An Adaptive Model of Child-Parent Relationships

Abstract

We develop game theoretic descriptions of attachment interactions between a child and a parent within an iterative, reinforcement learning based framework, whereby the parent's payoff matrix can evolve over time. We show how targeted reinforcement can lead an initially avoidant child in her relationship with the parent into a secure attachment style, and consider how this evolution is affected by both the relative value the parent places on immediate/future rewards, and the set of outcomes that are encouraged. Our method, which to the best of our knowledge is new, also provides a model of psychotherapy by mapping the child/parent relationship into the internal dynamics of the adult self and the inner child. We also consider a 'social game', whereby the individual parent-child dyads are part of a larger society and the parent is influenced by the actions of other parents within their neighbourhood.

1 Introduction

Attachment theory is a branch of psychology in which the bonds between people, and their lasting impact on psychological well-being, are studied. The theory outlines the need of every infant to develop an emotionally supportive, dependant relationship with a primary caregiver, to whom they become 'attached'; and the central tenet is that the type of attachment that emerges, with a life long enduring impact on the individual, is representative of the many interactions the parent and child have with each other during these early, formative years. It grew out of the work of Bowlby [Bowlby and Ainsworth, 1953], and was expanded on by Ainsworth who devised a 'strange situation' laboratory experiment in the 1970s, to investigate the attachment of a child to its mother and the effects of attachment on behaviour [Ainsworth *et al.*, 1978]. In these experiments, a series of separation episodes designed to activate attachment behaviour are carried out on children between 12 and 18 months. The observed quality of interaction with the caregiver upon separation and reunion, along with the degree of exploration of the unfamiliar environment, are then used as the basis for four attachment classifications: a secure category, along with three forms of insecure attachment (avoidant, ambivalent and disorganised).

Work in AI on computational modelling of attachments includes an agent-based simulation of the strange situation, which explains attachment styles as adaptations to care giving styles [Petters, 2005], and an exploration of the potential ability of robots to fulfil attachment related care giving functions [Petters *et al.*, 2010]. Directly relevant here is the work of Buono et al, who have devised a series of stage game models to describe the part of the strange situation at which point the parent returns to the room, in order to show how the behavioural characteristics of 'secure', 'ambivalent' and 'avoidant' might emerge for the child as equilibrium choices with respect to the way in which the parent responds to the child's needs [Buono *et al.*,]. We consider two of these games below.

		Parent	
		Attend	Ignore
Child	Go	4,2	2,3
	Don't Go	3,1	**3,4**

Figure 1: A Type-IIA game with an avoidant Nash equilibrium (child payoff is given first)

		Parent	
		Attend	Ignore
Child	Go	**4,4**	2,2
	Don't Go	3,1	**3,3**

Figure 2: A Type-IIB2a game with both secure and avoidant Nash equilibria (child payoff is given first)

The Type-IIA game (Figure 1) is the ordinal representation of a game for which the child's stress will increase if they go for attention but are ignored by the parent, and for which the parent prefers to ignore the child given that they have chosen to go for attention. We can see that this game is reflective of an avoidant attachment style: the parent's $Ignore$ strategy is dominant and, given this, the child is always better off choosing $Don't\,Go$. The game has a single Nash equilibrium in pure strategies at $(Don't\,Go, Ignore)$.

The Type-IIB2a (Figure 2) game is the ordinal representation of a game for which the child's stress will increase if they go for attention but the parent chooses to ignore them, but for which the parent prefers to attend to the child given that they have chosen to go for attention. This game has two Nash equilibria in pure strategies, at $(Go, Attend)$ and $(Don't\,Go, Ignore)$. The first equilibrium is Pareto optimal and representative of secure attachment, whereas the second equilibrium represents an avoidant relationship.

Only the ordinal ranks, i.e. the relative weight of the payoffs, are considered here. These payoffs can be regarded as the degree of pleasure-displeasure experienced by the child and the parent; their real values for each particular child-parent pair can be determined in future work by techniques in affective computing using for example the PAD (Pleasure-Arousal-Dominance) model [Mehrabian, 1995] [Scherer and Roesch, 2010].

2 Iterated Game

In the context of attachment behaviour, we are interested in how the repeated game can model a change from one attachment style to another. Evidence suggests that radical shifts in the way a parent interacts with their child can result in a change in attachment style [Waters *et al.*, 2000], and in this section we will focus on modelling the particular transitional case from insecure to secure attachment.

