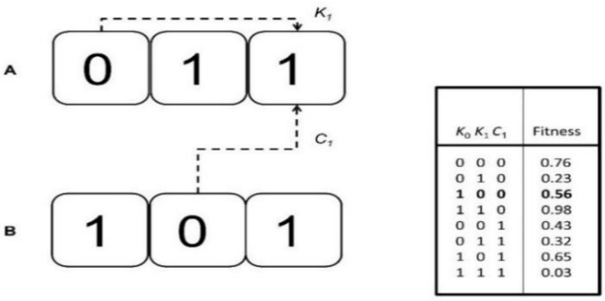
K_0 K_1	f(0)	f(1)	f(2)
0 0	0.76	0.02	0.99
0 1	**0.23**	0.34	0.82
1 0	0.56	**0.38**	**0.53**
1 1	0.98	0.86	0.19

$$Fitness = 1/N \Sigma f(n)$$

$$Fitness = (0.23+0.38+0.53)/3 = 0.38$$

Fig. 1. An example NK model showing how the fitness contribution of each gene depends on *K* random genes (left) and the table with $2^{(K+1)}$ rows used for the fitness calculation (right).

Kauffman and Johnsen [1992] subsequently introduced the abstract NKCS model to enable the study of various aspects of *co*evolution. Each gene is said to also depend upon *C* randomly chosen traits in each of the other *S* species with which it interacts. Altering *C*, with respect to *N*, changes how dramatically adaptive moves by each species deform the landscape(s) of its partner(s), where increasing *C* typically increases the time to equilibrium. Again, for each of the possible $K+(S\text{x}C)$ interactions, a table of $2^{(K+(S\text{x}C)+1)}$ fitnesses is created for each gene, with all entries in the range 0.0 to 1.0, such that there is one fitness for each combination of traits. Such tables are created for each species (Figure 2).



K_0 K_1 C_1	Fitness
0 0 0	0.76
0 1 0	0.23
1 0 0	**0.56**
1 1 0	0.98
0 0 1	0.43
0 1 1	0.32
1 0 1	0.65
1 1 1	0.03

Fig. 2. The NKCS model: Each gene is connected to *K* randomly chosen local genes and to *C* randomly chosen genes in each of the *S* other species. Connections and table shown for one gene in one species for clarity. Here *N*=3, *K*=1, *C*=1, *S*=1.

Following [Kauffman, 1993], a mutation-based hill-climbing algorithm is used here, where the single point in the fitness space is said to represent a converged species, to examine the properties and evolutionary dynamics of the NKCS models. That is, the population is of size one and a species evolves by making a random change to *M* randomly chosen genes per generation. The “population” is said to move to the genetic configuration of the mutated individual if its fitness is greater than the fitness of the current individual; the rate of supply of mutants is seen as slow compared to the actions of selection. Ties are broken at random.

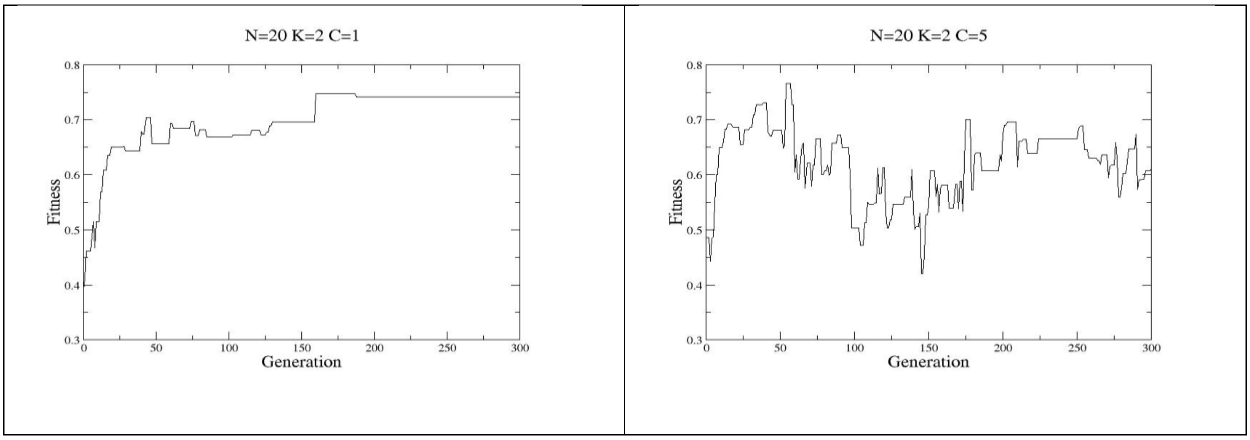


Fig. 3. Showing example single runs of the typical behaviour of the standard NKCS model of coevolution with different degrees of landscape coupling (*C*).

