# "If You Can't Be With the One You Love, Love the One You're With": How Individual Habituation of Agent Interactions Improves Global Utility

Adam P. Davies*,**
University of Southampton

Richard A. Watson**
University of Southampton

Rob Mills**
University of Southampton

C. L. Buckley†
University of Sussex

Jason Noble**
University of Southampton

**Abstract** Simple distributed strategies that modify the behavior of selfish individuals in a manner that enhances cooperation or global efficiency have proved difficult to identify. We consider a network of selfish agents who each optimize their individual utilities by coordinating (or anticoordinating) with their neighbors, to maximize the payoffs from randomly weighted pairwise games. In general, agents will opt for the behavior that is the best compromise (for them) of the many conflicting constraints created by their neighbors, but the attractors of the system as a whole will not maximize total utility. We then consider agents that act as *creatures of habit* by increasing their preference to coordinate (anticoordinate) with whichever neighbors they are coordinated (anticoordinated) with at present. These preferences change slowly while the system is repeatedly perturbed, so that it settles to many different local attractors. We find that under these conditions, with each perturbation there is a progressively higher chance of the system settling to a configuration with high total utility. Eventually, only one attractor remains, and that attractor is very likely to maximize (or almost maximize) global utility. This counterintuitive result can be understood using theory from computational neuroscience; we show that this simple form of habituation is equivalent to Hebbian learning, and the improved optimization of global utility that is observed results from well-known generalization capabilities of associative memory acting at the network scale. This causes the system of selfish agents, each acting individually but habitually, to collectively identify configurations that maximize total utility.

**Keywords**
Hebbian learning, Hopfield network, associative memory, game theory, self-organization, adaptive networks



## 1 Selfish Agents and Total Utility

This article investigates the effect of a simple distributed strategy for increasing total utility in systems of selfishly optimizing individuals. In closely related work we have developed a general model addressing this topic [42], and here we focus on a social agent system and the implications for social networks. The broader topic concerns many different types of systems.

* Contact author.
** Natural Systems group, ECS, University of Southampton, UK. E-mail: apd1e09@ecs.soton.ac.uk (A.P.D.); raw@ecs.soton.ac.uk (R.A.W.); rob.mills@soton.ac.uk (R.M.); jn2@ecs.soton.ac.uk (J.N.)
† CCNR, University of Sussex, UK. E-mail: c.l.buckley@sussex.ac.uk

 

For example, in technological systems, it is often convenient or necessary to devolve control to numerous autonomous components or agents that each, in a fairly simple manner, act to optimize a global performance criterion: Thus, communications routing agents act to minimize calls dropped, and processing nodes in a grid computing system each act to maximize the number of jobs processed [9, 21]. However, since each component in the network acts individually (i.e., using only local information), constraints between individuals can remain unsatisfied, resulting in poorly optimized global performance. In an engineered system one could, in principle, mandate that all nodes act in accord with the globally optimal configuration of behaviors (assuming one knew what that was)—but this would defeat the scalability and robustness aims of complex adaptive systems. The question for engineered complex adaptive systems, then, is how to cause simple autonomous agents to act "smarter" in a fully distributed manner such that they better satisfy constraints between agents and thereby better optimize global performance.

Meanwhile, in evolutionary biology it appears that in certain circumstances symbiotic species have formed collaborations that are adaptive at a higher level of organization [18], but it has been difficult to integrate this perspective with the assumption that under natural selection such collaborations must be driven by the selfish interests of the organisms involved [4, 19].

In social network studies there is increasing interest in adaptive networks [22] where agents in a network can alter the structure of the connections in it. Of particular interest is the possibility that by doing so they may increase the ability of the system to maintain high levels of cooperation [23, 26, 32, 36]. However, despite increased focus on models of *network reciprocity* [25], a general understanding of how agents in a network modify their interactions with others in a way that increases total cooperation is poorly understood.

In each domain we are interested, at the broadest level, in understanding/identifying very simple mechanisms that might cause self-interested agents to modify their behavior, or how their behaviors are affected by others, in a manner that increases adaptation or efficiency, either globally or at some other level of organization higher than the individual.

Taking an agent perspective, the obvious problem is this: If it is the case that agents collectively create adaptation that is not explained by the default selfish behaviors of individuals, then it must be the case that, on at least some occasions, agents make decisions that are detrimental to individual interests. If not, then there would be nothing to be explained over and above the selfish actions of individuals. But this conclusion appears to run counter to any reasonable definition of a rational selfish agent. In what sense could it be self-consistent to suggest that a *selfish* agent has adopted a behavior that *decreased* individual utility? One way to make sense of this is the possibility that, at the time that the agent takes this action, it appears to it to be the best one for it—that the agent is no longer making decisions according to the true utility function, but according to some distortion of it that alters the agent's perception of the utility of that action. If somehow the perception of an agent were distorted in the right way, so that the action that it preferred—the one that it thought was best for it—was in fact the action that was globally optimal, then a rational agent with this distorted set of preferences could increase global efficiency even at the cost of personal utility.

One might assume that this is easier said than done—but in this article we suggest that the reverse is true; it is easier to do than to explain how it works. However, the general problem and the essence of the strategy we investigate are straightforwardly introduced by means of the following simple parable. Although this makes the concepts intuitively accessible, it might tend to cast the model in a narrow interpretation. It is, of course, not really a model of scientists and their drinking habits, but a general model of interacting agents on a network with pairwise constraints between binary behaviors.

Consider a community of individuals (e.g., researchers) in a social network. Each has an intrinsic symmetric *compatibility*, or complementarity, with every other individual that determines the productivity (payoff) of collaborating with them. Each evening all researchers attend one of two intrinsically equal public houses (or other such collaborative projects), initially at random. Individuals must decide which to attend, based solely on who else attends that venue. Each individual seeks to maximize his or her scientific productivity by attending the pub that, on that night, maximizes the sum of compatibilities

with other researchers and minimizes incompatibilities. Assessing the company they find at any moment, individuals therefore (one at a time in random order) may choose to switch pubs to maximize their productivity according to the locations of others. Since each individual has compatibilities and incompatibilities with all other individuals, each must choose the pub that offers the best compromise of these conflicting interests. Since compatibilities are symmetric, the researchers will quickly reach a configuration where no one wants to change pubs [11]. However, this configuration will not in general be the arrangement that is maximal in total productivity, but merely a locally optimal configuration.

