Stereotyping, Groups and Cultural Evolution: A Case of "Second Order Emergence"?

David Hales

Department of Computer Science, University of Essex, Colchester, Essex, UK
daphal@essex.ac.uk

Abstract. An on-going project investigating group formation, stereotyping and cultural evolution using an artificial society is outlined. Agents culturally interact by exchanging behavioural rules and cultural markers. They economically interact by playing games of the Prisoners Dilemma. The mode of game play is novel because agents apply stochastic repeated game strategies not to individuals but to subjectively stereotyped groups (based on cultural makers). Agents consequently treat stereotyped groups as single players with whom they are involved in an on-going game of iterated PD. It is envisaged that such cultural processes may display a form of “second order emergence” [11] in which agents come to recognise the cultural groupings that have emerged within the society. Some initial experimental results are presented with tentative observations.

1 Introduction

In complex social worlds, individuals are required to interact with many strangers using limited knowledge and bounded rationality. Yet in human societies the outcome is rarely pure chaos and confusion. An important cognitive tool employed by humans to deal with this situation is the generation, sharing and confirmation of social categories. Social categories can be employed by an individual or group allowing them to partition their society into more-or-less distinguishable groupings. In modern societies there are many such categories in common currency. Some persist and become common knowledge (e.g. “Intellectuals”), others appear quickly then vanish (e.g. “Mods” and “Rockers”) some become official instruments of policy (i.e. governmental classifications of socio-economic class). The methods by which members of a category are identified are many and varied, as are the social behaviours for which such categories provide a rationalisation.

In this paper an Artificial Society is proposed which attempts to capture (in a highly abstracted form) the social construction, communication and dynamics of social categories and their effects on group formation and relations. Agents construct and use categories to inform a choice of behaviour towards others. Categories therefore take the form of “stereotypes” indicating the actions an agent should perform when encountering others. Agents interact culturally via the exchange of memes (see section 4.1) and economically via repeated games of the Prisoner's Dilemma (see section 2). The motivation for this study is to investigate two distinct yet possibly related phenomena: "second-order emergence" where agents come to recognise actual emergent macro-level groupings (see section 7) and sustained altruism and/or exploitation between emergent groupings (see section 6). Both these phenomena are examined within a minimal "memetic" [12] framework.

2 The Prisoner's Dilemma

The Prisoner's Dilemma (PD) game models a common social dilemma in which two players interact by selecting one of two choices: Either to "co-operate" (C) or "defect" (D). From the four possible outcomes of the game payoffs are distributed to the individuals. A reward payoff (R) and a punishment payoff (P) are given for mutual co-operation and mutual defection respectively. However, when individuals select different moves, differential payoffs of temptation (T) and sucker (S) are awarded to the defector and the co-operator respectively. Assuming that neither player can know in advance which move the other will make and wishes the maximise her own payoff, the dilemma is evident in the ranking of payoffs: $T > R > P > S$ and the constraint that $2R > T + S$.

Although both players would prefer T, only one can attain it. No player wants S. No matter what the other player does, by selecting a D move a player ensures he gets either a better or equal payoff to his partner. In this sense a D move can't be bettered since playing D ensures that the defector can not be suckered. This is the so- called "Nash" [10] equilibrium for the single round game. It is also an evolutionary stable strategy [24] for a population of randomly paired individuals playing the game where reproduction fitness is based on payoff. But the dilemma remains, if both individuals selected a co-operative move they would both be better off. But many societies (human and animal) appear to have solved (at least some) dilemmas similar to PD. How can this be explained either from purely rational action or evolutionary mechanisms?

3 Evolutionary Extensions

Evolutionary selection favours selfish individual replicators. When collections of these replicators form groups it's possible for them to co-ordinate their behaviour in ways which would make global optimisation possible. The kinds of behaviours that make this possible include, co-operation, altruism and specialisation. All of these are observed in animal and human societies. But evolutionary selection does not seem to offer an explanation for these behaviours. To address this problem three extensions of natural selection have been proposed: kin selection [14], group selection [31] and reciprocal co-operation [1]. Although each offers explanations of some of the kinds of the social behaviours of interest neither seems to offer a general framework applicable to human or artificial social systems. Kin selection only applies to highly genetically related individuals, group selection in it's simplest form is fundamentally flawed and reciprocal co-operation does not explain true altruism (i.e. co-operative behaviour in the one-shot PD). Neither does it scale-up well to large groups due to the cognitive demands from the requirement that all interactions be on-going with recognisable individuals and associated memory of past interactions.