In games where there are multiple Nash equilibria in pure strategies, it may be the case that the attachment style can move from an insecure to secure relationship simply by moving from one equilibrium to another. For example, in the game in Figure 2 moving from the equilibrium at $(Don't\,Go, Ignore)$ to $(Go, Attend)$ signifies a shift in attachment style from avoidant to secure. Once this shift has somehow been achieved neither the child nor parent would have an incentive to deviate unilaterally nor jointly, since this is a Pareto optimal Nash equilibrium. A realisation by the parent that this more efficient equilibrium exists, and a willingness to direct the interaction in this direction may thus be all that is required.

Iterated games in which the payoffs change have previously been considered in Stochastic game theory [Shapley, 1953]. Stochastic games are based on Markov Decision Processes [Howard, 1960] whereby the game can change between each round according to some probabilistic function over the action choices. The probabilistic state transition function has the Markov property, i.e. given the current state and action, the next state is independent of all previous states and actions. In our model, we use for our state representation the ordinal equivalence of the absolute, externally-reinforceable payoff matrix belonging to the parent. We show how, when combined with reinforcement learning, this model can facilitate the emergence of secure attachment in games for which there is initially just an avoidant Nash equilibrium. To the best of our knowledge, such evolving games in which the payoffs are externally reinforced have not been studied.

2.1 The Child's Game

We assume that the child is unaware of any change to the structure of the game being played and that they therefore continue to play according to the initial game payoff matrix. It is also assumed that the child will be playing a strategy with a reactive characteristic - i.e. that they pick their move in each round based on their analysis of the pattern of play of the parent, rather than, for example, deliberately trying to guide the interaction pattern into a new equilibrium.

Evidence for both statistical inference [Kushnir *et al.*, 2010] and event memory [Bauer, 2006] in strange situation aged children suggests that such a suitable reactive strategy may be one in which the child maximises their expected payoff based on a probability distribution over the last n moves that the parent chose. In this paper we concentrate on the simple case of this strategy for which n = 1, i.e. a 'best response to last move' (BRTLM) strategy, whereby the child picks their move corresponding to their highest payoff based on the assumption that the parent will play the same move that they played in the previous round.

2.2 The Parent's Game

Since the child's payoffs remain unaltered, the parent's payoff matrix must change in order for new Nash equilibria corresponding to secure attachment to emerge. This means that the parent must change the value they place on individual game outcomes, so that outcomes resulting in the parent choosing $Attend$ and the child choosing Go gradually come to be preferred over avoidant interaction outcomes. The parent therefore plays an evolving iterated game where their initial payoff matrix becomes their *absolute* payoff matrix, and subsequently changes according to a payoff reinforcement rule. The goal is to model a learning process for the parent, whereby the reinforcement of certain desirable action combinations gradually leads the parent (and thus the child) into a stable pattern of play resembling a secure attachment relationship.

The parent's side of the game is a state transition system consisting of a finite set of states S. Each state is represented by an ordinal payoff matrix $\begin{bmatrix} t & u \\ v & w \end{bmatrix}$, corresponding to the ordinal equivalence of the parent's absolute payoff matrix. The state therefore represents the relative value that the parent places over the corresponding stage game outcomes. There are 4! states for the permutations of all strict ordinal payoff matrices (i.e. for the case where $t \neq u \neq v \neq w$), plus 4!/2! states for the non-strict ordinal payoff matrices with two equal elements ($t \neq u \neq v = w$), plus 4!/3! states for the non-strict ordinal payoff matrices with three equal elements ($t \neq u = v = w$), plus 4!/4! states for the case where all the payoffs are equal ($t = u = v = w$), giving a state space size of 41.

At each discrete step in time (representing a moment in which a decision is to be made) the system is in some particular state $s \in S$, where each state s has a corresponding finite set of valid actions A for the parent to choose between. For simplicity we assume that each state has the same action set $A_p = \{Attend, Ignore\}$. The parent chooses an action $a_p \in A_p$, and the child chooses an action $a_c \in A_c$, where $A_c = \{Go, Don't\,Go\}$, resulting in an action-combination outcome for each stage game (a_c, a_p). It is this action-combination outcome which determines whether or not a reinforcement occurs on the parent's absolute payoff matrix, and it is these reinforcements which in turn are the driver of state transition.

