Chapter 6

Social Cues and Trust

In this chapter an artificial society model is described (the StereoLab) that attempts to capture in an abstracted form the social construction, communication and dynamics of stereotypes based on "social cues" and their effects on inter-agent trust. Agents interact culturally via the exchange of cultural units (or memes) which represent social cues and stereotypes. Agents economically interact via repeated games of the one-shot Prisoner's Dilemma. The motivation for this study is to investigate the conditions under which sustained co-operation emerges between agents under the influence of stereotyping. Many assumptions of the model are exogenously parameterised. The StereoLab therefore specifies a very large number of model variants over the space of parameters. Techniques by which that space may be explored are described in chapter 7. Actual experimental results are given in chapter 8.

6.1 Social Construction of Social Categories

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 aggregate individuals in 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. "Hippy", "Yippy", "Yuppy"), and 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. It has been argued by some social theorists that *all* categories are socially constructed. Although this view is not shared by the author (this debate is tangential to the work presented here), the focus here is on social categories that *are* socially constructed.

6.1.1 Never Trust a Hippy

In human societies people often have to interact co-operatively with others who they have never met before and therefore have no specific knowledge of. In those situations how does an individual select an appropriate behaviour? Specifically, in an economic transaction where trust is involved, when should a stranger be trusted? When should a stranger be cheated? Consider the following scenario:

> "Imagine you are driving across country for a family vacation when your car overheats. You have the car towed to a service station that has a repair shop. The mechanic says you need an expensive new radiator. It is a hot and humid August day, the kids are cranky, and you are in no mood to pay to have your car towed to another shop for a second opinion. You have no assurance that the mechanic is telling the truth or will charge a fair price or do proper work. What should you do? Meanwhile, the mechanic is equally worried that an out-of-town motorist may skip out on a bad check." [115].

In this scenario both you and the mechanic will benefit if a fair deal can be struck. But how can either party trust the other not to cheat? What knowledge can you both draw on to make a decision? It is argued that one mechanism for coping is to make use of "social cues". Both you and the mechanic assess the situation, observe each other and draw on socially or individually gained knowledge to come to a decision on how to act. If a "similar" mechanic in the past did a poor job and overcharged then you might be tempted to write a bad cheque since "this guy looks like a cowboy mechanic". Conversely, if the mechanic observes that you are wearing a kaftan and have long hair he may conclude you are a "no-good hippy" who is simply not to be trusted. He may overcharge for poor work or worse may refuse to help you. The mechanic may have never met a "no good hippy" in person before. But those he socially interacts with have told him anecdotes of bad deeds. He has been told to watch out for people like this. The point is that individuals may judge others based on personal experience or socially learned beliefs. Also, socially learned beliefs may or may not have some relationship to some real experience, they could simply be myths of uncertain origin and veracity.

6.1.2 Social Cues and Stereotypes

Stereotypes are defined here narrowly as knowledge that associates sets of attributes with sets of individuals based purely on observable characteristics (social cues, cultural markers or tags). It is assumed that stereotypes are constructed maintained and evolved through social interactions between individuals over time. It is also assumed that different individuals may posess different (even conflicting) stereotypes and that the processes that generate them are due to the need for cognitive efficiency and the selection of social strategies based on very limited information. The social psychological literature refers

to this characterisation of stereotyping as the "information processing error" explanation [130]. This is 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. 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[1] related rather than simply reflectively related or false. For example, if some set of individuals hold particular stereotypes (for whatever reason) then this may produce structural regularities which promote and confirm those stereotypes (a self-fulfilling prophecy).

6.1.3 Social Cues and Social Distance

Social cues in the form of dress, accent, physical characteristics etc. may be used by individuals to make comparisons of "social distance" between themselves and others. It is well documented that individuals often prefer to associate with others who are similar to themselves [154]. Social cues therefore may often be used as mechanisms to enforce forms of social exclusion (either economically or culturally) by creating in-groups and out-groups from populations of individuals. Some social cues (or tags) may be easily changed via social influence (e.g. dress or accent) but others are hard to change or disguise (e.g. sex or racial characteristics). So, two kinds of cues may be delineated: fixed traits and culturally learned traits. Either or both of these kinds of cues may be used in processes of social distance estimation. Extreme examples of such practices manifest themselves in communities such as the American Amish [91]. But less extreme forms of social and economic exclusiveness permeate most societies, often involving sets of overlapping, emerging and dissolving group-

[1] By "reflexively" related, I mean that the stereotyping process affects the very groupings which are represented by stereotypes.

ings. Numerous social psychological studies [130], [109] 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 [103] and adapting stereotyped and negative attitudes towards out-group members (so called "in-group bias").