What kinds of evolutionary process might produce the complex and co-operative social organisations found in human societies? It does not seem possible through natural selection applied to genetic replicators. The formation of highly specialised, co-operative human societies comparatively recently and their rapid evolution indicates a process much faster and more adaptive than genetic evolution. Are evolutionary perspectives of any value at this level of organisation?

4 Culture and Belief

Understanding individual or social "rationality" is not sufficient to explain complex social relations and groupings. Reason has to be applied to assumptions (beliefs) which may or may not be true. Different societies are based on distinctly different belief systems. These systems explain the society to it's members and structure their interactions. Such beliefs may emerge and evolve over long periods of time but sometimes change very rapidly. They are propagated by the combined actions of many individuals and are passed from one generation to the next. Many of these beliefs are not presented as falsifiable representations of reality but as foundational assumptions from which to reason and derive individual behaviours and explanations. To the extent that such beliefs become stable and shared they tend to define a group, we might say they form an essential element of the culture of that group. To understand the formation of complex social systems it is necessary to understand the processes which bring about these shared foundational beliefs.

4.1 Cultural Evolution

Can evolutionary selection be applied to culture? To the extent that the communication of beliefs can be seen as replication and that beliefs can be varied within individuals (through learning and creativity) and between individuals (through imperfect communication) many argue that it can [4], [5], [22], [30]. In order to apply an evolutionary perspective to culture it is necessary to identify a cultural unit, i.e. the cultural analogue of the gene. If ideas are seen as replicating and mutating entities (replicating through agent's brains via communication) then they can be viewed as "viruses of the mind". The analogy is that ideas spread through a population by "infecting" brains in a similar way to the spread of a virus. The word "meme" was introduced by Dawkins [4] as a convenient label for a unit of transmission. Dawkins states:

> The ’gene’ was defined, not in a rigid all-or-none way, but as a unit of convenience, a length of chromosome with just enough copying-fidelity to serve as a viable unit of natural selection. If a single phrase of Beethoven’s ninth symphony is sufficiently distinctive and memorable to be abstracted from the context of the whole symphony... then to that extent it deserves to be called one meme. ([4], p.210)

Many animals have nervous systems capable of supporting the spread of memes (any imitative behaviour). This capability increases the speed and flexibility of the adaptability of the population and paves the way for the development of complex cultural (global) entities to emerge, such as hierarchies, specialisation and group behaviour.

5 Group Selection of Memetic Kin

It is argued that previously proposed extensions to basic natural selection can be applied subtlety within a meme framework to explain seemingly evolutionary unstable behaviours such as altruism (self-sacrifice for the group) and it’s flip-side: exploitation.

Kin selection only applies when individuals share much of the same genetic material. Consider a group with a shared culture comprising a significant number of shared memes. Mathematical models [3] and computer simulations [2], [13] indicate that groups do emerge with relatively stable shared sets of memes. This is a result of the harmonisation of memes through a process of frequency dependant bias [3] and in the context of Axelrod's model "cultural barriers" [2]. When a group share memes, this makes them "memetic kin" [16]. Although they may be different in other respects, memetically they have a common link. Altruistic behaviours among "memetic kin" are selfish from the point of view of the shared memes, if they cause optimisation of individual behaviours which increase the size or stability of the group (sharing those memes). Selection can now operate at the group level if there are other groups with different sets of shared memes.

The argument is that many cultural practices and beliefs, are highly contingent and their evolutionary trajectory is heavily influenced by local chance conditions. If they become harmonised within a group this can result in reduced harmonisation between groups because of cultural barriers (regional languages for example). Even in small scale groups, numerous social psychological studies [27], [21] find that individuals within groups are highly oriented towards their own group both in terms of actively harmonising their beliefs and behaving in a more altruistic way towards in-group members [20] and adapting stereotyped and negative attitudes towards out-group members (so called in-group bias).

Axelrod [2] shows that simple harmonisation of cultural attributes between individuals in a 2-dimensional grid produces separate groups with different "cultures" split along "cultural barriers" of non-communication. An extension of this model into a resource sharing scenario shows that in-group altruism and out-group hostility readily emerge [13].

