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. Davies¹, Richard A. Watson¹, Rob Mills¹, C. L. Buckley², Jason Noble¹

¹Natural Systems group, ECS, University of Southampton, U.K. ²CCNR, University of Sussex, U.K. apd1e09@ecs.soton.ac.uk

Abstract

Simple distributed strategies that modify the behaviour 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 optimise their individual utilities by coordinating (or anti-coordinating) with their neighbours, to maximise the pay-offs from randomly weighted pair-wise games. In general, agents will opt for the behaviour that is the best compromise (for them) of the many conflicting constraints created by their neighbours, but the attractors of the system as a whole will not maximise total utility. We then consider agents that act as 'creatures of habit' by increasing their preference to coordinate (anti-coordinate) with whichever neighbours they are coordinated (anti-coordinated) with at the present moment. These preferences change slowly while the system is repeatedly perturbed such 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 maximise (or almost maximise) global utility. This counterintutitve 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 optimisation of global utility that is observed results from wellknown generalisation 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 maximise total utility.

Selfish Agents and Total Utility

This paper investigates the effect of a simple distributed strategy for increasing total utility in systems of selfishly optimising individuals. The broader topic concerns many different types of systems. 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, acts to optimise a global performance criterion: e.g. communications routing agents act to minimise calls dropped, or processing nodes in a grid computing system each act to maximise the number of jobs processed (1,2). However, since each component in the network acts individually, i.e., using only local information, constraints between individuals can remain unsatisfied, resulting in

poorly optimised global performance. In an engineered system one could, in principle, mandate that all nodes act in accord with the globally optimal configuration of behaviours (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 the question of 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 optimise 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 organisation (3), 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 (4,5). In social network studies there is increasing interest in adaptive networks (6) where agents in a network can alter the structure of the connections in the network. Of particular interest is the possibility that by doing so they may increase the ability of the system to maintain high levels of cooperation (7,8). However, a general understanding of how agents on a network modify their interactions with others in a way that increases total cooperation is poorly understood. In each domain we are, at the broadest level, interested in understanding/identifying very mechanisms that might cause self-interested agents to modify their behaviour, or how their behaviours are affected by others, in a manner that increases adaptation or efficiency either globally or at a higher-level of organisation 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 behaviours of individuals, then it must be the case that, on at least some occasions, agents take decisions that are detrimental to individual interests. If this were not the case then there is nothing to be explained over and above the selfish actions of individuals. But if it is the case, then this runs 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 behaviour 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 them be the best thing for them - that the agent is no longer making decisions according to the true utility function but some distortion of it that alters their 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 paper 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 is 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 about scientists and their drinking habits, but a general model of interacting agents on a network with pair-wise constraints between binary behaviours.

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/pay-off 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 maximise their scientific productivity by attending the pub that, on that night, maximises the sum of compatibilities with other researchers and minimises incompatibilities. Assessing the company they find at any moment, individuals therefore (one at a time in random order) may choose to switch pubs to maximise 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 (9), 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 behaviour of agents on the network. Our aim is to devise a simple individual strategy that causes researchers to make better decisions about when to change pubs such that total productivity is maximised. 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 and Nash put it "If you can't be with the one you love, honey, love the one you're with" (10). Since we already know the arrangements of researchers will be initially random and, most of the time, at best sub-optimal, this seems like a counter productive 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 maximise total productivity reliably every evening.

The agents that we model are therefore not wholly selfish agents – they sometimes take actions that do not maximise 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 paper 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 which does not require any teleological or, certainly, any centralised control and is therefore relevant to many domains.

In the next two sections we will detail an illustration of this strategy and the results we observe. In the Discussion section 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 latterly they are governed by a combination of these intrinsic constraints plus the interaction preferences that the agents have developed. The new behavioural dynamics of the agents caused by interaction preferences can therefore equally 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 (11.12). 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 (13). 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 behaviours on the network and behaviours of the network (i.e. changes to network structure) is essential for this result.

Methods

Default agents

Our model involves N=100 agents playing two-player games on a fully connected network (Table 1). 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 anti-coordination game ($\alpha=0$, $\beta=1$) with equal probability (Table 1).