6.2 Salient Features

From the above discussion and example scenario some salient features may be outlined:

- Individuals are boundedly rational social and individual learners.
- Individuals often need to interact with strangers to achieve their goals.
- Many interactions are only mutually satisfactory if trust can be established.
- Individuals may evaluate strangers with reference to observable social cues.
- The social cues an individual possesses may be culturally learned and propagated.
- Some cues may be fixed and unchangeable.
- The way cues relate to selected behaviours may be termed "stereotypes".
- Stereotypes can be socially learned and propagated.
- Individuals often prefer to interact with those holding similar cues.

The StereoLab artificial society attempts to minimally capture these salient features. Interaction involving trust is modelled by agents playing pair-wise games of the one-shot Prisoner's Dilemma (see section 2.1.1 in chapter 2 and section 6.3 below). Social learning and the propagation of cues and stereotypes is modelled in a minimal "memetic"

framework (see section 6.5.2 below). Exclusion practices based on cues are captured by the biasing of game and cultural interactions based on tags. Tags may be fixed or change via cultural interaction.

6.3 Modelling Trust as a Game

The Prisoner's Dilemma (PD) game (as discussed previously in section 2.1.1 of chapter 2) models a common social dilemma in which two players interact by selecting one of two choices: Either to "cooperate" (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 to maximise its 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.

The selection of a co-operative strategy by a player in the PD can be seen as a form of trust. The player exposes itself to exploitation by defection from the other player. Trust in this context represents some action that exposes the player to exploitation by another player when no binding agreement or contract is imposed. Trust, here, is seen as an interpretation placed on the action of an agent not a cognitive state of an agent. The StereoLab models economic interactions using pairwise singe-shot PD game interactions

between agents (players).

6.4 Modelling Social Cues as Tags

Labels or tags are defined as observable attributes attached to agents [5], [88], [143]. 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 tags used by Holland [88] 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. The role of tags as methods of increasing co-operation in Iterated PD games has been discussed by Holland [87], [90] and more recently Riolo [143], [29]. In these latter studies, experimentation with computational models demonstrate empirically that tags can increase co-operation in the iterated PD game.

However, tags have been used to represent cultural attributes which can be copied intra-generationally between agents in order to abstractly capture a form of cultural group formation [7], [52]. The interpretation in these cases is one of cultural characteristics gained through cultural interactions (e.g. style of dress, social demeanour etc.) which dynamically form identifiable cultural groups. Tags in the StereoLab may be either unchanging and fixed (the interpretation being of unchanging physical characteristics) or culturally learnable and mutable (the interpretation being of cultural traits such as style of dress).

6.5 The StereoLab Artificial Society

The aim of the StereoLab design is to capture, in a highly abstracted form, the salient features outlined in section 6.2 above.

Throughout the design of the society important assumptions 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 (see chapter 7 for more on this).

Individuals are represented as simulated agents which display tags represented as binary bit strings (social cues). Agents encounter each other dyadically and play a single round of the PD game which may be thought of as an economic interaction requiring trust. Agents store a set of rules which map tag patterns to PD strategies (stereotypes). Figure 6.1 shows a schematic diagram of a StereoLab agent. Section 6.5.1 below describes the various components of the agent. Cultural interaction between agents also occurs dyadically and involves the propagation of tags and rules (treated as memes). Both game and cultural interaction are described in section 6.5.2 below). Agents inhabit a one dimensional ring comprising a set of independent territories which may contain any number of agents including none (see figure 6.2).