With each reinforcement, the absolute value of the corresponding payoff for the parent will grow, and thus their payoff matrix will change over repeated interactions. It is the ordinal structure of the evolving payoff matrix that we are concerned with, since this can be seen to represent the parent's emotional state and thus forms our state representation. As a consequence of this ordinal state representation, we can see that an action-combination outcome that results in payoff reinforcement does not necessarily also result in a state transition $s \overset{a_p}{\to} s'$, since such a transition only occurs if the resulting *ordinal* payoff matrix differs to the current state.

The problem of the parent choosing the optimal action within each state of the evolving game is suited to a reinforcement learning framework [Sutton and Barto, 1998]. We apply the well known 'Q-Learning' reinforcement learning algorithm over the state transition system as a model of how the choices that the parent makes will adapt over time [Watkins, 1989]. The Q function $Q : S \times A \to \mathbb{R}$ calculates a Q value for each action associated with a particular state. These Q values are an estimate of the expected reward that the parent will receive from choosing this particular action when they come to make an action decision. Following the choice of action a_p in state s and an observed transition to state s', we update the Q values according to the conventional update rule:

$$Q(s, a_p) \leftarrow Q(s, a_p) + \ell[R_{a_p}(s, s') + \gamma]$$

where

$$\gamma = \delta \max_{a'_p} Q(s', a'_p) - Q(s, a_p)$$

This update rule represents the learning process by which the parent re-evaluates the quality of each state-action combination according to the rewards received from the action choices they make. The discount factor $0 \le \delta < 1$ signifies the relative value the parent places on immediate/future rewards, and the learning rate $0 < \ell \le 1$ determines the extent to which newly acquired information overrides old information. We associate a learning rate with each (s, a_p) pair and set it according to $\ell(s, a_p) = \dfrac{1}{n(s, a_p)}$, where $n(s, a_p)$ equals the number of times action a_p has been chosen in state s, so that initially $\ell(s, a_p) = 1$ and decreases with each subsequent selection of action a_p in state s.

Action Choice

In order to avoid the parent's Q values getting stuck in a local maximum, actions associated with non-optimal Q values should occasionally be played and as much of the state space as possible be explored. At the same time the parent should want to focus their exploration primarily on the most promising states. To resolve this exploration and exploitation tension we employ a probabilistic 'Boltzmann' action selection rule [Kaelbling *et al.*, 1996]: given that the parent is in state s and has exploration parameter k, the probability that they select action a_{p_i} is:

$$P(a_{p_i}|s) = \frac{k^{Q(s,a_{p_i})}}{\sum_i k^{Q(s,a_{p_i})}}$$

As k increases, the probability of selecting those actions with low Q values becomes smaller, and so one way of looking at the exploration parameter is to say that larger values of k are representative of a parent who has more 'embedded' behaviour.

Payoff Reinforcement Rule

Only desirable or helpful outcomes to each stage game should be praised and encouraged. We label this set of action-combinations, that will trigger reinforcements in the parent's absolute payoff matrix, η. In any single stage game the action-combination $(Go, Attend)$ is desirable, since we are hoping to guide the iterated game such that every game ultimately has this outcome. However, this is not necessarily the only action-combination that should be reinforced: we intuitively observe that the action-combination $(Don't\,Go, Attend)$ is also a helpful outcome, since the parent attending could encourage the child to Go in a following round. Conversely, the outcomes $(Go, Ignore)$ and $(Don't\,Go, Ignore)$ are neither desirable nor helpful outcomes to any stage game since they will encourage and reinforce 'Ignore' behaviours in the parent, and are never praised.

We define $r > 1$ to be the reinforcement parameter. In each round of the game the parent chooses their action according to the action selection rule outlined in the previous section, and the child picks theirs according to their game theoretic strategy (BRTLM), resulting in an action combination (a_c, a_p). The reward component of the Q value update rule, $R_{a_p}(s, s')$, corresponds to the payoff the parent receives from the action-combination (a_c, a_p). If $(a_c, a_p) \in \eta$ then this reward is a reinforcement of the corresponding payoff in their absolute payoff matrix, and the parent's ordinal payoff matrix may transition into a new state as a result of this reinforcement. If $(a_c, a_p) \notin \eta$ then the reward for the parent is a non-reinforced payoff and no state transition will occur, i.e. $s' = s$.