		Player 2	
		Α	В
Player	Α	α,α	β,β
1	В	β,β	α,α

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

Games are played in extensive form, i.e., initially all agents in the network are assigned a behaviour at random, and then each agent in random order is permitted to update its behaviour (to either A or B). Each agent does so according to a best response strategy, i.e., to adopt the behaviour (choose a pub) that maximises its utility, u_i (Eq.1) given the behaviours (pub choices) currently adopted by its neighbours:

$$u_{i}(t) = \sum_{j}^{N} U_{ij}(s_{i}(t), s_{j}(t))$$
 (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 above), and $s_n(t)$ is the strategy currently played by agent n. Behaviours are updated in this manner repeatedly. Each agent is involved in many games but can adopt only one behaviour at any one time, thus coordinating with one neighbour may preclude coordinating with another, and so each agent must therefore adopt the behaviour 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 (9), i.e. a configuration where no agent wants to change behaviour unilaterally (14). 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 (9) (Eq.2).

$$G_{(t)} = \sum_{i}^{N} \sum_{j}^{N} U_{ij}(s_{i}(t), s_{j}(t))$$
 (2)

However, in general, 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 behaviours to all agents) at infrequent intervals (here every 1000 time steps = one evening), and thereby allowed to settle, or relax, to many different local equilibria (on different evenings), the behaviour 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' (Fig. 1.c).

Creatures of habit

We seek a simple distributed strategy that causes agents to make different (hence unselfish) behavioural 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 neighbours 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 behaviour of the agent such that it chooses the behaviour that maximises its 'perceived utility', p_i , (Eq.3), instead of its true utility (Eq.2) alone:

$$p_{i}(t) = \sum_{j}^{N} \left[U_{ij}(s_{i}(t), s_{j}(t)) + P_{ij}(t)(s_{i}(t), s_{j}(t)) \right]$$
(3)

where P_{ij} is a pay-off matrix that represents an agent's preference for the combination of behaviours s_i and s_j . The perceived utility is thus simply the sum of the true utility plus the agent's preferences. Each agent has a separate preference pay-off matrix for each other agent. All preference payoff matrices are initially set to zero, such that the initial dynamics of the agents are as per the default agents. But as the values in these matrices change over time they may come to collectively overpower the tendency to maximise true utility and thereby cause agents to make different decisions about which behaviour 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 P_{ij} matrices that will cause agents to adopt behaviours that maximise global system utility instead of choosing behaviours that maximise individual utility and thereby failing to maximise 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 P_{ij} matrix is updated so as to increase the agent's perceived utility at the current moment. Specifically, whenever an agent's behaviour has just been updated (whether it changed behaviour or not), with probability $r_p = 0.0001$ all of its P_{ij} matrices will also be updated. To decide how to update each P_{ij} matrix, 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 $p_i(t)$ _given_ $P'_{ij} > p_i(t)$ _given_ P_{ij} then $P_{ij}(t+1) = P'_{ij}$ else $P_{ij}(t+1) = P_{ij}$.

		Player 2	
		Α	В
Player	Α	r	-r
1	В	-r	r

Table 2: adjustment matrix A(r=0.005)

This strategy has the effect of increasing agent i's preference for coordinating or anti-coordinating with agent j according to whether it is currently coordinating or anti-coordinating 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 behaviours irrespective of whether $U_{ij}(s_{i}(i),s_{j}(i))>0$. 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, so change my preference so I like being in the same pub with them a little more or dislike it less). This is a counterintuitive strategy in the sense that it can increase the

preference for coordinating with other agents even when U_{ij} defines an anti-coordination game, and vice versa. Note that this habituation does not alter the independent preference for playing behaviour A or B, but instead alters the preference for coordinating behaviours with others.

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 Fig1.a. The total utility at the end point of each relaxation is shown in Fig.1.b (first 1000 relaxations). The system is then run for 1000 relaxations with habituation (i.e. r=0.0005). As the preference utility matrices change over time the distribution of local optima found changes (Fig.1.b, relaxations 1001-2000). We see in these figures that the probability of finding the configurations with high total utility increases over time, such that the trajectories of the system after habituation (Fig.1.c) find high-utility configurations

reliably. Histograms of the total utilities found before and after habituation are shown in Fig.1.d.