6.5.1 Agents

Agents comprise a set of observable tags (bit strings), a set of behavioural rules and some state (memory) associated with each rule. The number of bits and rules stored are specified by exogenous parameters. Some proportion of the tag bits (specified by an exogenous parameter) and all rules are treated as a memes. This means that they can be communicated and mutated (see sections 6.5.1.3 and 6.5.2.2 below for specific mechanisms). For each meme held the agent maintains a "confidence value" [0..1] which indicates how "psychologically attached" the agent is to the meme. Confidence values are affected by

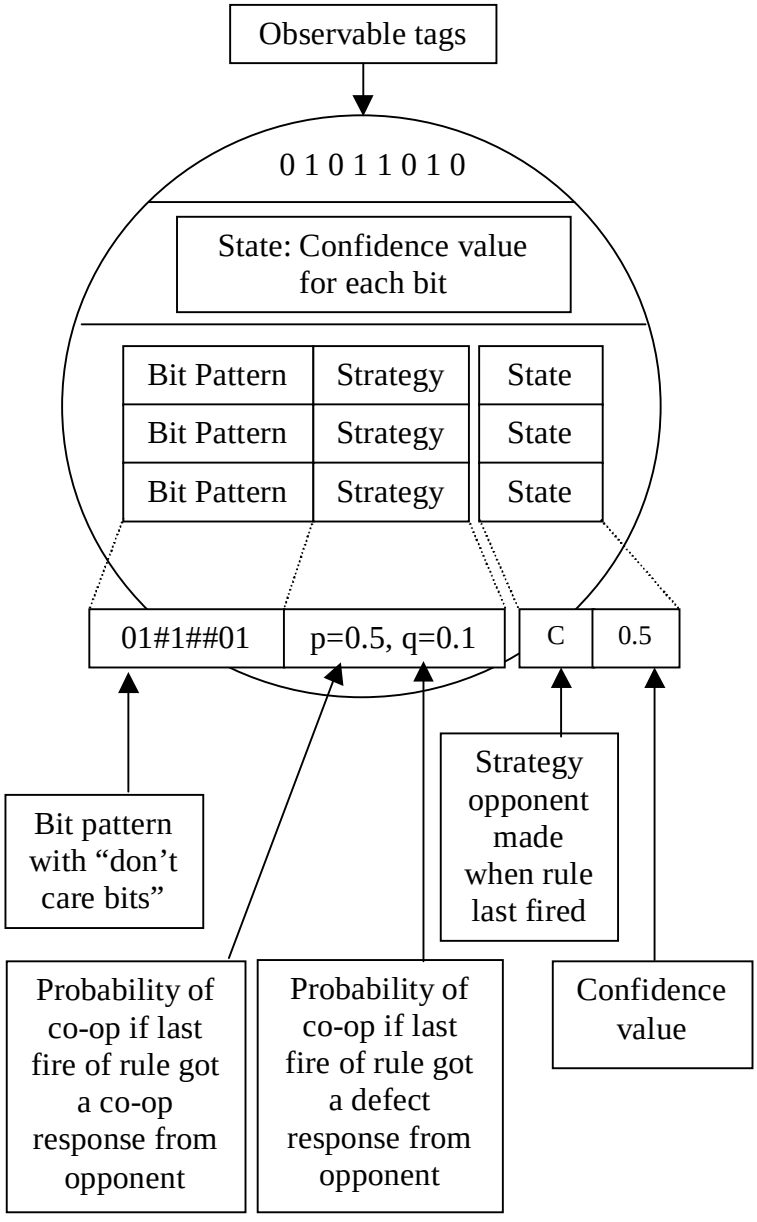

Figure 6.1: An agent in the StereoLab. An agent consists of a set of tag bits (observable by other agents) and a set of rules (stereotypes) mapping bit patterns to game strategies. Each tag and each rule has an associated confidence value.

cultural interactions and periodic satisfaction tests (see below). A proportion of the tag an agent holds may be fixed. The fixed bits never change. The proportion of fixed bits is specified by an exogenously defined parameter (BF). Figure 6.1 shows a schematic diagram of an agent in the StereoLab. In the following sections the components of the agent are described.

6.5.1.1 Tags and Rules

In order to implement "stereotyping", agents have the ability to generalise over observable tags 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 tags to strategy representations:

<tag_pattern> -> <strategy>

The tag pattern is a string of the same length as the tag bit string 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 tag pattern containing all "don't care " (#) digits, would match all possible tags.