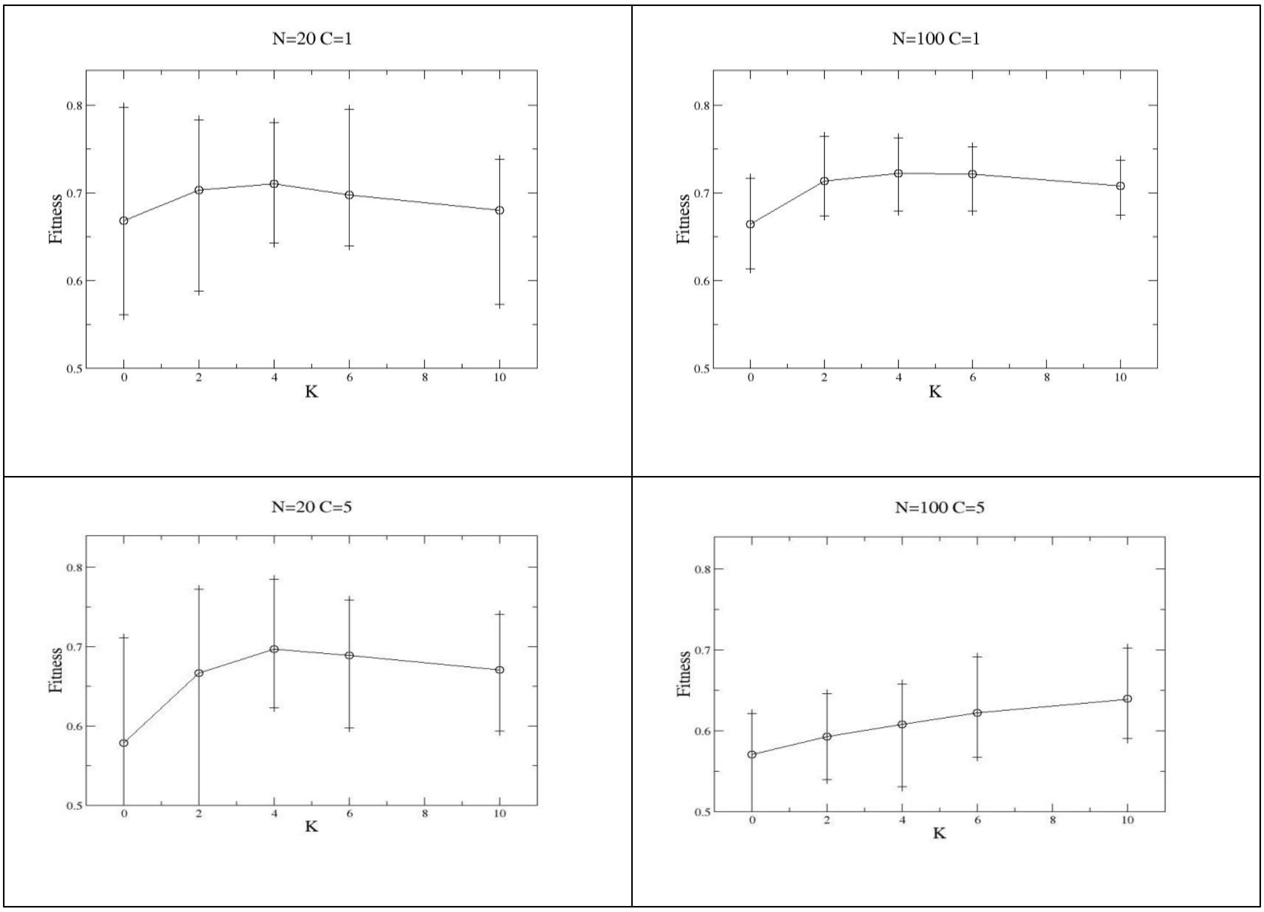


Fig. 4. Showing the fitness reached after 20,000 generations on landscapes of varying ruggedness (*K*), coupling (*C*), and length (*N*). Note how fitness drops as *K*→*N*, as in the traditional NK model. Error bars show min and max values.

Figure 3 shows example behaviour for one of two coevolving species where the parameters of each are the same and hence behaviour is symmetrical. The effects of mutual fitness landscape movement are clearly seen. All results reported in this paper are the average of 10 runs (random start points) on each of 10 randomly created NKCS fitness landscapes, i.e., 100 runs, for 20,000 generations, for each parameter configuration. The (average) final fitness of the converged population is used for comparisons. Here M=1, 0≤K≤10, 1≤C≤7, S=1, for N=20 and N=100. Figure 4 shows how increasing the degree of coupling (C) between the two landscapes causes fitness levels to fall significantly (T-test, p<0.05) when C≥K for N=20. Note this change in behaviour around C=K was suggested as significant in [Kauffman & Johnson, 1992], where N=24 was used throughout. However, Figure 4 also shows how with N=100 fitness *always* falls significantly with increasing C (T-test, p<0.05), regardless of K. As noted in [Kauffman & Johnson, 1992], increasing N increases the time to equilibrium; the duration of evolutionary search is increased.

3. Asymmetry in the NKCS Model

3.1 Epistasis *K*

Kauffman and Johnsen [1992] explored the effects of asymmetric K for varying C. In particular, they describe the changes in behaviour seen before Nash equilibria are reached. That is, in the symmetrical case, as C increases, the time taken to reach a Nash equilibrium increases (see Figure 3) and, for a given C, the time taken to reach an equilibrium decreases with increasing K. They show how, for higher values of C, a low K species has a higher pre-Nash average fitness against a high K species, and vice versa.

Figure 5 shows the final fitness reached is also affected by coevolving against a species with different K. Note the partner species' asymmetric parameter X is denoted X' here. With

N=100, when *C*>1, a significant decrease in fitness is typically seen when *K'*<*K* and a significant increase when *K'*>*K* (T-test, *p*<0.05). This appears to correlate with Kauffman and Johnsen's observations about pre-Nash fitness. When *N*=20 behaviour is less clear overall but with *C*=5 a similar result is seen when *K*<*C*. For *C*<5, increasing *K*' typically increases fitness when *K*<*C*. The point at which the increase becomes significant decreases with *C*. No significant effect is seen otherwise. For example, when *K*=0, fitness is significantly higher when *K*'≥8 and *C*=1 (T-test, *p*<0.05), whereas fitness is significantly higher when *K*'>0 with *C*>1 (T-test, *p*<0.05). Conversely, there is no significant change in fitness seen when *K*≥6 for any *K*' or *C* (T-test, *p*≥0.05).

3.2 Coupling *C*

As shown in Figure 4, in the traditional NKCS model, increasing *C* decreases the fitness level reached with *N*=100 – but only when *K*<*C* with *N*=20. Figure 6 shows the effects of allowing different degrees of coupling between the two species. With *N*=20, fitness typically decreases when *C'*>*C* when *K*<*C*. That is, the fitness of the less coupled species is significantly lowered when coevolving with a partner that is more coupled to them. With *N*=100 similar behaviour is seen, although the larger case again appears more sensitive to the coupling and fitness decreases for higher *K* compared to roughly when *K*<*C*.