Reinforcements on the parent's absolute payoff matrix are multiplicative, since each time a desirable action-combination $(a_c, a_p) \in \eta$ is observed following a round in the iterated game, the corresponding payoff element in the parent's absolute payoff matrix is reinforced by this factor r. Thus, if we had experienced N rounds of play, where the action combination $(Go, Attend)$ had occurred j times, $(Don't\,Go, Attend)$ had occurred m times, and $(Go, Ignore)$ and $(Don't\,Go, Ignore)$ occurred a total of $N - (j + m) \ge 0$ times, and our reinforcement rule is $\eta = \{(Go, Attend), (Don't\,Go, Attend)\}$, then the parent's absolute payoff matrix will have been reinforced to that in Figure 3 (such that the ordinal equivalence of this reinforced payoff matrix is the parent's current state in the state transition system). The payoff element u has not been reinforced since the action-combination $(Go, Ignore)$ has not occurred, and is not a reinforceable outcome anyway since $(Go, Ignore) \notin \eta$. Likewise, the element w has not been reinforced since $(Don't\,Go, Ignore) \notin \eta$.

An alternative state representation would be based on the parent's absolute payoff matrix as in Figure 3. In this case, the system would be an infinite transition system with states given by $(j, m) \in \mathbb{N}^2$ with deterministic rules of transition (i.e. a stochastic game with transition probabilities of either 0 or 1). In such a system, with $r > 0$ a constant, a state can never be re-visited. In contrast, when we use an ordinal payoff state representation, states can be re-visited and the parent can make use of their learning in any other round in which the absolute payoffs were ordinally equivalent.

$$\begin{bmatrix} tr^j & u \\ vr^m & w \end{bmatrix}$$

Figure 3: The reinforced absolute payoff matrix after $(Go, Attend)$ occurred j times and $(Don't\,Go, Attend)$ occurred m times, where t, u, v and w are the initial payoffs prior to reinforcement and r is the reinforcement parameter. The ordinal equivalence of this absolute payoff matrix forms the parent's current state.

The condition for convergence to a stable pattern of play is the emergence of a Nash equilibrium at $(Go, Attend)$. This is true for both the case where $\eta = \{(Go, Attend), (Don't\,Go, Attend)\}$ and $\eta = \{(Go, Attend)\}$, as the child prefers to choose Go over $Don't\,Go$ given that the parent has chosen $Attend$. Since we are using a probabilistic 'Boltzmann' action selection rule whereby the probability of the parent choosing $Attend$ will always be greater than a minimum positive number, the game will converge, in measure-theoretic terminology, almost surely (i.e., with probability one) to this Nash equilibrium over an infinite number of rounds. What we are really interested in, however, is not convergence per se, but the rate of convergence with regards to the learning parameters, since for a technique to be useful in psychotherapy its effects must be seen within a reasonably small time frame.

2.3 Simulation Results

Simulations of iterated games were run such that each iterated game consisted of 10,000 rounds, the initial game was the Type IIA game (Figure 1). Since the game begins with an avoidant attachment style, the Q values are initialised such that the parent's expectation of rewards associated with the action in each state corresponds to an expectation that the child will play $Don't\,Go$, and without a consideration for any potential reinforcements or state transitions. For example, for the state $\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$ the initial Q value for the action $Attend$ is 1 and for 'Ignore' it is 4. Similarly, for the state $\begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$ the initial Q value for $Attend$ is 1 and for 'Ignore' it is 2.

Each individual simulation was repeated independently 200 times, and the average number of rounds before the interaction converged towards a stable, secure relationship (i.e. where both the parent chose $Attend$ and the child chose Go without any deviation) was calculated. This process was repeated for various combinations of the parameters for the discount factor δ, for reinforcement rates $r = 1.01$, $r = 1.02$ and $r = 1.03$, and for the reinforcement rules $\eta = \{(Go, Attend), (Don't\,Go, Attend)\}$ and $\eta = \{(Go, Attend)\}$. In all games we set the parent's exploration parameter to $k = 2$. Below we chart our simulation results, where a solid line indicates that the reinforcement rule $\eta = \{(Go, Attend)\}$ was used, and a dashed line that the reinforcement rule $\eta = \{(Go, Attend), (Don't\,Go, Attend)\}$ was used. The resulting stable secure game that evolved was of Type IIB2a (Figure 2).