Since agents in certain circumstance may mutate the tag pattern this allows for generalisation and specialisation of stereotypes to take place. That is, rules may be widened or narrowed in their applicability. The number of rules an agent can hold is specified by an exogenously defined parameter (M). M is the same for all agents within a given society.

6.5.1.2 Strategies

Strategies are represented as pairs (p,q) of real values in the range [0..1] as used in [143], [126]. The (p) value represents the probability that the agent will co-operate given

that the opponent *co-operated* on the last application of the rule. The (q) value represents the probability that the agent will co-operate give that the opponent *defected* on the last application of the rule. Therefore for each rule an agent has an associated memory storing either C or D which indicates the move made by the opponent when the rule was last used. Initially these memories are set randomly. The (p,q) strategy representation is stochastic 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. Consequently, though agents actually play single round games, these are played by the agents as on-going games of the Iterated Prisoner's Dilemma (IPD) as if all agents in the category specified by the tag pattern in the rule were a single agent.

Given this arrangement it is possible for an agent to play tit-for-tat against a whole group of other agents as specified by the tag pattern associated with the strategy. This captures the notion that an agent may punish an agent within a stereotyped group for something which another agent from that same group did in the past. We should note that intuitively it appears that such a process would make co-operation very hard to achieve.

6.5.1.3 Mutation

Agents start with a set of randomly generated memes (tags and rules). Any fixed tag bits are also randomly initialised. Agents can only change their memes by mutation or by accepting a meme from another agent via communication. After a satisfaction test (see below) agents examine each of their memes to determine if mutation should take place. The susceptibility of a rule to mutate is inversely proportional to the associated "confidence" value. 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 probability MT (where MT is an

exogenously defined parameter) each digit from its 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) strategy 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]. MS is an exogenously defined parameter. Final values of p or q which are >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.

Here the notion of "cultural innovation" is minimally captured. An agent will tend to mutate a rule (stereotype) if its confidence in that rule is low (see below for description of confidence).

6.5.1.4 Satisfaction Tests and Confidence Values

Confidence values are changed during cultural interaction (see below) and periodically through the application of an all-or-nothing satisfaction test. If an agent is satisfied then all of its confidence values are increased by some factor, otherwise all values are reduced by some factor. An agent is said to be "satisfied" if its average payoff from game interactions is above some threshold (T) since the last satisfaction test. An agent performs a satisfaction test with some probability (P) after each game interaction. Both T and P are exogenous parameters. 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 have been discussed in the context of classifier systems [88]. 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 one of the assumptions of the StereoLab society that agents are highly bounded in their reasoning and that they don't know which individual memes are responsible for satisfactory outcomes [148].

6.5.2 Cultural and Game Interaction

There are two kinds of independent interaction that occur between agents: *game interaction* where a round of PD is played (see section 6.5.2.3 below) and *cultural interaction* where memes are exchanged (see section 6.5.2.2 below). The selection of partners for both kinds of interaction may be spatially or tag biased.

6.5.2.1 Tag and Spatial Biasing of Interaction Partners

Both spatial and tag biasing may be employed during the selection of partners for both game and cultural interaction types. Tag biasing consists in rejecting a potential interaction partner based on the number of differing bits between two tags - tag distance. Exogenous bias parameters specify the extent of biasing for both game (BG) and cultural (BC) interaction. They indicate the maximum tag distance allowable before an interaction