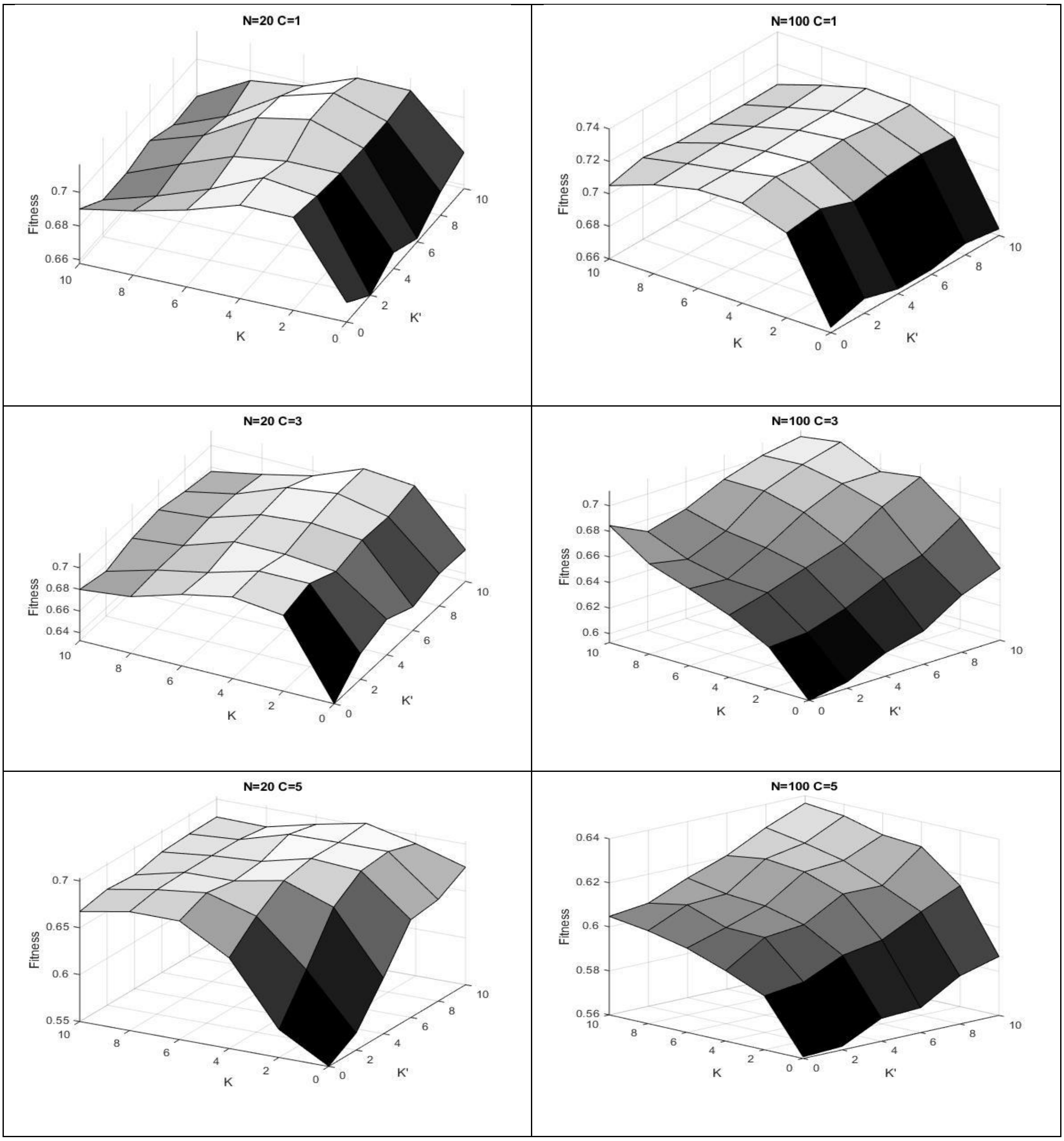
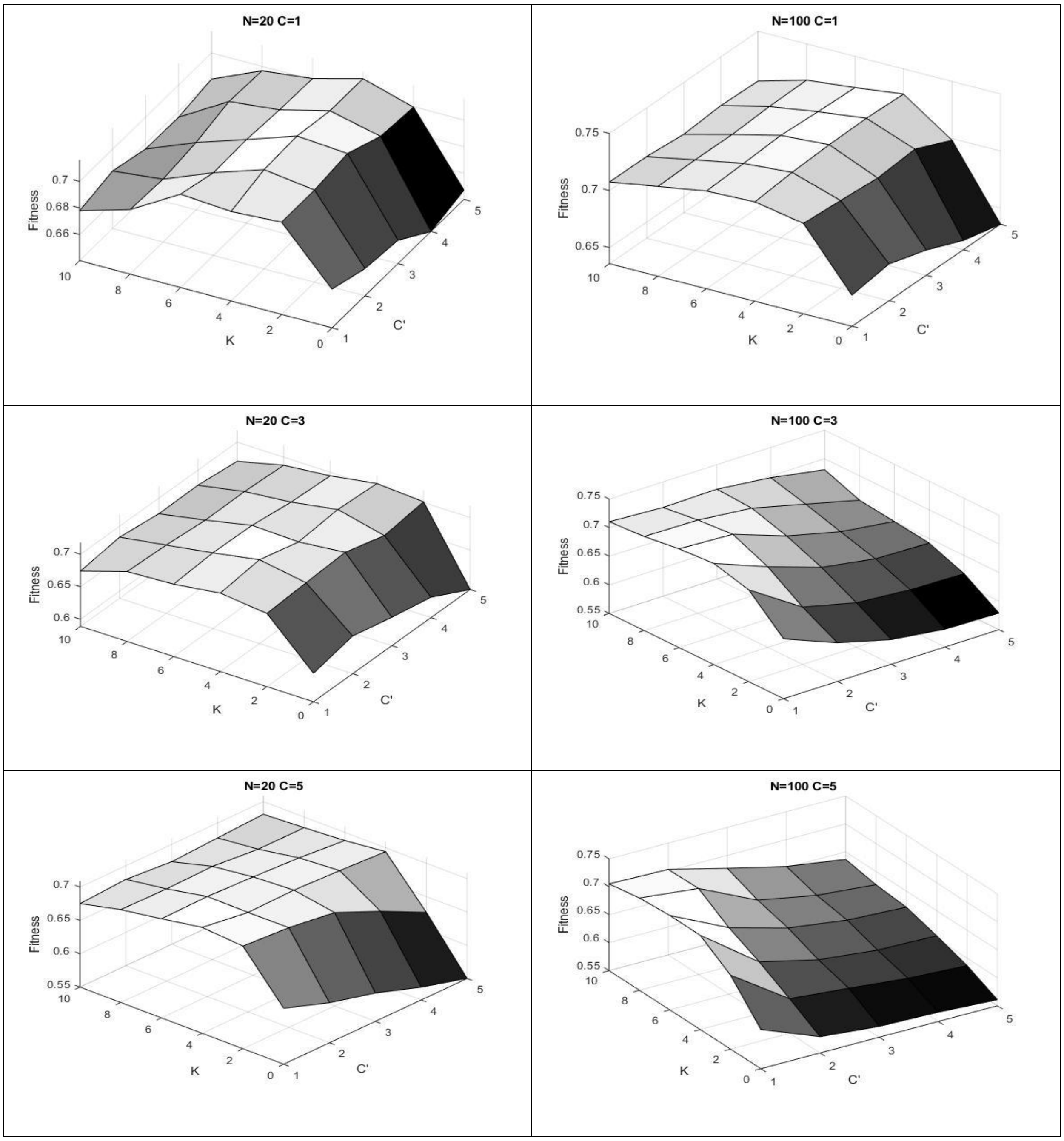