This describes the basic behavior of agents in the network. Our aim is to devise a simple individual strategy that causes researchers to make better decisions about when to change pubs so that total productivity is maximized. This will necessarily mean that some researchers, at some moments in time, must change pubs even though it decreases their individual productivity.

Surprisingly, we find that this can be achieved (over many evenings) by implementing a very simple rule—each individual must develop a preference for drinking with whichever other researchers they are drinking with right now. As Crosby, Stills & Nash put it, "If you can't be with the one you love, honey, love the one you're with" [34]. Since we already know the arrangements of researchers will be initially random and, most of the time, at best suboptimal, this seems like a counterproductive strategy. But, in fact, we find that it is capable, given enough evenings and slowly developed preferences, of causing all researchers to develop preferences that cause them to make decisions that maximize total productivity reliably every evening.

The agents that we model are therefore not selfish in an identical way to default agents—they sometimes take actions that do not maximize individual utility, which is the point of the exercise after all. But neither are they overtly cooperative or altruistic agents. They are simply *habitual selfish agents*. In this article we are not directly addressing *why* it might be that selfish agents act as creatures of habit, although we will discuss this briefly. But we suggest this type of distorted perception of a true utility function, one which agents come to prefer familiarity over otherwise obvious opportunities for personal gain, is one that does not require any teleological or, certainly, any centralized control and is therefore relevant to many domains. This *distorted perception* model differs from the work described in [42], where agents are assumed to be able to directly alter their real constraints with other agents. Watson et al. [42] also discuss the immediate selfish benefits that might motivate changes to constraints given that assumption, whereas in this article we simply mandate that agents act in a habitual manner (see Section 4). But the two articles have in common the idea of a self-modeling dynamical system and its equivalence with neural network models [39].

In the next two sections we will detail an illustration of this strategy and the results we observe. In Section 4 we will outline how this result can be interpreted in terms of adaptive network restructuring. Briefly: Initially, interactions between agents are governed by a network of intrinsic constraints (compatibilities), and later they are governed by a combination of these intrinsic constraints plus the interaction preferences that the agents have developed. The new behavioral dynamics of the agents caused by interaction preferences can therefore equally well be understood as a result of changes to connection strengths in the effective interaction network. The increased global utility observed can then be explained using theory from computational neuroscience. In particular, we can understand how the system as a whole improves global adaptation, via the observation that when each agent acts as a creature of habit it changes the effective dependencies in the network in a Hebbian manner [8, 10]. This means that through the simple distributed actions of each individual agent, the network as a whole behaves in a manner that is functionally equivalent to a simple form of learning neural network [39]. In this case, the network is not being trained by an external training set, but instead is "learning" its own attractor states, as we will explain. We discuss how a separation of the timescales for behaviors *on* the network and behaviors *of* the network (i.e., changes to network structure) is essential for this result, and consider conditions under which a habitual policy could spontaneously arise. We examine the concept of selfishness when applied to long term utility maximization mechanisms, and put forward an interpretation of selfishness that reframes the conundrum of cooperation, removing the apparent paradox by viewing cooperation as a by-product of long term selfishness.

Finally, we link the habituation policy with existing reinforcement-learning algorithms—in particular, those of generalized weakened fictitious play [16].

## 2 Methods

### 2.1 Default Agents

Our model involves $N = 100$ agents playing two-player games (Table 1) on a fully connected network. Specifically, for each game (i.e., each connection in the network), there is a single symmetric payoff matrix, $U_{ij}$, which defines for agents $i$ and $j$ either a coordination game ($\alpha = 1$, $\beta = 0$) or an anticoordination game ($\alpha = 0$, $\beta = 1$) with equal probability (Table 1).

Initially, all agents in the network are assigned a behavior at random, and then the game progresses in extensive form with perfect monitoring, where each agent, in a random order, is permitted to update its behavior (to either A or B) after observing the current actions of the other players. Each agent does so according to a best-response strategy, namely, to adopt the behavior (choose a pub) that maximizes its utility $u_i$ given the behaviors (pub choices) currently adopted by its neighbors:

$$u_i(t) = \sum_j^N U_{ij}(s_i(t), s_j(t)) \tag{1}$$

where $U_{ij}(x,y)$ is the payoff received by player $i$ when player $i$ plays strategy $x$ and player $j$ plays strategy $y$ (according to Table 1), and $s_n(t)$ is the strategy currently played by agent $n$. Behaviors are updated in this manner repeatedly. Each agent is involved in games with all neighbors simultaneously, but can adopt only one behavior at any one time; thus coordinating with one neighbor may preclude coordinating with another, and so each agent must adopt the behavior that is the best compromise of these constraints. By using a symmetric game, $U_{ij} = U_{ji}$, we can ensure that the system will reach a stable fixed point [11], that is, a configuration where no agent wants to change behavior unilaterally. Moreover, this configuration will be a local optimum in the total or global utility $G$ of the system, which is simply the sum of individual utilities [11]:

$$G(t) = \sum_i^N \sum_j^N U_{ij}(s_i(t), s_j(t)) \tag{2}$$

Equivalently, since each game is a potential game [20] and the global game is simply an addition of all local games, then the global game must also be a potential game [2] (the potential function given by Equation 2), and hence will have only pure Nash equilibria [20]. In general, however, the stable configuration reached from an arbitrary initial condition will not be globally maximal in total utility. If the system is repeatedly perturbed (reassigning random behaviors to all agents) at infrequent intervals (here every 1000 time steps = one evening), and thereby allowed to settle, or *relax*, to many different local

Table 1. Payoff for (player 1, player 2).

| | | Player 2 | |
|---|---|---|---|
| | | A | B |
| Player 1 | A | $\alpha,\alpha$ | $\beta,\beta$ |
| | B | $\beta,\beta$ | $\alpha,\alpha$ |