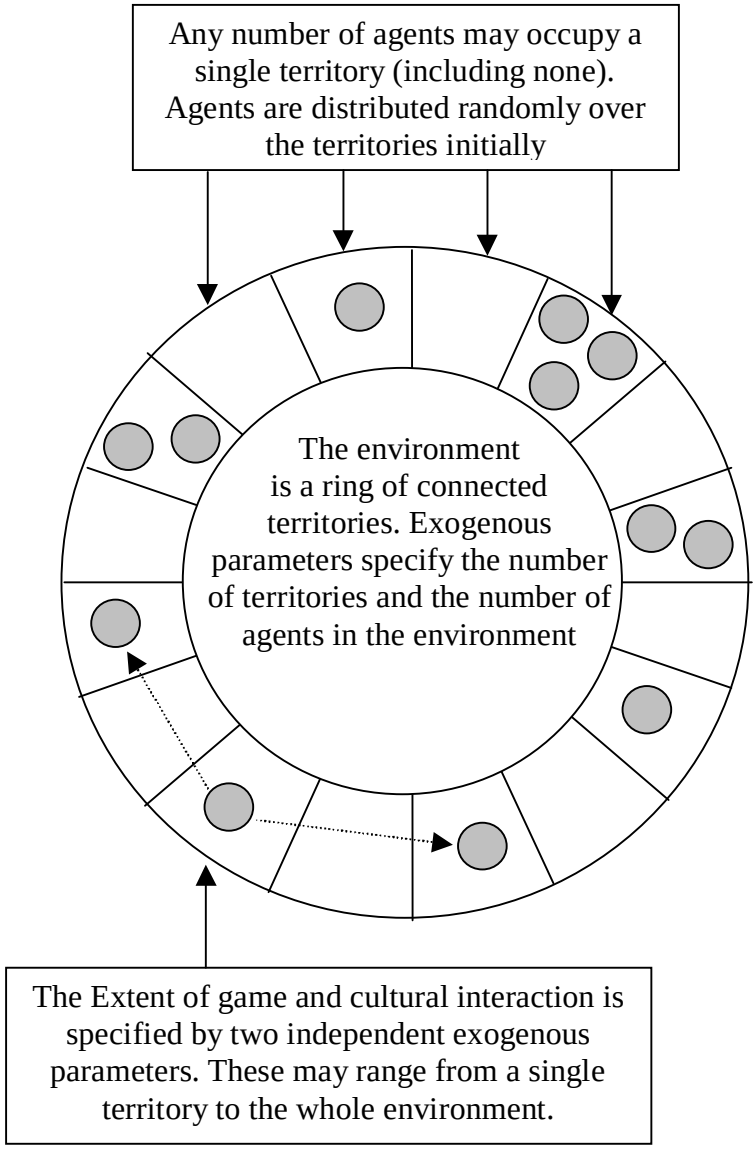

Figure 6.2: The StereoLab interaction environment. Agents inhabit a ring of connected territories. Each territory may contain any number of agents (including none). An agent culturally and game interacts over some proportion of territories specified by exogneously set parameters.

rejection is triggered. The total number of rejections allowed in succession by a single agent before interaction is forced is also specified as exogenously defined parameters (TG, TC).

Agents also limit their interactions to a subset of the population who are spatially close (within their "interaction window"). The justification for this is that cultural and economic interactions are often localised spatially within real societies. Agents inhabit a one dimensional space (see figure 6.2). Each end of the line is joined to form a ring topology. Along the line are a finite number of locations or "territories". The number of territories is specified by an exogenous parameter (S). Each territory can hold any number of agents. Agents are distributed along the line initially at random from a uniform distribution. The interaction window for all agents for a given society is determined by exogenously specified parameters (VG, VC). Both game (VG) and cultural interaction (VC) are mediated by independent "interaction window" size parameters. The largest interaction window specifies that agents in all territories are reachable from any other, the smallest indicates that only agents within the same territory can interact. This spatial arrangement allows for different cultural and game mixing methods to be implemented from pure random mixing (when VG and VC are at a maximum) to highly restricted or niche mixing (when VG and VC are at a minimum). This parameterisation allows for a large set of different localisation types to be explored minimally in one dimension.

Both game and cultural interaction is dyadic. Each kind of interaction is implemented separately: the same pair of agents do not culturally and game interact at the same time. Selection of a pair of agents for either kind of interaction follows the same pattern. Firstly an agent is selected from the population at random, then an interaction partner is selected at random from within the appropriate interaction window (implementing the spatial bias). Then tag bits are compared and the interaction partner is rejected if the

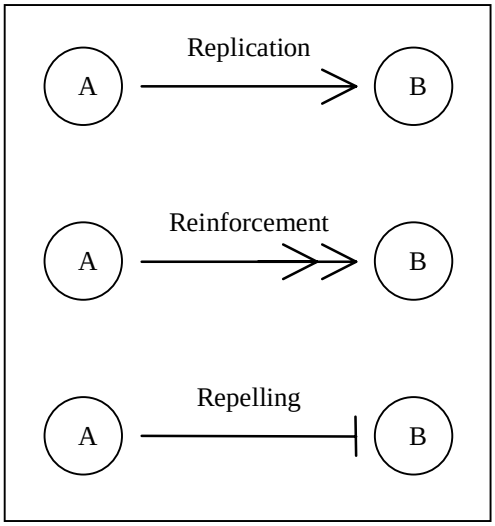

Figure 6.3: Cultural interaction in the StereoLab. Replication: agent A replicates a meme to agent B. Reinforcement: agent A reinforces a meme already held by agent B. Repelling: agent B repels an attempted replication of a meme by agent A.