Fig. 5. Showing the fitness reached by one species after 20,000 generations on landscapes of varying ruggedness (*K*), coupling (*C*), and length (*N*). The partner species had the same parameters except for the landscape ruggedness (*K'*).



N=20 C=1

N=100 C=1

N=20 C=3

N=100 C=3

N=20 C=5

N=100 C=5

Fig. 6. Showing the fitness reached by one species after 20,000 generations on landscapes of varying ruggedness (*K*), coupling (*C*), and length (*N*). The partner species had the same parameters except for the degree of coupling (*C*').

3.3 Size *N*

Since the larger symmetrical *N* case shows different behaviour to the smaller *N*, the effects of coevolving two different sized species have also been explored. As noted in the introduction, this can be seen as the typical case for more intimate forms of symbiotic relationships, for example. Figure 7 shows how, when N=20 and C=5, increasing the length of the partner significantly increases fitness when $K<C$ (T-test, $p<0.05$). No significant effect is seen otherwise. When N=100, no significant effect is seen for C=1 (T-test, $p\geq 0.05$). For C>1, fitness typically decreases with decreasing N', when $K<C$. However, when $K>C$, the lowest fitness levels are typically reached when $N=N'$, increasing again for increasing and decreasing values of N'. Figure 8 shows two examples in more detail. The reason for such inflection points is not immediately clear amongst the interactions between the number of local optima, the typical walk length to such optima, and the duration for which evolution is able to search amongst them before an equilibrium is reached. The same effect is seen when N=20 when $K<C$.

3.4 Mutation *M*

Kauffman and Johnsen [1992] explored the effects of varying the number of mutations applied per generation, showing that generally the optimal *M* decreases with increasing *C*. Figure 9 shows the effects of varying the number of unique mutations per offspring in the partner species, that is, M=1 and $1\leq M'\leq 5$. When N=20 and C>1, increasing M' decreases fitness when $K<C$, with no significant effect otherwise. When N=100 and C>1, fitness typically decreases with increasing M' for all *K* and *C*. That is, generally, being coupled to a species varying significantly is, perhaps not unsurprisingly, detrimental.