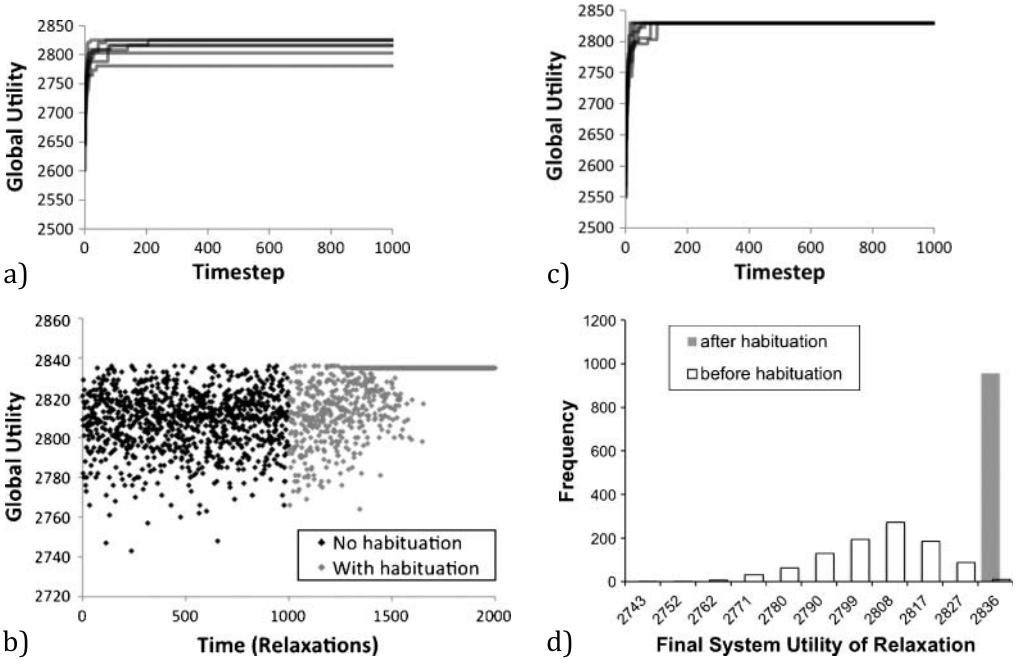


Figure 1. Behavior of the system using default (no habituation) and habituating agents. (a) Some example trajectories of system behavior before habituation. Each curve represents one relaxation ($N = 100$, relaxation length $10N$). Vertical axis is the total system utility ($G$, Equation 2). (b) Utilities of attractor states visited (i.e., end points of curves like those in (a)) without habituation (relaxations 1–1000) and during habituation (relaxations 1001–2000, $r = 0.005$). (c) Example trajectories after habituation. (d) Histogram of attractor utilities before habituation (relaxations 1–1000) and after habituation (relaxations 2001–3000), showing that after habituation the system reliably finds one of the highest-total-utility configurations from any initial condition.

equilibria (on different evenings), the behavior of the system given these default agents can be described by the distribution of total utilities found at the end of each of these relaxations (Figure 1c).

## 2.2 Creatures of Habit

We seek a simple distributed strategy that causes agents to make different (hence unselfish) behavioral choices in particular contexts in such a manner that configurations of higher global utility are attained or high-global-utility configurations are attained with greater reliability (i.e., from a greater number of random initial conditions). To this end, we investigate agents that act as creatures of habit by increasing their preference to coordinate with whichever neighbors they are coordinated with at the present moment (regardless of whether this is presently contributing positively or negatively to their utility). Specifically, in addition to the *true* utility matrix $U_{ij}$, each agent also possesses a *preference* matrix $P_{ij}$ for each of its connections. These are used to modify the behavior of the agent so that it chooses the behavior that maximizes its *perceived utility*, $p_i$, instead of its true utility (Equation 2) alone:

$$p_i(t) = \sum_{j}^{N} [U_{ij}(s_i(t), s_j(t)) + P_{ij}(t)(s_i(t), s_j(t))] \tag{3}$$

where $P_{ij}$ is a payoff matrix that represents an agent's preference for the combination of behaviors $s_i$ and $s_j$. The perceived utility is thus simply the sum of the true utilities plus the agent's preferences. Each agent has a separate preference payoff matrix for each other agent. All preference payoff matrices are initially set to zero, so that the initial dynamics of the agents are per the default agents. But

as the elements in these matrices change over time, they may come to collectively overpower the tendency to maximize true utility and thereby cause agents to make different decisions about which behavior is best for them to adopt.

It should be clear that it is possible in principle, with knowledge of the globally optimal system configuration, to assign values to each of the matrices $P_{ij}$ that will cause agents to adopt behaviors that maximize global system utility instead of choosing behaviors that maximize individual utility and thereby failing to maximize total utility. But our question then becomes how to enable agents to develop, via a simple distributed strategy (without knowledge of the global optimum, of course), such a perception of interactions with others that causes them to make these globally optimal decisions.

The strategy we investigate is very simple—we assert that each matrix $P_{ij}$ is updated so as to increase the agent's perceived utility at the current moment. Specifically, whenever an agent's behavior has just been updated (whether it changed behavior or not), with probability $r_p = 0.0001$ all of its matrices $P_{ij}$ will also be updated. To decide how to update each $P_{ij}$, one of two possibilities is considered (chosen at random): Either $P_{ij}' = P_{ij}(t) + A$, or $P_{ij}' = P_{ij}(t) - A$, where $A$ is the adjustment matrix defined in Table 2. If $[p_i(t) \text{ given } P_{ij}'] > [p_i(t) \text{ given } P_{ij}]$ then $P_{ij}(t+1) = P_{ij}'$ else $P_{ij}(t+1) = P_{ij}$.

This strategy has the effect of increasing agent $i$'s preference for coordinating or anticoordinating with agent $j$ according to whether it is currently coordinating or anticoordinating with agent $j$, respectively. Note that this preference is not sensitive to whether the interaction between these two agents is currently contributing positively to the utility of agent $i$; an agent increases its preference for the current combination of behaviors irrespective of whether $U_{ij}(s_i(t),s_j(t)) > 0$. It is thereby simply reinforcing a preference for doing more of what it is currently doing with respect to coordinating with others (i.e., I'm in the same pub with them now; therefore I change my preference so that I like being in the same pub with them a little more or, at least, dislike it less). Hence the reference to the Crosby, Stills & Nash song in our title. This is a counterintuitive strategy in that it can increase the preference for coordinating with other agents even when $U_{ij}$ defines an anticoordination game, and vice versa. Note that this habituation does not alter the independent preference for playing behavior A or B, but instead alters the preference for coordinating behaviors with others.

## 3 Results

The system is run for 1000 relaxations, of 1000 time steps each, without habituation (i.e., default agents). Example trajectories of total utility for individual relaxations are shown in Figure 1a. The total utility at the endpoint of each relaxation is shown in Figure 1b (first 1000 relaxations). The system is then run for 1000 relaxations with habituation ($r = 0.005$). As the preference utility matrices change over time, the distribution of local optima found changes (Figure 1b, relaxations 1001–2000). We see in these figures that the probability of finding the configurations with high total utility increases over time. After this habituation period the system is then tested with 1000 further relaxations (using the habituated agents, but with no further habituation). The plotted results show that the trajectories of the system after habituation (Figure 1c) find high-utility configurations reliably. Histograms of the total utilities found before and after habituation are shown in Figure 1d.