Nowak and May [25] reduce the number of interaction partners to small overlapping spatially determined groups playing only pure strategies in single round PD. Agents are given the ability to determine (after each round) which agent within their spatial neighbourhood of interaction got the highest pay-off and copy their strategy. Through the application of a strict spatial structure, co-operation could evolve without the need for complex strategies or memories of past interactions. However, spatial models only produce co-operation when learning is limited to some small and fixed group [17].

Macy & Skvoretz [23] show that situating individuals within a large society divided between a small fixed in-group and a much larger out-group of strangers can result in co-operation (in single round games of PD with an exit option) emerging through a form of learning based on a modified GA which captures a form of social learning. Selection operates at the group level because in-groups that do well (by establishing co-operative conventions) become models from which strangers learn and subsequently spread co-operative behaviours.

6 Group Formation

However, in these studies, the group formation processes themselves are not fully explored. The agents have no explicit representations of the groups that comprise their societies. In Axelrod's [2] model group emergence is identified by the observer of the model but is effectively a side-effect of simple agent behaviours: although groups emerge the agents have no explicit cognitive input into this process. In Macy's & Skvoretz' [23] model agents are pre-assigned in-group and out-group membership. Also, these models only address divisions between an in-group and an out-group relative to each agent. There is no mechanism allowing for agents to generate their own internal representations of possibly multiple groupings towards which differential behaviours can be applied. In the context of group formation the Sugarscape [8] models a society in which groups (or "tribes") do emerge and agents do have explicit representations of them. However, the representations in the form of behavioural rules are hard-coded and unchanging. In this sense the representations are exogenous to the model and effectively specify a priori what groups the society can divide into.

Doran & Palmer [6] and Doran [7] present models which furnish agents with explicit representations of group membership. Here agents construct subjective groups based on direct perception or interaction with other uniquely identifiable agents. "Meme-like" harmonisation processes are also implemented. This has the effect of producing "collective beliefs" concerning group membership among collections of agents. However, the groups are defined by explicit representations of individual group members within each agent's memory. Consequently these models do not explore stereotyping processes.

In order to explore these issues, this work outlines an on-going project which attempts to capture some of the dynamics which link group formation and explicit internal agent representations of groups. Minimally agents need to be able to construct categories which partition the population and use these categories to influence interactions with others. The artificial society constructed captures these features via sets of (initially) arbitrary tags (or cultural markers) attached to agents and internal agent behavioural rules (stereotypes) which map tag patterns to strategies for playing the repeated form of the PD. Both tags and rules are treated as memes and as such can be harmonised between individuals during cultural interaction.

7 Second Order Emergence?

Such mechanisms as tags and rules, it is claimed, may capture a minimal form of "second order emergence" (Gilbert 1995). Put simply, the previous models discussed are strictly "bottom-up". Macro structures and behaviours emerge from the micro-level interactions of agents but the agents have no explicit representation of those structures. Consequently any feedback from macro to micro is indirect and happens "behind the backs" of the agents or (in the context of the Sugarscape) was a priori decided by the agent designers. Here an attempt is made to allow agents to evolve macro structures and representations of those structures using a minimal uniform mechanism of cultural imitation (or social learning).

It is not just the actual macro structures that emerge (i.e. groups and actions) that are of interest but the internal representations of those structures (stereotypes) stored within each agent and the relationship between the two.

8 Stereotypes

For the purposes of this study stereotypes are defined narrowly as knowledge which associates attributes with agents based purely on observable characteristics (labels or tags). I exclude assumed, deduced, believed or relative characteristics. The role of tags as methods of increasing co-operation in iterated co-operation games has been discussed by Holland [18] and more recently in [28].

I start from the assumption that stereotypes are constructed maintained and evolved through individual agent interactions over time. Also that different agents may posses different (even conflicting) stereotypes and that the processes that generate them are due to the need for cognitive efficiency and selection of social strategies based on incomplete information. The social psychological literature refers to this as the "information processing error” explanation [27] as opposed to the "kernel of truth" position which proposes that stereotypes are based (at least in part) on true group differences embedded in the structure of society (e.g. that women are innately better homemakers). However, it can be argued that the "structural differences" from which stereotypes may be generated may themselves be the result of processes involving stereotyping (among other cognitive and social processes) and hence are reflexively related.