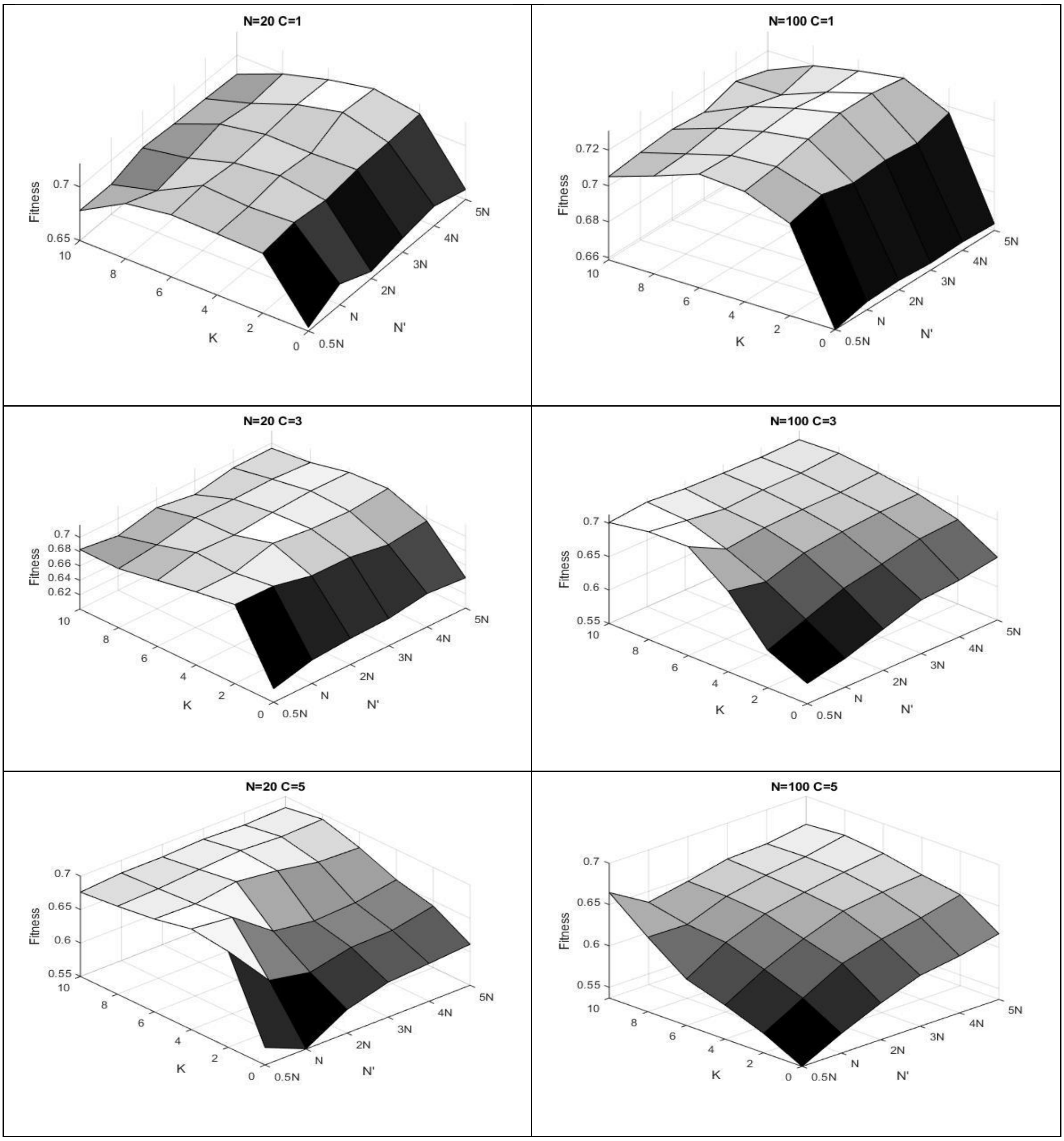


Fig. 7. Showing the fitness reached by one species after 20,000 generations on landscapes of varying ruggedness (*K*), coupling (*C*), and length (*N*). The partner species had the same parameters except for the size (*N*’).

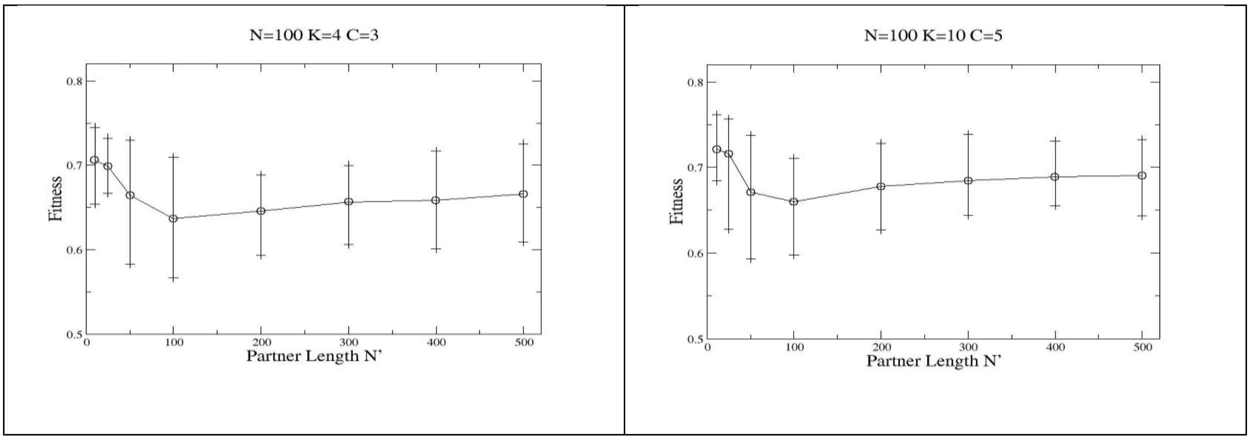


Fig. 8. Showing examples of how varying the length of the coevolving partner can both increase and decrease fitness reached after 20,000 generations under some conditions.

3.5 Rate *R*

The traditional NKCS model assumes all species coevolve at the same rate. Following [Bull, et al. 2000], a new parameter *R* can be added to the model to represent the relative rate at which one species evolves - by undertaking *R* rounds of mutation and selection to one round in the other(s). Figure 10 shows how with *N*=100, generally, increasing *R* decreases the effects of *C* for a given *K* when *C*>1, from around *R*=-*5*, where -20≤*R*≤20. That is, evolving at a faster rate is beneficial (+*R*) whereas evolving more slowly (-*R*) is detrimental. In contrast, no notable effect is seen by increasing *R* for any *K* and *C* values tried when *N*=20, except for *K*=0 when *C*>1 and *R*>1. It can be noted that *N*=64 (only) was used in [Bull et al., 2000].