Table 2. Adjustment matrix $A$ ($r = 0.005$).

| | | Player 2 | |
|---|---|---|---|
| | | A | B |
| Player 1 | A | $r$ | $-r$ |
| | B | $-r$ | $r$ |

These results therefore show that habituation of agent interactions, created by developing a preference for whatever combination of behaviors is currently observed, has the effect of causing agents to adopt different behaviors in some situations (essentially because the resulting combination of behaviors has been experienced more often in the past). Specifically, since *without* habituation agents adopt behaviors that maximize their individual (true) utility, the different behaviors adopted *with* habituation decrease (at least temporarily) their true utility—otherwise, the trajectories would not be different (neutral changes are very rare in this system). Over time, agents therefore come to choose behaviors that decrease their individual utility in certain circumstances, but that allow the system to ultimately reach states of global utility higher than would have otherwise been possible. Accordingly, trajectories before and after habituation are different, but more specifically, the behavioral choices that agents make after habituation increase total system utility and are in this well-defined sense more cooperative.

The results collected for 50 independent simulations (each consisting of 1000 relaxations before habituation, 1000 relaxations during habituation, and 1000 relaxations after habituation) show that with these parameters, the global utility of system configurations found after habituation is on average in the 93rd percentile of global utilities of system configurations found before habituation. This represents a considerable increase in the likelihood of finding a high-utility system configuration, but the current learning rate ($r = 0.005$) does not always cause the system to ultimately settle exactly on the global optimum. In general, as one might expect, there is a tradeoff between the speed with which learning can improve global adaptation and the probability that the system will ultimately settle on the globally maximal attractor. But a sufficiently low learning rate will find the global optimum utility configuration with high probability [38, 39].

## 4 Discussion

### 4.1 Adaptive Networks

An agent system where actions are governed by a perceived utility (rather than the true utility) is formally equivalent to a system where actions are governed by a new network of constraints (rather than the original network of constraints) [6]. Here we have been modeling a system that is fully connected, with coordination and anticoordination games played on the edges of that network. This is equivalent to a weighted network, where edges are weighted by $\omega_{ij} = \pm 1$, and all games are coordination games ($\alpha = 1$, $\beta = 0$) with payoff $\omega_{ij}U_{ij}$ (i.e., each of the entries in $U_{ij}$ is multiplied by the scalar $\omega_{ij}$). The structure of the games defined by the payoff matrices is thus converted into the weighted connections of the network (with identical payoff matrices). Further, the addition of a preference matrix (restricted to the limited form investigated here) is equivalent to an alteration of this weighting; specifically, $(\omega_{ij} + k_{ij}r)U_{ij}$, where $r$ is the learning rate (as above) and $k_{ij}$ is the number of times agents $i$ and $j$ have been coordinated in the past minus the number of times they have been anticoordinated (note that $k_{ij}$ will always equal $k_{ji}$, ensuring that the connections remain symmetric if they start symmetric). Thus, although conceptually contrasting, changing the perception of payoffs for agent $i$ via a preference matrix is functionally identical to altering the connection strengths between the agents. We chose not to introduce the model in these terms, in part because it is important to realize that although an agent's behaviors will be governed by the new connections, the effects on global true utility that we are interested in must be measured using the original connection strengths [39] (it should be clear that if this were not the case, it would be trivial for agents to alter connections in a manner that would make satisfying constraints easier for them and thereby increase total utility).

Nonetheless, this perspective helps us to connect the current work with studies of adaptive networks [7, 22, 37] where agents on a network can alter the topology (here, connection strengths) of connections in it. We can thereby understand the system we have illustrated to be an example of how agents on a network can restructure the network in a manner that enhances the resolution of conflicting constraints and thereby global efficiency. Other works in this area include that of [26, 32],

where agents on a network, playing a variety of games, rewire their links when their utility is low, but keep the local topology unchanged if their utility is high. Although it displays several important technical differences from the current work, the basic intuition that agents should alter network topology to make themselves happier (or at least, alter it if they are unhappy) appears in both [42].

In essence, the form of habituation we model is a very simple form of restructuring; it simply asserts that connections between agents increase or decrease in strength in a manner that reinforces the current combinations of behaviors observed. The effects of this habituation are put into context by considering the problem at hand: We are dealing with a limited form of global optimization problem [41] in which local optima (and the global optimum) are created by the inability to resolve many overlapping low-order dependencies [15, 39]. When using simple local search on this problem (i.e., agents without habituation), there is only a small probability of finding configurations with high global utility (Figure 1a,b); however, they are found nonetheless. Habituation outcompetes local search, not by finding new configurations of higher utility (although this may occur in some cases), but instead by progressively increasing the *probability* of finding high-utility configurations, until only one configuration is ever found (which is very likely to be one of high utility). We can therefore view habituation as a mechanism that gradually transforms the search space of the problem from one with many varied local optima to one with a single (and very likely high-utility) optimum, which will always be reached; furthermore, it does so via a simple distributed strategy [39, 42].

Specifically, although it is not immediately obvious from a static analysis of the connection matrix which connections should be increased and which decreased in order to cause selfish agents to solve the problem better, the necessary information is naturally revealed by allowing the system to repeatedly settle to local optima and reinforcing the correlations in behaviors so created. These correlations are determined by the connections of the original network in an indirect manner. For example, a particular constraint may often remain unsatisfied in locally optimal configurations even though the direct connection defining this constraint states that it is just as valuable to satisfy it as any other connection. Then, if a constraint is often easily satisfied, its importance is strengthened; if it is equally often satisfied and unsatisfied, it remains unchanged on average; and when agents are on average unable to satisfy it, its importance is weakened and eventually its sign can be reversed. This causes the system to (gradually over time) pay more attention to the connections that can be simultaneously satisfied and weaken or soften the constraints that cannot be satisfied. One way to understand the result of this adaptive constraint relaxation and exaggeration is that agents become complementary specialists, that is, selectively attuned to some constraints more than others. That is, whereas the default agents are generalists that persist in trying to satisfy all constraints whether satisfiable or unsatisfiable, habituating agents, through the self-organization of the behaviors in the network, come to specialize in a manner that "for their own comfort" (i.e., for the immediate increase of their perceived utility) fits them together better with one another but thereby actually satisfies more of the system constraints overall.