Labels or tags are defined as observable attributes attached to agents [1], [18], [28]. In a binary string representation of a tag, each bit can be interpreted as representing the presence or absence of some observable characteristic. The definition of labels used by Holland [18] specifies that they are fixed and unchanging intra-generationally but evolve inter-generationally. The interpretation here, therefore is one of physically observable properties linked to genetic material. However, tags have been used to represent cultural attributes which can be copied between agents in order to abstractly capture a form of cultural group formation [2], [8]. The interpretation here, therefore is one of cultural characteristics gained through cultural interactions (e.g. style of dress, social demeanour etc.) which dynamically form identifiable groups.

9 The Artificial Society

Throughout the design of the society important constants have been parameterised. Specified as exogenous parameters they open-up the system to an exploration of a space of behaviours linked to the assumptions (parameter values). This allows for searching and sensitivity analysis in order to link the assumptions to system behaviour.

Agents comprise a set of observable labels (bit strings), a set of behavioural rules (stereotypes) and some state (memory) associated with each rule. The number of bits and rules stored are exogenous parameters. Each bit of the label and each stereotype is treated as a meme. This means they can be communicated and mutated. For each meme held the agent maintains a "confidence value" [0..1] which indicates how attached the agent is to the associated meme. Confidence values are affected by cultural interactions and periodic satisfaction tests (see below).

9.1 Stereotypes and Observable Labels

In order to implement "stereotyping" agents have the ability generalise over observable labels using their behavioural rules. This is achieved by a simple form of pattern matching. Agents store some fixed number of rules which map patterns of observable labels to strategy representations:

```
<label_pattern>  ->  <strategy>
```

The label pattern is a string of the same length as the label bit strings but may comprise digits of zero (0), one (1) and don’t care (#). A don’t care digit matches both zero and one digits. This mechanism allows for generalisation. A label pattern containing all don’t care (#) digits, would match all possible labels.

9.2 The Representation of Strategies

Strategies are represented as pairs (p,q) of real values in the range [0..1] as used in [28], [26]. These values represent the probabilities that an agent will co-operate on a move proceeded by co-operation from the last player to whom the rule was applied in a previous game (p), or co-operate on a move proceeded by defection (q). This is a stochastic representation with a memory of one. It captures many variations of reciprocity and provocability: (1,0) represents tit-for-tat-like reciprocity, (0,0) represents pure defection and (1,1) represents pure co-operation. Each stereotype rule has an associated memory storing either C or D which indicates the move made by the other agent when the rule was last used. Initially these memories are set randomly. Consequently, though agents actually play single round games, these are played by the agents as on-going games of IPD (as if all agents in the category specified by the label pattern were a single agent).

9.3 Cultural and Game Interaction

There are two kinds of independent interaction that occur between agents: game interaction (where a round of PD is played) and cultural interaction (where memes are exchanged). Both spatial and label biasing may be employed to either of these interaction types.

Label biasing consists of rejecting a potential interaction based on label distance. The number of differing bits between the two labels indicates distance. Exogenous bias parameters specify the extent of biasing. They indicate the maximum label distance allowable before an interaction rejection is triggered. Two independent parameters control label biasing for game and cultural interaction.

In order to capture the phenomena of some agents interacting with each other regularly and some less regularly and some not at all, a one dimension spatial relationship has been selected for the population structure. Each end of the line is joined to form a ring topology. Along the line are a finite number of locations (an exogenous parameter), but each location can hold any number of agents. Agents are distributed along the line initially at random. The probability that an agent interacts with another agent is a function of distance representing a gaussian-like distribution around the agents. Both the "flatness" and "width" of the curve are determined by exogenously specified parameters. Both game and cultural interaction are mediated by independent curve parameters.

9.4 Mutation

Agents start with a set of randomly generated memes. Agents can only change their memes by mutation or by accepting a meme from another agent via communication. In order to reduce complexity agents don't generate memes from scratch based on personal interactions. After a satisfaction test (see below) agents examine each of their memes to determine if a mutation should take place. The susceptibility of a rule to mutate is inversely proportional to it's associated "confidence" factor.