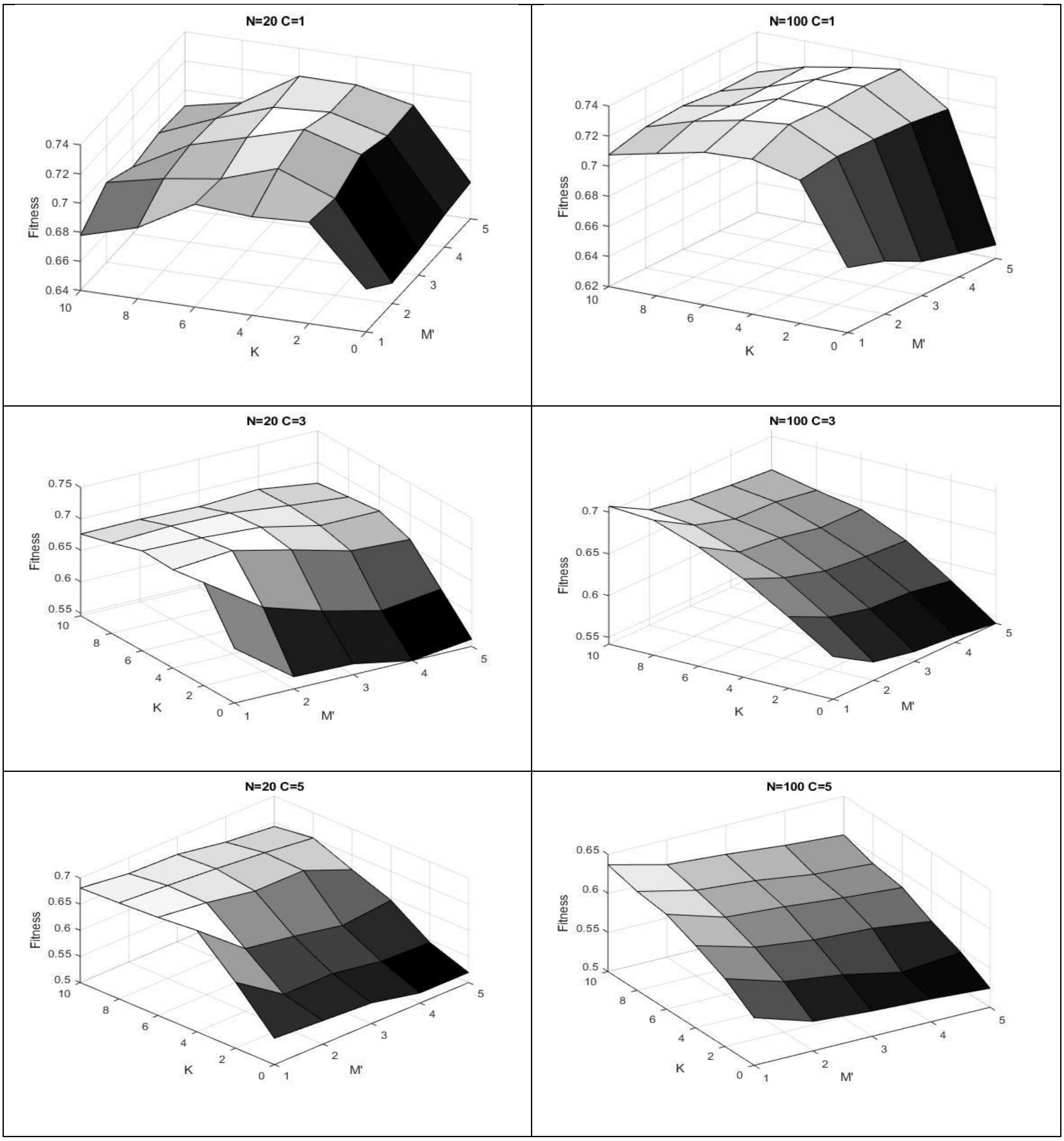


Fig. 9. Showing the fitness reached by one species after 20,000 generations on landscapes of varying ruggedness (*K*), coupling (*C*), and length (*N*). The partner species had the same parameters except for the mutation rate (*M*’).