### 4.2 Self-Structuring Adaptive Networks, Neural Network Learning, and Associative Memory

How this type of adaptive network, with very simple, local modification of connections, comes to maximize global utility can be explained formally using theory from computational neuroscience. Specifically, the behavior of the network of default agents detailed above is identical to the behavior of the discrete Hopfield network [11] (which is just a bit-flip hill climber [2]), and when connections between nodes increase or decrease in strength in a manner that reinforces the current combinations of behaviors, this is formally equivalent to *Hebbian* learning [42]. Hebb's rule, in the context of neural network learning, is often represented by the saying "neurons that fire together wire together," meaning that synaptic connections between neurons that have correlated activation are strengthened. This learning rule has the effect of transforming correlated neural activations into causally linked neural activations, which, from a dynamical systems perspective, has the effect of enlarging the basin of attraction for the current activation-pattern/system configuration. This type of learning

can be used to train a recurrent neural network to store a given set of training patterns [11], thus forming what is known as an *associative memory* of these patterns. A network trained with an associative memory then has the ability to "recall" the training pattern that is most similar to a partially specified or corrupted test pattern.

Formally, a common simplified form of Hebb's rule states that the change in a synaptic connection strength $\omega_{ij}$ is $\Delta\omega_{ij} = \delta s_i s_j$, where $\delta > 0$ is a fixed parameter controlling the learning rate, and $s_n$ is the current activation of the $n$th neuron. Here, by changing the payoff matrix of each individual by $k_{ij}(t)rU_{ij}$, where $k_{ij}(t)$ is the correlation of behaviors at time $t$, we are effecting exactly the same changes. Thus the habituating agents each modify their perceived utilities in a manner that effects Hebbian changes in connection strengths—which they must if these preferences are to mean that this behavior combination is preferred more. This equivalence at the agent level has the consequence that the system of agents as a whole implements an associative memory. Since this is a self-organized network, not a network trained by some external experimenter, this is not an associative memory of any externally imposed training patterns. Rather, this is an associative memory of the configuration patterns that are commonly experienced under the network's intrinsic dynamics—and (given the perturbation and relaxation protocol we have adopted, which means that the system spends most of its time at locally optimal configurations) it is these configurations that the associative memory stores (or enlarges).

From a neural network learning point of view, a network that forms a memory of its own attractors is a peculiar idea (indeed, the reverse is more familiar [12]). Forming an associative memory means that a system forms attractors that represent particular patterns or state configurations. For a network to form an associative memory of its own attractors therefore seems redundant; it will be forming attractors that represent attractors that it already has. However, in forming an associative memory of its own attractors, the system will nonetheless alter its attractors; it does not alter their positions in state configuration space, but it does alter the size of their basins of attraction (i.e., the set of initial conditions that lead to a given attractor state via local energy minimization).

Specifically, the more often a particular state configuration is visited, the more its basin of attraction will be enlarged and the more it will be visited in future, and so on. Because every initial condition is in exactly one basin of attraction, it must be the case that some attractor basins are enlarged at the expense of others. Accordingly, attractors that have initially small basins of attraction will be visited infrequently, and as the basins of other, more commonly visited attractors increase in size, these infrequently visited attractors will decrease. Eventually, with continued positive feedback, one attractor will outcompete all others, resulting in there being only one attractor remaining in the system.

But what has this got to do with resolving the constraints that were defined in the original connections of the system? One might expect, given naive positive feedback principles, that the one remaining attractor would have the mean or perhaps modal global utility of the attractor states in the original system; but this is not the case (Figure 1d). In order to understand whether the competition between attractors in a self-modeling system enlarges attractors with especially high total utility or not, we need to understand the relationship between the size of attractor basins and the total utility of their attractor states. At first glance it might appear that there is no special reason why the largest attractor should be the best (highest-utility) attractor—after all, it is not generally true in optimization problems that the basin of attraction for a locally optimal solution is proportional to its quality. But in fact, existing theory tells us that this is indeed the case [15] for systems that are additively composed of many low-order interactions. Specifically, in systems that are built from the superposition of many symmetric pairwise interactions, the height (with respect to total utility) of an attractor state is positively related to its width (the size of its basin of attraction), and the globally optimal attractor state has the largest basin of attraction [39]. One must not conflate, however, the idea that the global optimum has the largest basin with the idea that it is a significant proportion of the total configuration space and therefore easy to find: In particular, the global optimum may be unique, whereas there will generally be many more attractors that lead to inferior solutions, and importantly, the basins of these suboptimal attractors will collectively occupy much more of the configuration space than the basin of the global optimum.

Given that high-utility attractors have larger basins than low-utility attractors, they are therefore visited more frequently and therefore outcompete low-utility attractors in this self-modeling system. Thus (in the limit of low learning rates such that the system can visit a sufficient sample of attractors) we expect that when a dynamical system forms an associative memory model of its own utility maximization behavior, it will produce a model with ultimately only one attractor, and this attractor will correspond to the globally optimal minimization of constraints between variables in the original system [39].

This is not an entirely satisfactory conclusion, however. It implies that the system only fixes on the global optimum because the global optimum has already been visited many times in the past. But this is not the full story. A final part of the puzzle is provided by the well-known ability of Hebbian learning to generalize training patterns and create learned attractors that represent new combinations of common features from the training patterns rather than the training patterns per se. In associative memory research the creation of such *spurious attractors* is generally considered to be a nuisance [5], but it in fact represents a simple form of generalization that is important for our results [42]. Producing new attractor states that are new combinations of features (subpatterns) observed in the training patterns [13] enables the globally optimal attractor to be enlarged even though it has not yet been visited. Basically, this occurs because when Hebbian learning is applied to a training pattern, it not only has the effect of enlarging the basin of attraction for this pattern, but also enlarges the basin of attraction for all configurations in proportion to how many behavior pairs they share. The global optimum is, by definition, the configuration that has the most simultaneously satisfied constraints, and this ensures that, on average at least, it tends to share many behavior combinations with locally optimal configurations (each of which has many constraints simultaneously satisfied, but not as many as globally possible).