Since the LHS of a rule (pattern label) is a bit string (perhaps including don't care symbols), mutation takes the form of changing (with some exogenously defined probability - Mt) each digit from it's current value to one of the other two values with equal probability. When a specific bit value (0 or 1) is replaced by a don't care (#) digit then the rule is generalised. Conversely when a # is replaced by a 0 or 1 the rule is specialised.

On the RHS of the rule (the (p,q) tuple representation), mutation takes the form of changing, with probability Mt, the values of each variable by some +ve or -ve value (in the range -Ms..+Ms inclusive, where Ms is an exogenously defined parameter) Values of >1 or <0 are reset to 1 and 0 respectively. After a rule is changed by either mutation or communication the confidence associated with the rule is set to a random value.

9.5 Confidence Values

Confidence values are changed during communication (see below) and periodically through the application of an all-or-nothing satisfaction test. If an agent is satisfied all confidence values are increased by some factor, else all values are reduced by some factor. Such a scheme implements a crude form of reinforcement learning: if an agent is satisfied it increases the confidence of all memes (by a factor of Ci) otherwise confidence is reduced (by a factor of Cr). Both Ci and Cr are exogenously defined parameters.

Two limitations emerge from this crude method: 1) no attempt is made to promote or demote individual memes based on their contribution to the outcome of the satisfaction test. 2) delayed rewards may not be credited to the memes which generated them since they may have changed in the intervening time. In the context of the game theoretical scenario presented here these limitations are not considered overly restrictive (In more complex scenarios methods of tackling these limitations are discussed in [18] in the context of classifier systems). Since the outcome of each game interaction results in an instant payoff it would not be difficult to accumulate payoffs against the rules that generated them. In this way, confidence values could be differentially updated. However, it is part of the assumption of the society that individuals are highly bounded in their reasoning. Agents don't know which individual memes are responsible for satisfactory outcomes [29].

9.6 Satisfaction Test

An agent is said to be "satisfied" if it's average round payoff is above some threshold T since the last satisfaction test. A satisfaction test is performed with some probability P after each round played. Both T and P are exogenous parameters.

9.7 Cultural Interaction

Each individual rule and label bit is viewed as a meme. The labels bits can be considered to be "surface memes" the rules "hidden memes". Both are communicated in a similar manner.

Two agents are selected to communicate rules and label bits using the interaction method outlined above. Given two agents have been selected, at random, one becomes the sender, the other the receiver. Each meme held by the sender is proposed to the receiver with a probability of Pm (this is an exogenous parameter, 0 indicates no meme propagation, 1 indicates all memes are proposed). The fundamental mechanisms of meme spread are those of:

- replication: the sender replicates a meme to the receiver overwriting an existing meme.
- reinforcement: the receiver already possesses the meme proposed by the sender and this results in an increase in confidence associated with that meme by the receiver
- repelling: the receiver is likely to reject an attempted replication when the associated confidence value of the meme to be overwritten is high.

In order to implement such mechanisms each agent must posses the ability (given a proposed meme by a sending agent) to classify it's memes into one of three types with respect to the proposed meme: a) Identical memes - which can be reinforced; b) Contradictory memes - which need to be removed if the new meme is accepted; c) Other memes - which are neither identical nor contradictory. The label bits are naturally either identical or contradictory (the bits match or they don't). Two behavioural rules (stereotypes) are deemed to be identical if both the pattern and the strategy match exactly and contradictory if the patterns match exactly but the strategies don't. The process of rule communication can be characterised as:-

```
the sender proposes a meme to the receiver
IF the receiver finds an "identical" meme THEN
   the confidence associated with meme is increased
ELSE
   IF the receiver finds a "contradictory" meme THEN
      attempt replication (see 1. below)
   ELSE
      choose a meme at random
      attempt replication (see 1. below)
   ENDIF
ENDIF
```

```
(1) attempt replication:

  the receiver draws a random number between 0 and 1.
  IF draw > confidence associated with it’s meme THEN
```

```
      overwrite receivers meme with senders meme
      (replication)
   ELSE
      sender keeps it's meme intact (repel)
   ENDIF
```

9.8 Game Interaction

When two agents are selected for game interaction (see above) the following occurs:

- each agent reads the other agents label bit string
- using this label each agent searches it's set of rules
- each rule with a LHS label pattern that matches the label is marked as "active"
- each "active" rule is assigned a score based on the number of actual bits (1 or 0) that match (specific rules are therefore scored higher than general rules)
- the rule with the highest score is "fired" and the appropriate action performed as dictated by the strategy represented on the RHS of the rule and the associated memory

If more than one rule has the same highest score then the rule with the highest confidence is used. If more than one rule has the same highest confidence then a random selection is made between them. There will always be at least one "active" rule since each agent is is forced to maintain a default rule (all don't care states on the LHS).