These results therefore show that habituation of agent interactions, created by developing a preference for whatever combination of behaviours is currently observed, has the effect of causing agents to adopt different behaviours in some situations (essentially because the resulting combination of behaviours has been experienced more often in the past). Specifically, since without habituation agents adopt behaviours that maximise their individual (true) utility, so the different behaviours adopted with habituation are therefore behaviours that (at least temporarily) decrease 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 behaviours that decrease their individual utility in certain circumstances, but that allow the system to ultimately reach states of global utility higher than would have been otherwise possible. Accordingly, trajectories before and after habitation are different, but more specifically, the behavioural choices that agents make after habituation increase total system utility and are in this well-defined sense more cooperative.

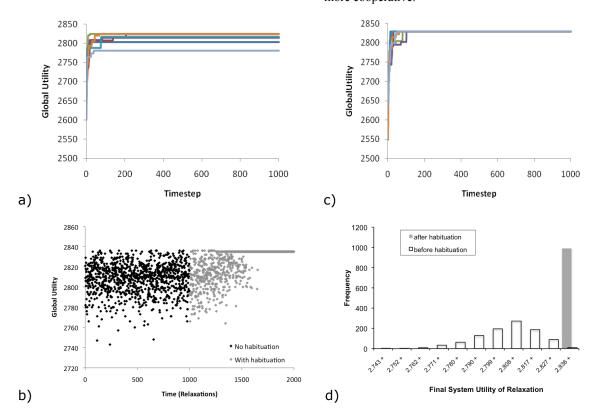


Fig.1. Behaviour of the system using default (no habituation) and habituating agents. a) Some example trajectories of system behaviour before habituation – each curve represents one relaxation (N=100, relaxation length 10N) – vertical axis is the total system utility (G, Eq 2); b) utilities of attractor states visited (i.e. end points of curves like those in (a)) without habituation (relaxations 1-1000) and during habituation (relaxations 1001-2000, r=0.0005); 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.

Results collected for 50 runs (each consisting of 1000 relaxations before habitation, 1000 relaxations during habituation and 1000 relaxations after habituation) show that with the current 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. While this represents a considerable increase in the likelihood of finding a high utility system configuration, it is clear that with the current learning rate (r=0.0005) habituation will not always cause the system to ultimately settle at the global optimum. However, it is important to note that this is simply due to the learning rate used; with a sufficiently low learning rate, after habituation the system will only ever find the global optimum utility configuration (13,15).

Discussion

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) (28). Here we have been modelling a system that is fully connected with coordination and anti-coordination games played on the edges of that network. This is equivalent to a weighted network, where edges are weighted by $\omega_{ii}=\pm 1$, and all games are coordination games ($\alpha=1$, $\beta=0$) with pay-off $\omega_{ii}U_{ii}$. (i.e. each of the table entries in U_{ii} is multiplied by the scalar ω_{ii}). The structure of the games defined by the pay-off matrices is thus converted into the connections of the network (with identical pay-off 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_{ii} is the number of times agents i and j have been coordinated in the past minus the number of times they have been anti-coordinated (note that k_{ii} will always equal k_{ii} , ensuring that the connections remain symmetric if they start symmetric). Thus, although conceptually contrasting, changing the perception of pay-offs 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 realise that although an agents' behaviours 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 (13) (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 (6) where agents on a network can alter the topology (here, connection strengths) of connections in the network. We can thereby understand the system we have illustrated to be an example of how agents on

a network can 're-structure' the network in a manner that enhances the resolution of conflicting constraints and thereby global efficiency. Other works in this area include that of (7,8) 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 there are several important technical differences with 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) is common to both.