In addition, it is important to recognize how the separation of the timescales for behaviors *on* the network and behaviors *of* the network (i.e., changes to network structure) influence this result. Getting the timescale of the changes to network structure correct is equivalent to setting the learning rate correctly in a neural network. If connections are modified too slowly, then learning is unnecessarily slow. And if learning happens too quickly, the network will only learn the first local optimum it arrives at, or worse, if the learning rate is really high, the system could get stuck on some transient configuration that is not even locally optimal [39]. More generally, if most learning happens at or near random initial conditions, then the patterns learned will be similarly random. It is therefore essential that the system be allowed to relax to local optima, and that most learning therefore happen at local optima, so that the patterns learned are better than random. But if the system is not perturbed frequently enough or vigorously enough, and consequently spends all of its time at one or a few local optima, the system will simply learn these attractor configurations and will not generalize correctly.

Lastly on this equivalence, it should be noted that the Hopfield model is not new [11], and its capabilities for Hebbian learning are well known [12]. However, here we provide a reinterpretation of the system, staging it in a generic, game theoretic network scenario, which opens up the possibility of reinterpretation of some of the analytically solved variants of the Hopfield model (e.g., [3, 14]).

### 4.3 Why Adopt a Habitual Policy?

In the model presented here we mandate that agents follow a habitual policy. Habituation, or a preference for the status quo, is not an uncommon phenomenon in natural systems, but if such a policy were optional, would it be in the self-interest of agents to adopt it? Related work examines this question in detail and shows that when agents can directly alter their utility by altering their constraints with other agents, self-interested agents will always alter those constraints in a Hebbian manner, thereby causing habituation [42]. However, in the current model we are examining a subtly different scenario where agents can alter their perception of an interaction but cannot alter the true value of that interaction. We have shown successfully in this article that a (mandated) policy of altering perceptions in a habituating manner has the effect of increasing the future likelihood of the system reaching attractors that have high true utility. But since altering a *perception* per se has no direct effect on *true* utility, a selfish agent has no direct or immediate reason to alter its perceptions unless (a) that change in

perception causes the agent to immediately alter its behavior, (b) this change in behavior immediately increases utility. So, since altering perceptions in a habituating manner necessarily has the effect of reinforcing or stabilizing the current behavior, selfish changes to perceptions will only be habituating when reinforcing the current behavioral state is desirable.

In general, it is not guaranteed that the current behavioral state, nor therefore reinforcement of the current behavioral state, is desirable for an agent. But, assuming that an agent updates its behaviors rapidly compared to its perceptions (separation of timescales [39]), as is the case here, the agent will most of the time exhibit the behavioral choice that is currently the best compromise of its conflicting interests. Whenever this is the case, any change to perceived utility that increases the magnitude of the current behavior will be preferred over a change to perceived utility that decreases the current behavior. Thus, although it is not the case for all behavioral configurations that the changes to perceptions that increase utility are Hebbian, they will be Hebbian for the behavioral configurations that are most common under the condition of separation of timescales. In short, under these conditions, Hebbian learning (i.e., preferring the current state) becomes equivalent to reinforcement learning (i.e., preferring what is good). This reasoning suggests that, given the option, selfish agents would change preferences in the same manner as the habituation policy we mandate in this article—not because of the future increases in total utility they afford, but because of immediate individual utility gains.

However, one final complication is introduced by the possibility that the current behavioral state may already be locally optimal in magnitude as well as sign (in particular, if behaviors are discrete). In this case, there is no change to behavior that can increase utility. But even here, note that an anti-Hebbian change to perceptions might cause behaviors to move away from this locally optimal state, causing a decrease in utility, whereas a Hebbian change to perceptions is at worst neutral. This selection *against anti*-Hebbian changes (plus stochastic variation) is sufficient under some circumstances for a selfish agent to exhibit a systematic trend toward Hebbian changes [40, 42, 43]. We have been studying these mechanisms in several evolutionary scenarios: coevolving species in an ecosystem that can alter the coefficients of a Lotka-Volterra system [17, 28], coevolving symbiotic relationships that affect codispersal probabilities [43], and the evolution of a gene interactions that affect the correlation of phenotypic traits [40]. This work provides examples of the more general concept we refer to as *social niche construction*—the ability of an agent to create social contexts that subsequently alter social behavior [24, 27, 29, 30, 33]. Collectively these works suggest that reinforcing the status quo, and thereby exhibiting a tendency to recall or recreate behavioral patterns exhibited in the past, may be a widespread property of adaptive multi-agent systems [42]. Formal neural network models (where a Hebbian assumption is normal) then help us to better understand the conditions where this tendency will increase total welfare [39, 42].

### 4.4 Timescales of Selfishness

Habituating agents start out behaving exactly the same as default agents—making individual utility-maximizing behavioral decisions. However, as they change their perceived utilities, this begins to influence their decisions, causing them to adopt different behaviors than default selfish agents. Thus, even though the previous section shows that habituation can be motivated by immediate selfish interests, the habituated policy so created can cause agents to subsequently adopt behaviors that decrease individual utility. Since ultimately these new behaviors result in higher total utility, they are arguably cooperative, not selfish. However, since the habituated policy leads to higher global utility in the long term, and global utility is simply the sum of individual utilities, it must mean that under the habituated policy, on average each agent is likely to have higher individual true utility in the long term than under the default selfish policy. So, which agents are the selfish ones—the default agents or the habituated agents?

We propose a novel resolution to this inconsistency: *Both* policies are selfish, but simply selfish over different timescales. That is, a rational selfish agent that was attempting to maximize its utility *in the immediate term* would adopt the default policy rather than the habitual policy. But, given a choice and the appropriate information, a rational selfish agent that was attempting to maximize its utility *in the long term* would adopt the habitual policy rather than the default policy because, although this

policy might cause it to suffer utility decreases in the short term, on average it leads to higher-total-utility attractors at equilibrium.

This implies a refined definition of "selfishness": A behavioral policy can be described as selfish if it maximizes individual utility (given the available information) on *some timescale or other*. Such selfish behavioral policies can be subcategorized based on the timescale over which they maximize utility: the behavioral policy's *timescale of selfishness*. This classification leads to the possibility of having two or more behavioral policies that can equally be said to be selfish (i.e., they each maximize utility on some timescale or other), but nonetheless, because they are selfish on different timescales, may cause different behaviors—even when faced with identical environments.