9.9 Consistency & Redundancy

In the scheme described, "contradictory" and "identical" rules are not allowed to coexist within a single agent rule set (see above for functions that define these two terms). Basically, the LHS of each rule must be unique. If a mutation event causes two LHS' to become identical it is reversed. A communication event is so defined (see above) that it can not result in either contradiction or redundancy. Note: This does not mean that more than one rule can not match a single agent. This is resolved via specificity, then confidence, then ultimately a random choice.

9.10 The Time Unit

In a given time unit the following events occur:

- With probability Fg and refusal bias Bg, two agents have a game interaction
- With probability Fc and refusal bias Bc, two agents culturally interact
- With probability Fm one agent moves

A "cycle" of the system is defined as 10N where N is the number of agents in the society (an exogenously defined parameter).

9.11 The Parameter Space Search

Because of the complexity of the system and the number of exogenously defined parameters, an experimental automated searching system has been constructed. Currently the system is simplistic employing multiple hill-climbers over the parameter space. These are executed in parallel across multiple processors working to optimise single pre-specified statistical values derived from the output of the simulation runs.

10 Terminology & Measures

In order to explore the relationship between the agents' internal representations of groups and the actual state of the society the following terminology in introduced.

A group is defined by a rule that partitions the population into group members and non-group members based on observable labels only. An objective group is any set of agents with identical label bits. Since the label bits mediate all interactions between agents, two agents with identical labels will elicit identical responses from any given agent in the population at a given instant in time. A subjective group intention is a string of label bits or some generalisation over the label bits (that is, any combination of "0","1" or "don't care" bits). Each agent holds exactly M rules at all times. Each rule comprises a subjective group intention (or definition) and an associated game strategy. A subjective group extension is defined as those agents that are covered by the intention at a given time for a given population. The most general subjective group intention (comprising all don't care bits) always has an extension comprising the entire population. The size of a subjective group extension is defined as the number of agents in the extension which may be zero. The order of intentional recognition of a subjective group intention is defined as the number of agents in the population at a given time possessing a rule containing the intention. Subjective group intentions can be subsets and supersets of one another. Given an incomplete range of labels in the population at a given time, different subjective group intentions may have identical extensions. An existing subjective group has an extension with a size of more than zero. A recognised subjective group has an order of recognition of more than one. A recognised existing subjective group conforms to both these constraints. For a given existing subjective group, the order of intentional self-recognition of the group is defined as the proportion of members of the group extension that store a rule containing the group intention.

The distinctive homogeneity (HM) of an existing subjective group is defined as the average “relative proportion” of the top M shared rules in the group. The "relative proportion" of a rule is the proportion of agents inside the group holding the rule less the proportion of agents outside the group holding the rule. A value of 1 indicates a rule shared by all members of the group but none outside the group. A value of 0 indicates that the proportions of the rule are equal in and outside the group. With this measure smaller groups will tend to have higher distinctive homogeneity than larger groups. It's a metric of the distinctive cohesiveness of the group with respect to behavioural rules (see equation 1 below).

In order to identify a set of groups with distinctive homogeneity the space of all existing subjective groups (of size >= 10) is searched for the gtop (10) groups with the highest HM values. The HM values for these gtop groups are averaged to give an overall measure of multiple group formation MG (see equation 2 below).