In essence, the form of habituation we model is a very simple form of re-structuring; it simply asserts that connections between agents increase or decrease in strength in a manner that reinforces the current combinations of behaviours 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 optimisation problem (16) in which local optima (and the global optimum) are created by the inability to resolve many overlapping low-order dependencies (17, 13). 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 (Fig.1.a and b); however, they are found nonetheless. Habituation outcompetes local search, not by finding new configurations of absolute 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.

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 behaviours so created. These correlations are created 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/exaggeration is that agents become specialists, i.e. selectively attuned to some constraints more than others. That is, whereas the default agents are generalists who persist in trying to satisfy all constraints whether satisfiable or unsatisfiable, habituating agents, through the self-organisation of the behaviours on the network, come to specialise in a manner that 'for their own comfort' (i.e. for the immediate increase of their perceived utility) fits together better with one another but thereby actually resolves more of the system constraints in total.

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 maximise global utility can be explained formally using theory from computational neuroscience. Specifically, the behaviour of the network of default agents detailed above is identical to the behaviour of the discrete Hopfield network (9) (which is just a bit-flip hillclimber (15)) and when connections between nodes increase or decrease in strength in a manner that reinforces the current combinations of behaviours this is formally equivalent to Hebbian learning (13). Hebb's rule, in the context of neural network learning, is often represented by the slogan 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 (9) 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 ω_{ij} is $\Delta \omega_{ij} =$ $\delta s_i s_i$ 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 pay-off matrix of each individual by $k_{ii}(t)rU_{ii}$ where $k_{ii}(t)$ is the correlation of behaviours 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 to connection strengths – which they must if these preferences are to mean that this behaviour 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 selforganised 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 networks 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.

From a neural network learning point of view, a network that forms a memory of its own attractors is a peculiar idea (indeed, the converse is more familiar (18)). 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 minimisation).

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, so these infrequently visited attractors will decrease. Eventually, with continued positive feedback, one attractor will out-compete 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 naïve 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 (Fig.1.d). In order to understand whether the competition between attractors in a self-modelling system enlarges attractors with especially high total utility or not, we need to understand the relationship between attractor basin size 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 optimisation 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 (17) for systems that are additively composed of many low-order interactions. Specifically, in systems that are built from the superposition of many symmetric pair-wise interactions, the height (with respect to total utility) of an attractor basin is positively related to its width (the size of the basin of attraction), and the globally optimal attractor state has the largest basin of attraction. 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 sub-optimal 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 out-compete low utility attractors in this self-modelling 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 maximisation

behaviour it will produce a 'model' with ultimately only one attractor, and this attractor will correspond to the globally optimal minimisation of constraints between variables in the original system (13).

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 generalise 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 (18,19), but it in fact represents a simple form of generalisation that is important for our results. Producing new attractor states that are new combinations of features (subpatterns) observed in the training patterns (20) 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 it enlarges the basin of attraction for all configurations in proportion to how many behaviour-pairs they share in common. 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 behaviour combinations in common with locally optimal configurations that have many constraints simultaneously satisfied (but not as many as globally possible).

Lastly on this equivalence, it is essential to recognise how the separation of the timescales for behaviours on the network and behaviours 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 the problem of 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. More generally, if most learning happens at or near random initial conditions then the patterns learnt will be similarly random. It is therefore essential that the system is allowed to relax to local optima, and that most learning therefore happens 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 generalise correctly.

Limitations and further work

Why would agents be creatures of habit? In this paper we have mandated that (otherwise selfish) agents behave as creatures of habit and examined the consequences of this simple local mechanism on global system behaviour. But we

are also interested in the question of whether selfish agents, given the opportunity to alter their preferences according to their own self-interest, would alter them in a Hebbian/habit forming manner. Intuitively, we suggest that this is indeed the case – that forming preferences for the *status quo* is a natural strategy for any agent that favours exploitation over exploration, as any non-teleological agent must.

There is some interesting subtlety involved here however. If an agent's perceptions only alter the *perceived* utility of its actions, and not its *true* utility, then an agent can only assess a proposed change in perception as having some real consequence for its utility if that change in perception causes it to change its behaviours and hence its true utility. Note that when the system is at a locally optimal configuration all changes to behaviours are deleterious, whereas Hebbian changes to preferences never cause a change in behaviour and are therefore neutral. This indicates a preference for Hebbian changes in a somewhat subtle sense. However, when behaviours are discrete (and deterministic) as in the current model, most changes to preferences, either Hebbian or non-Hebbian, will not cause a change in behaviour and will therefore be neutral.