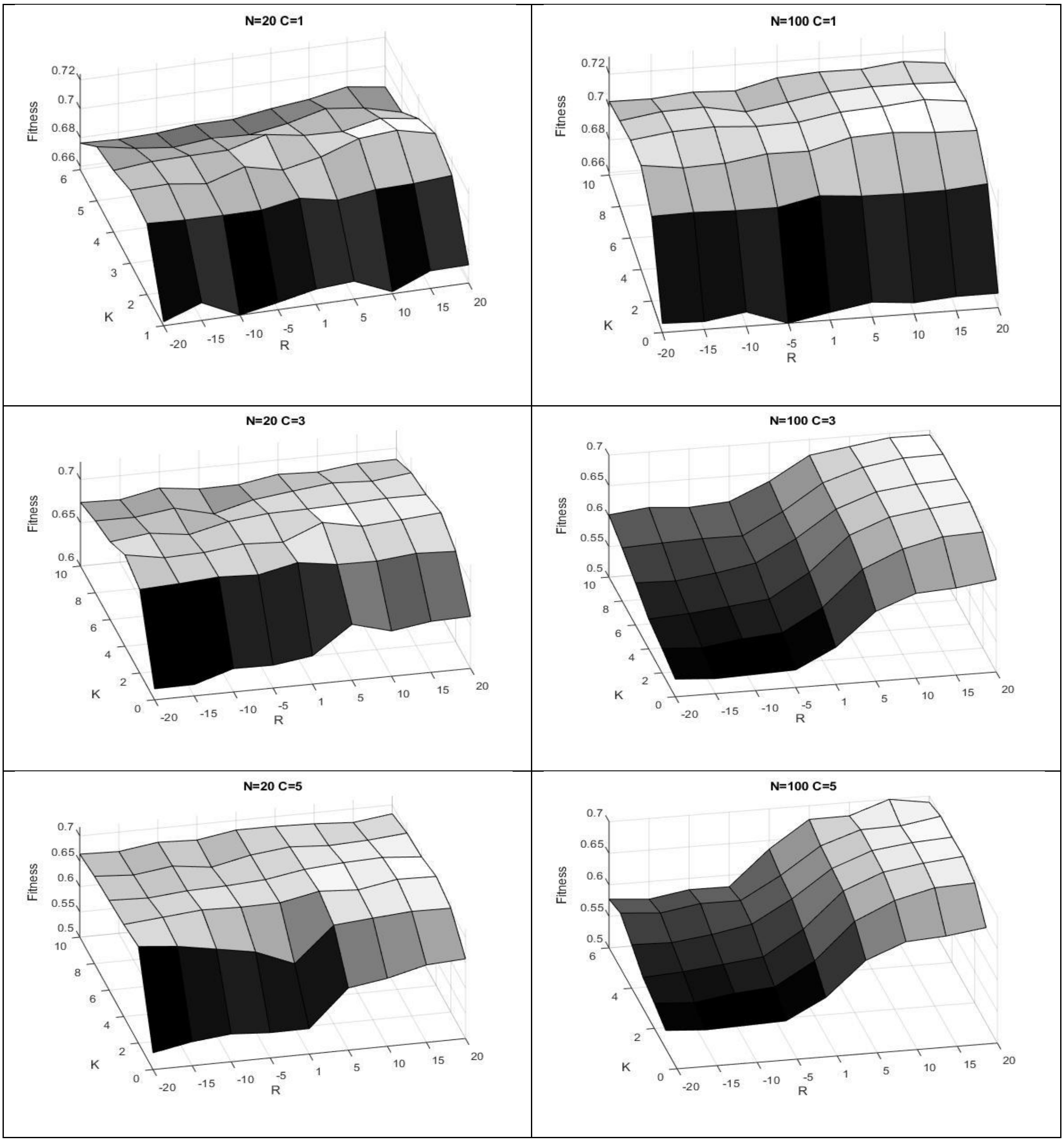


Fig. 10. Showing the fitness reached by one species after 20,000 generations on landscapes of varying ruggedness (*K*), coupling (*C*), and length (*N*). The partner species had the same parameters except the shown species evolves at a different rate (*R*).

3.6 Learning *L*

Hinton and Nowlan [1987] were the first to investigate the Baldwin effect [Baldwin, 1896], showing that enabling genetically specified neural networks to alter inter-neuron connections randomly during their lifetime meant the evolutionary system was able to find an isolated optimum, something the system without learning struggled to achieve. That is, the ability to learn "smoothed" the fitness landscape into a unimodal hill/peak. They also found that over time more and more correct connections became genetically specified and hence less and less random learning was necessary; the evolutionary process was guided toward the optimum by the learning process (e.g., see [Sznajder et al., 2012] for an overview of the Baldwin effect). Using the NK model, it was later shown how the most beneficial frequency and amount of learning varies with the ruggedness of the underlying fitness landscape [Bull, 1999]. Whilst learning and evolution have long been used together within coevolutionary scenarios (after [Ackley & Littman, 1992]), no previous systematic exploration is known.

Following [Bull, 1999], a very simple (random) learning process to enable phenotypic plasticity can be added to evolution by allowing a new individual to make a further L (unique) mutations after the first. If the averaged fitness of this "learned" configuration and that of the first mutant is greater than that of the original, the species is said to move to the *first* mutant configuration but assigned the *averaged fitness* of the two configurations, here $1 \leq L \leq 5$.

Figure 11 shows how when N=20, with C=1, increasing L is beneficial when K>0, with the most benefit typically seen around L=3. However, L>0 proves detrimental when C>K thereafter (T-test, p<0.05). When N=100, with C=1, only L=1 is beneficial when K>0 (T-test, p<0.05), with the the local learning decreasing fitness for all other parameter combinations. Hence it appears increased movement in the fitness landscape disrupts the smoothing process seen in the non-coevolutionary case.