tag bias constraint is not met. If interaction was rejected another interaction partner is selected. This re-selection is continued until an appropriate interaction partner is found, or until the maximum number of rejections is reached after which interaction is forced with the next randomly chosen partner.

6.5.2.2 Cultural Interaction

Each individual rule and (non-fixed) tag bit is viewed as a meme. The tag bits can be considered as "surface memes" or "social cues" visible to other agents. The rules can be considered as "hidden memes" which are not visible to others (see section 2.6.2 in chapter 2). Both are communicated (i.e. propagated from agent to agent) in the same manner.

Two agents are selected for cultural interaction using the selection method described previously (see section 6.5.2.1 above). Given two agents have been selected, one becomes the sender, the other the receiver (decided by a fair coin toss). Each meme held by the sender is proposed to the receiver with a probability of PM (this is an exogenous

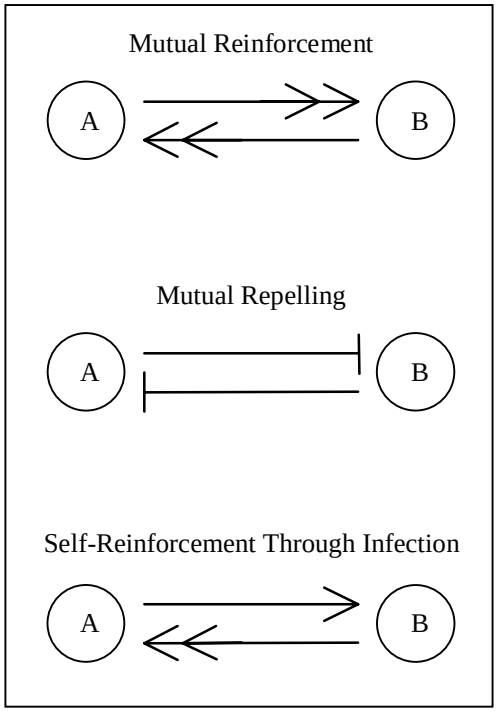

Figure 6.4: Some of the possible interactions that two agents can be involved in over time based on application of the three main meme spread mechanisms (see figure 6.3).

parameter, 0 indicates no meme propagation, 1 indicates all memes are proposed). The fundamental mechanisms of meme spread (see figure 6.3 and 6.4) 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 possess the ability to classify its 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

```
                        Meme propagation:

the sender proposes a meme to the receiver
IF the receiver holds an "identical" meme THEN
        receiver increases associated confidence (reinforcement)
ELSE
        IF the receiver holds a "contradictory" meme THEN
                attempt replication - replacing the contradictory meme
        ELSE
                receiver selects a held meme at random
                attempt replication - replacing the chosen meme
      ENDIF
ENDIF

                        Attempt replication:

The receiver draws a random number [0..1]
IF draw > confidence associated with the meme to replace THEN
        receiver overwrites meme with the senders meme (replication)
ELSE
        receiver keeps its meme intact (repel)
ENDIF
```

Figure 6.5: The steps involved in the propagation of a meme from one agent to another. During cultural interaction between two agents the sender propagates each of its memes with PM probability.

the new meme is accepted; c) Other memes - which are neither identical nor contradictory. The tag bits are naturally either identical or contradictory (the bits match or they do not). 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. In this latter situation the rules are considered contradictory because they would both fire for an identical set of opponents but give different strategies to apply. The process of meme propagation involves the steps shown in figure 6.5.