Investigations using alternate behavioural models are therefore being developed elsewhere to address this question. This relates to work we are developing in the context of coevolving species in an ecosystem where species may evolve the coefficients of a Lotka-Volterra system (21,22) or evolve symbiotic relationships (23). This connects the current work with concepts we refer to as 'social niche construction' (24,25,26,27).

Altruism in populations of self-interested individuals has been well researched (e.g. 29); however, very few previous studies investigate games on adaptive networks. Those that do (7,8) differ in a number of ways from the current model, in that here, we: a) only address one type of game (coordination/anti-coordination games), b) play games in normal form, and c) only allow strategies to be adopted to a best-response strategy, rather than by replication equations.

However, despite the novelty of the current model, there appears to be an important similarity between this and many other game theoretic models (network or otherwise) which observe flourishing altruism. Whether they do so by giving agents memory of their past games (30), allowing 'reputation' (31), rewiring links (7,8) or changing link weightings (15), all of these models promote altruism by giving the system a method of passing information from one game to the next, that is not available in the simple, non-altruistic case. This information passing effectively forms a distributed system-level memory, allowing optimisation over multiple games – a mechanism that unites these disparate mechanisms under a common theme.

Finally, it should be noted that the Hopfield model is not new (9), and its capabilities for Hebbian learning are well known (18). However, here we provide a reinterpretation of the system, staging it in a generic, game-theoretic network scenario. This opens up the possibility of reinterpretation of some of the analytically solved variants of the Hopfield model (e.g. 32,33).

Conclusions

This paper 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 behaviours with those they are coordinating with at the present moment, and henceforth adopt behaviours that maximise 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 behaviour that may (at least temporarily) decrease their personal utility but that in the long run increases (the probability of arriving at configurations that maximise) 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 generalisation properties of associative memory, enlarges the basin of attraction of the global optima. This work thereby helps us to understand self-organisation in networks of selfish agents and very simple processes that subtly deviate selfish agents in the direction that maximises global utility without overtly prescribing cooperation or using any form of centralised control.

Acknowledgments: Alex Penn, Simon Powers and Seth Bullock.

References

- Heylighen, F., Gershenson, C., Staab, S., Flake, G.W., Pennock, D.M., Fain, D.C., De Roure, D., Aberer, K., Wei-Min Shen, Dousse, O. and Thiran, P. (2005). Neurons, viscose fluids, freshwater polyp hydra-and self-organizing information systems. Intelligent Systems, 18:4. 72-86. IEEE.
- Nettleton, R.W. and Schloemer, G.R. (1997). Self-organizing channel assignment for wireless systems, Communications Magazine, IEEE 35:8. 46-51. IEEE.
- Smith, J.M. and Szathmáry, E. (1997). The major transitions in evolution, Oxford University Press, USA.
- Dawkins, R. (2006). The selfish gene, Oxford University Press, USA.
- Michod, R.E. (2000). Darwinian dynamics: evolutionary transitions in fitness and individuality, Princeton University Press.
- Newman, M. E. J., Barabasi, A. L. and Watts, D. J. editors (2006). The Structure and Dynamics of Networks. Princeton University Press.
- Pacheco, J., Lenaerts, T. and Santos, F. (2007). Evolution of Cooperation in a Population of Selfish Adaptive Agents. Advances in Artificial Life, 535–544.
- Santos, F.C., Pacheco, J.M. and Lenaerts, T. (2006). Cooperation prevails when individuals adjust their social ties. PLoS Comput Biol, 2(10), e140: 1284–1291
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities, PNAS USA, 79 (8): 2554-2558.