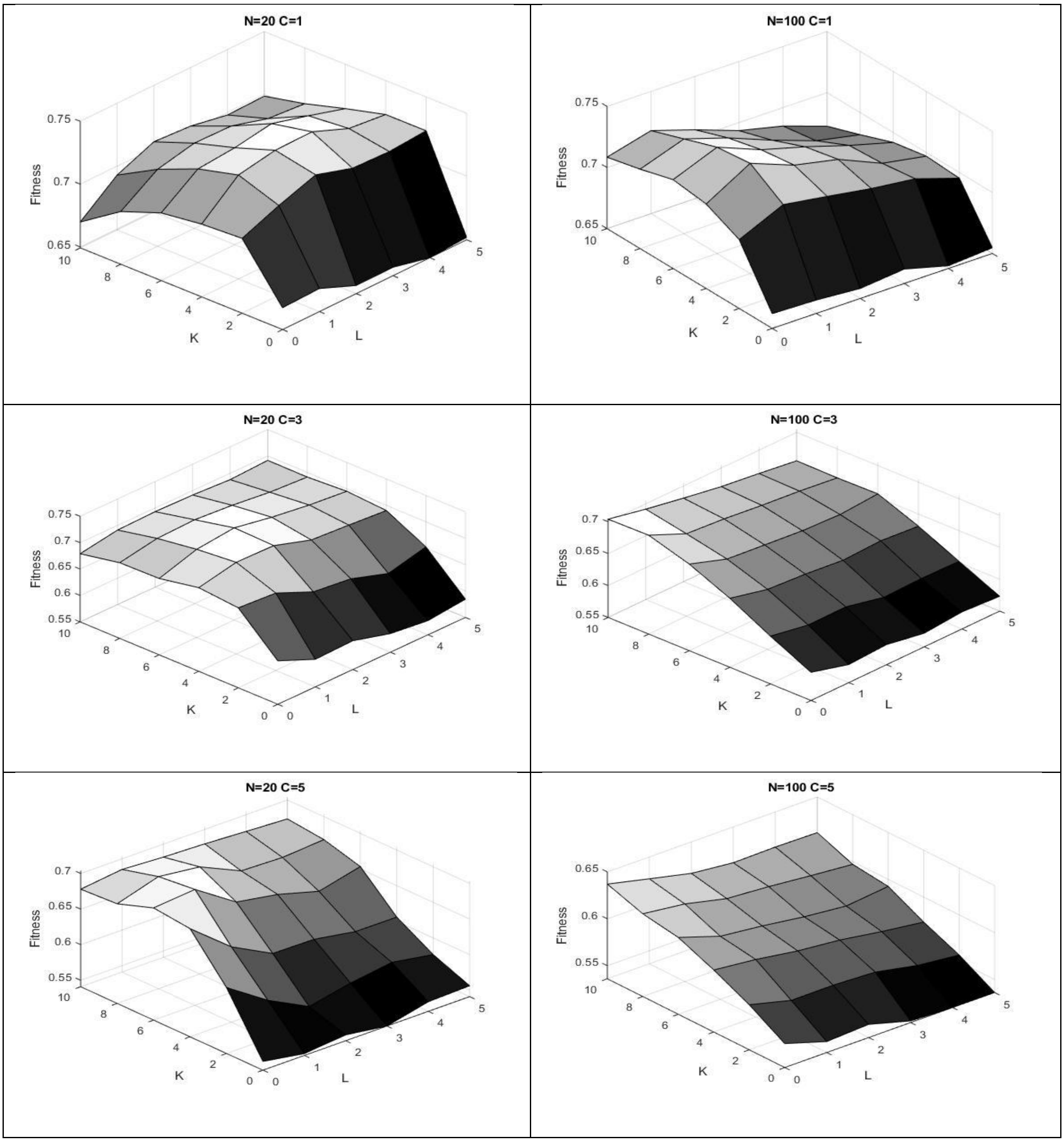


Fig. 11. Showing the fitness reached by one species after 20,000 generations on landscapes of varying ruggedness (*K*), coupling (*C*), and length (*N*). The partner species had the same parameters except the shown species also utilises local random learning at rate (*L*).

4. Conclusion

The original NKCS model - and almost all known uses of it thereafter - assumes symmetry in the species. It has been shown here that introducing asymmetry in the degree of landscape ruggedness (*K*), the degree of landscape coupling (*C*), the size of the partners' genomes (*N*), the amount of mutation experienced by a species (*M*), the relative rate of reproduction of the partners (*R*), and the inclusion of local search (*L*) can all affect fitness in comparison to the symmetrical case.

One of the few previous studies to consider asymmetry revisited the use of the NKCS model to explore the conditions under which hereditary (endo)symbioses may emerge [Bull, 2010]. It was shown how, assuming that forming the closer relationship slowed the relative speed of evolution of one partner (*R*'), the host would benefit. This extended the previously identified conditions of benefit when $C>K$, noting how interspecies epistasis effectively becomes intraspecies epistasis in such cases, i.e., *C* becomes *K* [Bull & Fogarty, 1996]. Section 3.5 above showed how a phase transition-like phenomena typically exists where the fitness of the slower species sharply decreases as the relative rate of reproduction of the partner increases.

The other asymmetric partnering scenarios explored here can also be seen as relevant to symbioses, as well as other close ecological relationships in general: section 3.1 showed how partnering with a species on a more rugged fitness landscape (*K*') can be beneficial, as is the case with bacteria and multicellular eukaryotes; section 3.2 showed how being partnered with a more dependent species (*C*') can lower fitness, something which might also encourage the formation of hereditary symbiotic relationships; section 3.3 showed that coevolving with a partner of greater size (*N*') is beneficial, especially for low complexity (*N* and *K*) species, again the typical case of bacteria and multicellular eukaryotes; section 3.4

explored the sensitivity to the mutation rate (*M*') of a partner, with more complex (*N*) species being particularly vulnerable, again this would appear to also motivate the direct inheritance of bacterial symbionts by eukaryotes; and finally, section 3.6 showed how fitness landscape smoothing through individual learning is generally not beneficial for closely coupled species.

References

Ackley, D. & Littman, M. (1992). Interactions between learning and evolution. In C G Langton, C Taylor, J D Farmer & S Rasmussen (Eds.) *Artificial Life II*. Addison-Wesley, pp487-509.

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

Bull, L. (2010). Artificial symbiogenesis and differing reproduction rates. *Artificial Life,* 16(1): 65-72.

Bull, L. (2020). *The Evolution of Complexity: Simple Simulations of Major Innovations*. Springer.

Bull, L. & Fogarty, T.C. (1996). Artificial symbiogenesis. *Artificial Life,* 2(3):269-292.

Bull, L. (1999). On the Baldwin effect. *Artificial Life,* 5(3):241-246.

Bull, L., Holland, O. & Blackmore, S. (2000). On meme-gene coevolution. *Artificial Life,* 6(3): 227-235

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

Kauffman, S. A. (1993). *The Origins of Order: Self-organisation and Selection in Evolution*. Oxford University Press.

Kauffman, S. A. (1995). *At Home in the Universe: The Search for Laws of Complexity*. Viking Press.

Kauffman, S. A., & Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128: 11–45.

Kauffman, S. A., & Johnsen, S. (1992). Co-evolution to the edge of chaos: Coupled fitness landscapes, poised states and co-evolutionary avalanches. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial Life II* (pp. 325–370). Addison-Wesley.

Levitan, B., Lobo, J., Schuler, R., & Kauffman, S.A. (1997). Evolution of organizational performance and stability in a stochastic environment. *Computational & Mathematical Organization Theory*, 8:281-313.

Preen, R., You, J., Bull, L. & Ieropoulos, I. (2019). Design mining microbial fuel cell cascades. *Soft Computing*, 23(13): 4673-4683.

Stanley, K.O. & Miikkulainen, R. (2004). Competitive coevolution through evolutionary complexification. *Journal of Artificial Intelligence Research,* 21: 63-100.

Sznajder, B., Sabelis, M.W. & Egas, M. (2012). How adaptive learning affects evolution: Reviewing theory on the Baldwin effect. *Evolutionary Biology*, 39: 301-310.