6.5.2.3 Game Interaction

Game interacting involves the pairing of two agents for a game of the one-shot PD. Two agents are selected for game interaction with relevant tag and spatial biasing mechanisms as previously described (see section 6.5.2.1 above). Each agent decides whether to co-operate or defect in the following way (figure 6.6 summarises the process below):

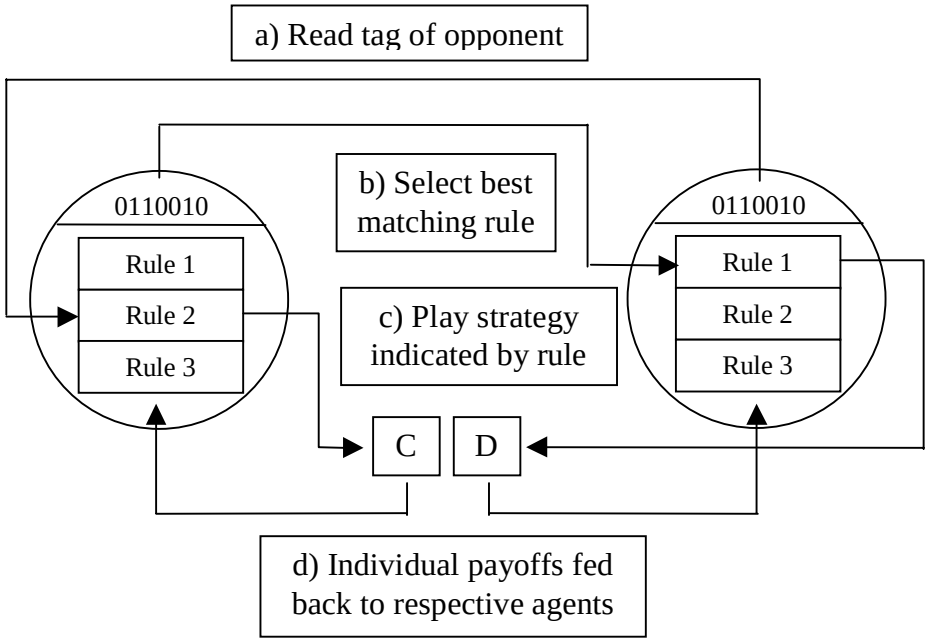

Figure 6.6: Game interaction in the StereoLab. Game interactions take place between selected pairs of agents. The following stages are executed: a) agents read the tag string of the other and use this to selected the rule which best matches the tag. b) the selected rule is then "fired" by playing the strategy indicated. c) the payoffs from the game are fed back to the individual agents.

- Each agent reads the other's tag string.
- Using this tag each agent searches its 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 both the strategy represented on the RHS of the rule and the associated memory.
- If more than one rule has the same highest score (i.e. there is a tie) 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 forced to maintain a default rule - that being, all "don't care" states on the LHS.

6.5.2.4 Rule Consistency & Redundancy

A cultural interaction event is defined such that it cannot result in either contradiction or redundancy within an agent rule set. This does not mean that more than one rule from the rule set of an agent cannot match the tag pattern of a single agent. This is resolved via specificity, then confidence, then ultimately a random choice. As described above, "contradictory" and "identical" rules are not allowed to coexist within a single agent rule set. Basically, the LHS of each rule must be unique. If a mutation event causes two LHS' to become identical it is reversed.

6.5.3 The Time Unit

In a given time unit the following events occur:

- With probability FG two agents game interact.
- With probability FC two agents culturally interact.
- With probability FM one agent moves spatially.

Movement involves a randomly selected agent moving to a randomly selected location. A single cycle of the system is defined as the number of time units required until 10N game interactions have occurred, where N is the number of agents in the society (an exogenously defined parameter). FG, FC, FM and N are exogenously defined parameters. Notice that since FG, FC and FM are probabilities, the events of game interaction, cultural interaction and movement will generally be asynchronous. If all these parameters are set to one there is an ordering of the events as each event must occur. However, in each case the agent(s) which are selected for each event are selected randomly within the tag and spatial biasing constraints described in section 6.5.2.