We therefore describe behavioral policies that optimize utility over some timescale other than the immediate term as (to some extent) *long-term selfish*. They are selfish policies that have nonimmediate timescales of selfishness. For a given system, if there exists a higher utility optimum that cannot be reached by following local gradients, then a policy that systematically reaches this nonlocal utility optimum would be described as long-term selfish. In addition, to reach a nonlocal utility optimum, it must be that such a policy entails occasionally accepting temporary decreases in utility (otherwise it would be forced to follow local gradients, and hence end up at the local utility optimum). Furthermore, the ability to systematically reach a nonlocal, higher utility optimum logically implies that such a policy must have access to information—in some form or other—about that system beyond that of local utility gradients. This information must be stored somewhere, and it must influence the behavior of the agents under this policy. Another way of describing such an information store that affects decision making is as a *memory*: Essentially, we claim that such a long-term selfish policy (or a mechanism, like our habituation mechanism, that causes an agent to act in a manner consistent with a long-term selfish policy) necessarily requires (overtly or otherwise) a functional memory that contains or acquires information about the long-term utility of behavioral choices in the system.

Under this classification, the habitual agents of the current model act in a manner consistent with a long-term selfish behavioral policy, in the sense that it maximizes utility, but does so on a timescale longer than the immediate. The information about long-term utility that such a policy requires—the memory—is distributed and is stored in the preference matrices of all of the agents. Habitual agents (and not default agents) always possess the *machinery* of such a memory, although at the start of the simulation it contains no information (and hence at this stage it has no effect on behavior, and so habitual agents act as default agents). It is composed of the storage medium (preference matrices), a write mechanism (habituation), and a read mechanism (behavior decisions of habitual agents are based on perceived utility). And it is only once this memory has been populated with information about nonlocal utility optima (i.e., after repeated perturbations, so that the system has had a chance to experience more than one optimum) that it begins to cause behavioral differences from the default policy. Default agents, on the other hand, do not possess such memory machinery: They have no mechanism that can store information about the system, and hence their behavioral choices are always based solely on immediate utility gradients, limiting them to simple, immediate-term utility optimization, which is always destined to find only the local utility optimum.

### 4.5 Reinforcement Learning and Fictitious Play

In the previous sections we have discussed two points that indirectly link the habituation policy to mechanisms of reinforcement learning [35]. The first point is that in the system modeled in this article, due to the separation of timescales (i.e., that agent behaviors change much faster than agent perceptions), the policy of habituation, or "do more of what you are currently doing," is effectively equivalent to "do more of what is good." The result is that agents learn to prefer behaviors that have been successful in the past—which is the central mechanism of reinforcement learning. The second point is that making decisions that are influenced to some degree by historical information (such as by preference matrices in the current model) can be viewed in some cases as if agents were "attempting to" predict future outcomes. In some systems, even if agents are myopic (such as habituated agents, which have no "intention" of predicting long-term outcomes of their actions), the structure of the

system itself can be enough to effectively convert this myopic behavioral policy into one that is indistinguishable from a policy where agents overtly reason about the future by generalizing from the past (i.e., reinforcement learning agents). The model shown here has the form of such a myopic system that, when considered as a whole, has the effect of optimizing in a manner that appears very similar to reinforcement learning.

In particular, the behavior of the model can be linked to the fictitious play algorithm [1, 31]. Under a policy of pure fictitious play, each agent records past opponent behaviors and then makes the assumption that opponents will adopt a randomized strategy that selects each behavior with a probability given by the frequency at which it has been played by that opponent in the past. The agent then best-responds to the joint strategy (i.e., cross product) of all of its opponents' strategies. Ignoring the distributed nature of our model, the habitual policy appears to lie under the umbrella of fictitious play algorithms as described for extensive form games (i.e., asynchronous behavior updating) [31]. Even from this perspective, however, habituation differs from the standard description of extensive form fictitious play, largely due to the strong discrepancy in the habitual model between the influence of a given opponent's *current* behavior and any of its individual *historical* behaviors: Habituation effectively involves a fixed historical discounting of influence, where opponents' current behaviors always have significantly more influence on decision making than behaviors from any individual past move (although when acting as a whole, recorded historical behaviors can come to overwhelm the influence of opponents' current behaviors). This therefore links habituation to generalized weakened fictitious play [16], a more general class of fictitious play algorithms that can include similar forms of historical discounting.

There are therefore clear parallels between the current model and traditional systems of reinforcement learning, but there are also important differences to be highlighted. First, theoretically, the recognition that this type of individual learning between agents is equivalent to associative learning familiar in neural networks is important for understanding the global adaptation that results. And second, mechanistically, the current model uses the repeated relaxation protocol and the separation of timescales to enable the associative memory thus formed to generalize over many locally optimal equilibria. Without these dynamical conditions, the behavior of habituating agents would not be equivalent to reinforcement learning, and progressive system-level optimization would not occur.

## 5 Conclusions

This article has investigated the effect of a simple distributed strategy for increasing total utility in systems of selfish agents. Specifically, habituating selfish agents develop a preference for coordinating behaviors with those they are coordinating with at the present moment, and henceforth adopt behaviors that maximize the sum of true utility and these preferences. We show that this causes agents to modify the dynamical attractors of the system as a whole in a manner that enlarges the basins of attraction for system configurations with high total utility. This means that after habituation, agents sometimes make decisions about their behavior that may (at least temporarily) decrease their personal utility but that in the long run increase (the probability of arriving at configurations that maximize) global utility. We show that the habituating agents effectively restructure the connections in the network in a Hebbian manner, and thus, through the simple distributed actions of each individual agent, the network as a whole behaves in a manner that is functionally equivalent to a simple form of learning neural network. This network improves global adaptation by forming an associative memory of locally optimal configurations that, via the inherent generalization properties of associative memory, enlarges the basin of attraction of the global optima. This work thereby helps us to understand self-organization in networks of selfish agents and very simple processes that subtly deviate selfish agents in the direction that maximizes global utility without overtly prescribing cooperation or using any form of centralized control.

## Acknowledgments

We thank Archie Chapman, Alex Penn, Simon Powers, Seth Bullock, and Anna Jewitt for helpful discussions.