- Crosby, D., Stills, S. and Nash, G. (1971). Love the One You're With. 4 Way Street. Atlantic.
- Hebb, D.O. (1949). The organization of behaviour. Wiley, New York.
- Hinton, G.E. and Sejnowski, T.J. (1983). Analyzing Cooperative Computation. In Proceedings of the 5th Annual Congress of the Cognitive Science Society, Rochester, NY.
- Watson R.A., Buckley, C.L. and Mills, R. (2010). Optimisation in 'Self-modelling' Complex Adaptive Systems, Technical Report, ECS, University of Southampton.
- Nash, J.F. (1950). Equilibrium points in n-person games. Proceedings of the National Academy of Sciences of the United States of America, 36(1): 48–49.
- Watson R.A., Buckley, C.L. and Mills, R. (2009). The Effect of Hebbian Learning on Optimisation in Hopfield Networks, Technical Report, ECS, University of Southampton.
- Wolpert, D.H. and Macready, W.G. (1997). No Free Lunch Theorems for Optimization, IEEE Transactions on Evolutionary Computation 1, 67.
- Kryzhanovsky, B., Kryzhanovsky, V. and Mikaelian A.L. (2007) Binary optimization: A relation between the depth of a local minimum and the probability of its detection. ICINCO-ICSO 2007: 5-10
- Hopfield, J.J., Feinstein, D. and Palmer, R. (1983). 'Unlearning' has a Stabilizing Effect in Collective Memories. Nature, 304:158-159.
- Gascuel, J.D., Moobed, B. and Weinfeld, M. (1994). An Internal Mechanism for Detecting Parasite Attractors in a Hopfield Network, Neural Computation 6:902-915.
- Jang, J.S., Kim, M.W. and Lee, Y. (1992). A Conceptual Interpretation of Spurious Memories in the Hopfield-type Neural Network. Neural Networks, 1992. IJCNN., International Joint Conference on. 1:21-26
- Lewis, M. (2009). An Investigation Into The Evolution Of Relationships Between Species In An Ecosystem. MSc Dissertation, ECS, University of Southampton.
- Poderoso, F.C. and Fontanari, J.F. (2007). Model ecosystem with variable interspecies interactions. J. Phys. A: Math. Theor. 40:8723-8738
- Watson, R.A., Palmius, N., Mills, R., Powers, S.T., and Penn, A.S. (2009). Can Selfish Symbioses Effect Higher-level Selection? European Conference on Artificial Life 2009.
- Odling-Smee, F.J., Laland, K.N. and Feldman, M.W. (2003). Niche Construction. The Neglected Process in Evolution. Monographs in Population Biology. 37. Princeton University Press.
- Penn, A. S. (2006). Ecosystem Selection: Simulation, Experiment and Theory. PhD thesis, University of Sussex.
- Powers, S.T., Mills, R. Penn, A.S. and Watson, R.A. (2009). Social niche construction provides an adaptive explanation for new levels of individuality (ABSTRACT), Proceedings of Workshop on Levels of Selection and Individuality in Evolution, European Conference on Artificial Life 2009.
- Powers, S.T., Penn, A.S. and Watson, R.A. (2007). Individual Selection for Cooperative Group Formation. Proceedings of European Conference on Artificial Life 2007. pp. 585-594.
- 28. Gross, T. & Blasius, B., (2008). Adaptive coevolutionary networks: a review. Journal of The Royal Society Interface, 5(20), 259.
- West, S.A. and Gardner, A., (2010). Altruism, Spite, and Greenbeards. Science, 327(5971), 1341-1344.,
- Axelrod, R., (1987). The evolution of strategies in the iterated prisoner's dilemma. Genetic algorithms and simulated annealing, 32–41
- 31. Nowak, M.A., Page, K.M. and Sigmund, K., (2000). Fairness versus reason in the ultimatum game. Science, 289(5485): 1773.
- 32. Coolen, A.C.C. and Sherrington, D., (1993). Dynamics of fully connected attractor neural networks near saturation. Physical review letters, 71(23): 3886–3889.
- Kryzhanovsky, B., Simkina, D. and Kryzhanovsky, V., (2009). A vector model of associative memory with clipped synapses. Pattern Recognition and Image Analysis, 19(2): 289-295.