Recognition of the MG groups both inside and outside of those groups is given by averaging the number of rules which match those groups inside and outside of the groups. MGSR (see equation 3 below) indicates the average number of rules matching each group in MG stored by agents within those respective groups. MGIR (see equation 4 below) indicates the average number of rules matching each MG group stored by agents outside those respective groups.

$$HM(g) = \frac{\sum_{r=1}^{topM} \left(Pi(g,r) - Po(g,r) \right)}{M} \tag{1}$$

where: Pi(g,r) = proportion of out (g) group agents holding rule (r)
Po(g,r) = proportion of in (g) group agents holding rule (r)

$$MG = \frac{\sum_{g-1}^{gtop} HM(g)}{gtop} \tag{2}$$

where: gtop = the highest 10 groups based on HM measure

$$MGIR = \frac{\sum_{g=1}^{gtop} IR(g)}{gtop} \tag{3}$$

where: IR(g) = number of rules outside group g that match g

$$MGSR = \frac{\sum_{g=1}^{gtop} SR(g)}{gtop} \tag{4}$$

where: SR(g) = number of rules inside group g that match g

11 Theory of Group and Category Formation

A speculative theory of group and related category formation can now be characterised within the terminology of the Artificial Society presented. From initially randomly placed agents and randomly generated rules and markers cultural interaction produces distinct groupings with shared memes. Cultural markers become predictors (or tell-tale signs) of groups. Rules which identify such groupings give the agent holding them an advantage (consequently stabilising its memes). Meme bundles of markers and rules which recognise those same markers are likely to propagate if they produce positive in-group behaviours. Multiple groups emerge with high self-recognition and recognition of other emerged groupings.

From this theory the following initial hypothesis was derived: In regions of the parameter space where groups with high HM form these groupings will be highly recognised by others (high MGIR) and have a high order of self-recognition (high MGSR).

In order to test the hypothesis the parameter space was first randomly sampled over 100 points giving an average measure of MG, MGIR and MGSR. These values where compared to 100 points which were each found via 100 steps of hill-climbing (maximising MG) from random starting points. Here therefor the hypothesis that high MG is correlated with high MGIR and high MGSR is being tested.

12 Results

Table 1 (below) shows the results of the runs described above. As can be seen the opposite of that predicted by the hypothesis has occurred. When MG was maximised (2) MGIR and MGSR are reduced. After inspection of the runs it was found that some 20% of the runs in (2) resulted in domination of the entire population by a single objective group (all share same marker bits). The MG measure does not exclude this. However, even when these 20% were excluded (3) there was still a negative effect on MGIR and MGSR over (1).

Table 1. Comparison of average MG, MGIR and MGSR values between random sample (1) and maximised MG (2). Each value the result of the average of 100 individual runs at different points in the parameter space.

Description	MG	MGIR	MGSR
(1) random sample	0.22	0.68	0.08
(2) maximise MG	0.53	0.22	0.03
(3) excluding singles	0.48	0.46	0.07

13 Conclusion

These initial results indicate that the simple hypothesis that high group distinctive homogeneity (MG) is correlated with high recognition within and without the group (MGSR and MGIR) is false. High MG is not a necessary and sufficient condition alone to produce such recognition. Of course there may be areas of the parameter space where the hypothesis holds – in which case those areas need to be located and identified. This is part of the on-going project here described.

The method present here for the exploration of artificial societies involves the exogenous parameterisation of each unjustified assumption and the explicit specification of what constitutes behaviour of interest within the society. The parameter space can then be searched automatically to locate those behaviours and link them to the assumptions which give rise to them.

The problems presented by such an approach are firstly, precise specification of "behaviours of interest" (in this paper these are given by MG, MGIR and MGSR) and secondly, how to search a large parameter space practically and ultimately how to identify regions of that space which produce behaviours of interest.