## References


1. Brown, G. W. (1951). Iterative solution of games by fictitious play. In T. C. Koopmans (Ed.), *Activity analysis of production and allocation* (pp. 374–376). New York: Wiley.
2. Chapman, A. C., Rogers, A., Jennings, N. J., & Leslie, D. S. (in press). A unifying framework for iterative approximate best response algorithms for distributed constraint optimisation problems. *The Knowledge Engineering Review*.
3. Coolen, A. C. C., & Sherrington, D. (1993). Dynamics of fully connected attractor neural networks near saturation. *Physical Review Letters*, *71*(23), 3886–3889.
4. Dawkins, R. (2006). *The selfish gene* (3rd ed.). Oxford, UK: Oxford University Press.
5. Gascuel, J. D., Moobed, B., & Weinfeld, M. (1994). An internal mechanism for detecting parasite attractors in a Hopfield network. *Neural Computation*, *6*(5), 902–915.
6. Gross, T., & Blasius, B. (2008). Adaptive coevolutionary networks: A review. *Journal of the Royal Society Interface*, *5*(20), 259–271.
7. Gross, T., & Sayama, H. (Eds.). (2009). *Adaptive networks*. Cambridge, MA: NESCI.
8. Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
9. Heylighen, F., Gershenson, C., Staab, S., Flake, G. W., Pennock, D. M., Fain, D. C., De Roure, D., Aberer, K., Shen, W. M., Dousse, O., & Thiran, P. (2005). Neurons, viscose fluids, freshwater polyp hydra—And self-organizing information systems. *Intelligent Systems*, *18*(4), 72–86.
10. Hinton, G. E., & Sejnowski, T. J. (1983). Analyzing cooperative computation. In *Proceedings of the Fifth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
11. Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8), 2554–2558.
12. Hopfield, J. J., Feinstein, D. I., & Palmer, R. G. (1983). "Unlearning" has a stabilizing effect in collective memories. *Nature*, *304*, 158–159.
13. Jang, J. S., Kim, M. W., & Lee, Y. (1992). A conceptual interpretation of spurious memories in the Hopfield-type neural network. In *IJCNN, International Joint Conference on Neural Networks*, Vol. 1 (pp. 21–26).
14. Kryzhanovsky, B. V., Simkina, D. I., & Kryzhanovsky, V. M. (2009). A vector model of associative memory with clipped synapses. *Pattern Recognition and Image Analysis*, *19*(2), 289–295.
15. Kryzhanovsky, B. V., Magomedov, B. M., Mikaelian, A. L., & Fonarev, A. B. (2006). Binary Optimization: A relation between the depth of a local minimum and the probability of its detection. *Optimal Memory and Neural Networks*, *15*(3), 170–182.
16. Leslie, D. S., & Collins, E. J. (2006). Generalised weakened fictitious play. *Games and Economic Behavior*, *56*(2), 285–298.
17. Lewis, M. (2009). *An investigation into the evolution of relationships between species in an ecosystem*. Unpublished M.Sc. dissertation, University of Southampton.
18. Maynard Smith, J., & Szathmáry, E. (1997). *The major transitions in evolution*. Oxford, UK: Oxford University Press.
19. Michod, R. E. (2000). *Darwinian dynamics: Evolutionary transitions in fitness and individuality*. Princeton, NJ: Princeton University Press.
20. Monderer, D., & Shapley, L. S. (1996). Potential games. *Games and Economic Behavior*, *14*(1), 124–143.
21. Nettleton, R. W., & Schloemer, G. R. (1997). Self-organizing channel assignment for wireless systems. *Communications Magazine*, *35*(8), 46–51.
22. Newman, M. E., Barabasi, A. L., & Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton, NJ: Princeton University Press.
23. Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560–1563.
24. Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution*. Princeton, NJ: Princeton University Press.

25. Ohtsuki, H., & Nowak, M. A. (2006). The replicator equation on graphs. *Journal of Theoretical Biology*, *243*(1), 86–97.

26. Pacheco, J. M., Lenaerts, T., & Santos, F. C. (2007). Evolution of cooperation in a population of selfish adaptive agents. In F. Almeida e Costa (Ed.), *Proceedings of the 9th European Conference on Advances in Artificial Life* (pp. 535–544).

27. Penn, A. (2006). *Ecosystem selection: Simulation, experiment and theory*. Unpublished doctoral dissertation, University of Sussex.

28. Poderoso, F. C., & Fontanari, J. F. (2007). Model ecosystem with variable interspecies interactions. *Journal of Physics A: Mathematical and Theoretical*, *40*(30), 8723.

29. Powers, S., Mills, R., Penn, A., & Watson, R. A. (2009). Social niche construction provides an adaptive explanation for new levels of individuality. In *Proceedings of Workshop on Levels of Selection and Individuality in Evolution (ECAL 2009)*.

30. Powers, S., Penn, A., & Watson, R. (2007). Individual selection for cooperative group formation. In F. Almeida e Costa (Ed.), *Proceedings of the 9th European Conference on Advances in Artificial Life* (pp. 585–594).

31. Robinson, J. (1951). An iterative method of solving a game. *Annals of Mathematics*, *54*(2) 296–301.

32. Santos, F. C., Pacheco, J. M., & Lenaerts, T. (2006). Cooperation prevails when individuals adjust their social ties. *PLoS Computational Biology*, *2*(10), e140.

33. Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge, UK: Cambridge University Press.

34. Stills, S. (1970). *Love the one you're with*. New York: Atlantic.

35. Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.

36. Taylor, C., & Nowak, M. A. (2007). Transforming the dilemma. *Evolution*, *61*(10), 2281–2292.

37. Van Segbroeck, S., Santos, F. C., Pacheco, J. M., & Lenaerts, T. (2010). Coevolution of cooperation, response to adverse social ties and network structure. *Games*, *1*(3), 317–337.

38. Watson, R. A., Buckley, C. L., & Mills, R. (2009). *The effect of Hebbian learning on optimisation in Hopfield networks* (ECS Technical report). Southampton, UK: University of Southampton.

39. Watson, R. A., Buckley, C. L., & Mills, R. (2010). Optimization in "self-modeling" complex adaptive systems. *Complexity*, http://onlinelibrary.wiley.com/doi/10.1002/cplx.20346/pdf.

40. Watson, R. A., Buckley, C. L., Mills, R., & Davies, A. P. (2010). Associative memory in gene regulation networks. In H. Fellerman et al. (Eds.), *Proceedings of the 12th International Conference on Artificial Life* (pp. 194–203).

41. Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for search. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82.

42. Watson, R. A., Mills, R., & Buckley, C. L. (2011). Global adaptation in networks of selfish components: Emergent associative memory at the system scale. *Artificial Life*, *17*(3), 147–166.

43. Watson, R. A., Palmius, N., Mills, R., Powers, S., & Penn, A. (2009). Can selfish symbioses effect higher-level selection? In G. Kampis et al. (Eds.), *Proceedings of 10th European Conference on Artificial Life (ECAL 2009)*, LNCS 5778, Part II, pp. 27–36. Berlin: Springer.