6.5.4 Summary of the Parameters

A summary of the exogenous parameters used by the StereoLab is given in table 6.1. The range column indicates the range from which values can be selected. Parameters with a single value in the range column indicate that they are fixed at the stated value. The satisfaction threshold T, the probability of a satisfaction test P and the Prisoner's Dilemma payoffs PT, PR, PP and PS are all fixed. The fixing the PD payoffs means game interactions follow the constraints for the PD. However, note that the satisfaction threshold T and the probability of satisfaction testing P are fixed such that a game interaction only produces

satisfaction if an agent receives a temptation (PT) or reward (PR) payoff (see section 6.5.5 below for a justification of this). The reasons for the odd fixed values for the number of agents (N) and the size of the environment (S) are discussed in section 8.1 of chapter 8. Several of the parameters were fixed or limited in range for practical reasons. For example, N and S are fixed since large numbers of agents would significantly increase the time taken to execute a simulation run and very sparse distributions of agents in the environment (which would result from large values of S or small values of N) would limit the application of spatial biasing. Large values for B (the number of tag bits) also significantly increases execution time and small values (below two) would not allow for the proper functioning of the tag processes previously described. The minimum value for FG is set to 0.1 rather than zero since some level of game interaction is required in order to obtain meaningful results.

6.5.5 What Kind of Society Has Been Proposed?

Considering the fixed parameters and the nature of the agents it can be seen that since the agents are satisficers rather than optimisers, and since the satisfaction threshold T = PR (the reward payoff from a PD interaction), the dilemma of the PD is partially resolved. That is, if all agents choose to co-operate then all will be satisfied. The assumption expressed here is that for all StereoLab societies *a state of total satisfaction through complete co-operation is possible.* To put this more anthropomorphically: each agent is happy to sustain a convention of co-operation if all other game interaction partners encountered also co-operate. This assumption intuitively makes co-operation appear more likely. However, this can be contrasted with the previous assumption that agents may retaliate against others that are categorised within the same stereotype as a previous agent that was not co-operative - they make a generalisation. This generalisation means that agents subjectively

stereotyped as members of the same group are treated as if they were a single individual. Taking both of these aspects into account the StereoLab consists of agents who (quite reasonably) are prepared to co-operate if all others do so but (perhaps less reasonably) may retaliate against any stereotyped group member when some member of that group does not co-operate. Such agent behaviour is very reasonable if the stereotyped groups are viewed as single agents rather than some subjectively categorised grouping.

6.6 Summary

In this chapter an artificial society (the StereoLab) has been presented. In the StereoLab agents store, communicate and mutate rules which determine how they should behave when they economically interact with each other (playing pair-wise games of the Prisoner's Dilemma). The rules take the form of "stereotypes" which map observable cultural markers (tags) to game strategies. Agents also have the ability to bias the selection of both game and cultural interaction partners based on "cultural distance" defined as the number of differing tag bits between two agents. Key assumptions of the society (e.g. the number of tags, the extent of spatial and tag biasing) have been parameterised. This parameterisation means that a large space of possible variants of the society are available. In chapter 7 an overview of parameter exploration techniques is given. In chapter 8 the StereoLab parameter space is explored with particular emphasis on regions where high levels of co-operation are found.

Name	Description	Range
B	Number of bits in tag string	4..8
M	Number of stereotypes an agent can store (memory size)	2..10
S	Number of locations in environment	101
N	Number of agents in the society	101
T	Satisfaction threshold	3
PM	Probability of meme propagation	0..1
P	Probability of satisfaction test	1
MT	Mutation rate	0..1
CI	Factor by which to increase confidence	0..1
CR	Factor by which to reduce confidence	0..1
MS	Mutation size for strategy parts	0..1
FG	Probability of game interaction over a time unit	0.1..1
FC	Probability of of cultural interaction over time unit	0..1
FM	Probability of random agent movement over time unit	0..1
BF	Proportion of tag bits that are fixed	0..1
BG	Proportion of tag bits that must be shared to avert game interaction refusal	0..1
BC	Proportion of tag bits that must be shared to avert cultural interaction refusal	0..1
TG	Number of refusals allowed before forced game interaction	1..10
TC	Number of refusals allowed before forced cultural interaction	1..10
VC	Proportion of locations which form cultural interaction window	0..1
VG	Proportion of locations which form game interaction window	0..1
PP	The Punishment payoff from the PD matrix	1
PT	The Temptation payoff from the PD matrix	5
PR	The Reward payoff from the PD matrix	3
PS	The Sucker payoff from the PD matrix	0

Table 6.1: The parameters which charaterise the StereoLab artificial society.