References

1. Axelrod, R.: The Evolution of Cooperation. Basic Books, New York (1980)
2. Axelrod, R.: The Convergence and Stability of Cultures: Local Convergence and Global Polarization. Working Paper 95-03-028. Santa Fe, N.M.: Santa Fe Institute (1995)
3. Boyd, R. & Richerson, P.: Culture and the Evolutionary Process. University of Chicago Press, Chicago (1985)
4. Dawkins, R.: The Selfish Gene. Oxford University Press, New York (1976)
5. Dennett, D.: Darwin's Dangerous Idea. Simon & Schuster, New York (1995)
6. Doran, J. & Palmer, M.: The EOS Project: Integrating Two Models of Palaeolithic Social Change. In: Gilbert, N. and Conte, R. (eds.): Artificial Societies: the Computer Simulation of Social Life. UCL Press, London (1995).
7. Doran, J.: Simulating Collective Misbelief. In: Journal of Artificial Societies and Social Simulation vol. 1 no. 1, http://www.soc.surry.ac.uk/JASSS/1/1/3.html (1998)
8. Epstein, J., & Axtell, R.: Growing Artificial Societies: Social Science from the Bottom Up. MIT Press. London (1996).
9. Gabora, L.: The Origin and Evolution of Culture and Creativity. In: Journal Of Memetics - Evolutionary Models of Information Transmission, 1. http://www.cpm.mmu.ac.uk/jom-emit/vol1/gabora_1.html (1997)
10. Gibbons, R.: A Primer In Game Theory, Harvester, New York (1992)
11. Gilbert, N.: Emergence in social simulation. In: Gilbert, N. and Conte, R. (eds.): Artificial Societies: the Computer Simulation of Social Life, UCL Press, London (1995)
12. Hales, D.: Modelling Meta-Memes. In: Conte,R. & Hegselmann,R. (eds.): Simulating Social Phenomena – LNEMS 456, Springer, Berlin (1997)
13. Hales, D.: Selfish Memes and Selfless Agents: Altruism in the Swap Shop. In: The Proceedings of the ICMAS 1998 (ICMAS'98), IEEE Computer Society, California (1998)

14. Hamilton, W.: Altrusim and Related Phenomena, Mainly in Social Insects. In: Annual Review of Ecology and Systemics 3 (1972) 193-232
15. Hardin, G.: The tradigy of the commons. In: Science. 162 (1968) 1243-1248
16. Heylighen,F. & Campbell,D.: Selection of Organization at the Social Level: Obstacles and Facilitators of Metasystem Transitions. In: The Journal of General Evolution: Special Issue on "The Quantum of Evolution: Toward a Theory of Metasystem Transitions" at ftp://ftp.vub.ac.be/pub/projects/Principia_Cybernetica/WF-issue/Social_MST.txt (1995)
17. Hoffmann, R. & Waring, N.: The Localisation of Interaction and Learning in the Repeated Prisoner's Dilemma. Santa Fe Institute Working Paper 96-08-064. Santa Fe, NM (1996)
18. Holland, J.: The Effect of Lables (Tags) on Social Interactions. Santa Fe Institute Working Paper 93-10-064. Santa Fe, NM (1993)
19. Huberman, B. A, & Lukose, R. M.: Social Dilemmas and Internet Congestion. Science, 277 (1997) 535-537
20. Kramer, R. & Brewer, M.: Effects of Group Identity on Resource Use in a Simulated Commons Dilemma. Journal of Personality and Social Psychology, Vol. 46, No.5, (1984) 1044-1057.
21. Leyens, J. et al.: Stereotypes and Social Cognition. Sage, London (1994)
22. Lynch, A.: Thought Contagion: How Belief Spreads Through Society. Basic Books, New York (1996)
23. Macy, M. & Skvoretz, J.: Trust and Cooperation Between Strangers. Presented at Santa Fe Institute, August 6, Santa Fe, N.M (1996)
24. Maynard Smith, J.: Evolution and the Theory of Games. Cambridge University Press, Cambridge (1982)
25. Nowak, M. & May, R.: Evolutionary Games and Spatial Chaos. Nature, 359, (1992) 532-554
26. Nowak, M. & Sigmund, K.: Tit for Tat in Heterogeneous populations. Nature, 355, (1992) 250-253
27. Oakes, P. et al.: Stereotyping and Social Reality. Blackwell, Oxford (1994)
28. Riolo, R.: The Effects of Tag-Mediated Selection of Partners in Evolving Populations Playing the Iterated Prisoner's Dilemma. Santa Fe Institute Working Paper 97-02-016. Santa Fe, NM (1997)
29. Simon, H. A.: A Mechanism for Social Selection and Successful Altruism. Science, 250 (1990) 1665-1668
30. Sperber, D.: The Modularity of Thought and the Epidemiology of Representations. In: Hirschfeld, L. & Gelman, S. (eds): Mapping the Mind, Domain Specificity in Cognition and Culture. Cambridge University Press. Cambridge (1994)
31. Wilson, E.: Sociobiology: The New Synthesis. Harvard University Press, Cambridge Mass (1975)