As the size of the reinforcement r on the parent's payoff matrix is increased, the average number of rounds required before a stable, secure attachment style emerges decreases. We also observe that, in general, parents with lower discount factors (i.e. parents who prefer immediate rewards) see the emergence of a secure attachment style more quickly than those who favour future rewards. As would be intuitively expected under our model, the results also show us that secure attachment relationships are more quick to emerge when $\eta = \{(Go, Attend), (Don't\,Go, Attend)\}$ (i.e. when any outcome resulting from the parent selecting the 'Attend' action is encouraged, represented by a dashed line) than when $\eta = \{(Go, Attend)\}$ (i.e. when only the $(Go, Attend)$ outcome is encouraged, represented by a solid line). However, we note that this effect becomes less pronounced as the size of the reinforcement r increases.

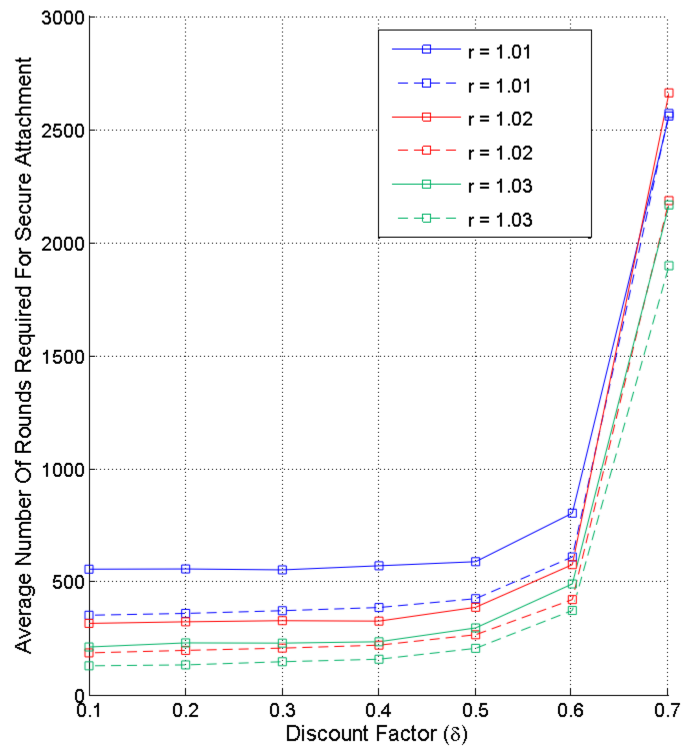

Figure 4: The average number of rounds required for a stable, secure attachment style to evolve when the child played BRTLM, for various discount factors (δ) and reinforcement rates (r). Solid lines are results for the reinforcement rule $\eta = \{(Go, Attend)\}$ and dashed lines for $\eta = \{(Go, Attend), (Don't\,Go, Attend)\}$. We have postulated that a lower k would be required in order to reduce convergence time beyond $\delta = 0.6$

An unexpected observation was that parents who had a large discount factor $\delta > 0.7$ (i.e. parents who placed more of a preference on future rewards) did not on average converge to a stable, secure attachment style within 10,000 rounds of play, and we will attempt to explain why here. If the action-combination outcome from some arbitrary round $(a_c, a_p) \in \eta$ then the reward issued is based on a reinforcement of the parent's underlying payoff matrix. The parent will receive a reward even when no state transition occurs, i.e. $R_a(s, s') > 0$ for $s = s'$. Since state transitions only occur when these reinforcements result in a new ordinal payoff matrix for the parent, we can have the situation whereby Q values for $Q(s, a)$ are updated even if no state transition has occurred. A parent with a high discount factor $\delta > 0.6$ who initially selects actions in state s that do not result in any state transitions will learn relatively larger Q value updates for these non-reinforceable actions than would a parent with a lower discount factor. In addition, since the learning rate l decreases with each action choice a in state s, early action choices which do not result in state transitions are even more pronounced. It appears as though there is a threshold for δ for which these initial non-transitional Q value updates are so large that the probability of exploration becomes too small to result in a change in attachment style. Therefore, to see a quicker convergence toward secure attachment in parents with a discount factor of $\delta > 0.7$, we postulate that a lower value of k would be required, resulting in more state-action exploration. This is interesting when considered within the context of psychotherapy and the exploration of new states, since it suggests that the more relative value a parent places on future rewards, the more encouragement they will require initially with regards to breaking embedded patterns of behaviour in order for a stable, secure form of attachment to be achieved.

3 An Adaptive Model for Psychotherapy

We map our model into the dynamics of the inner world of a person undergoing psychotherapy in order to obtain an iterative internal game within the person's internal psyche. The inner world of the person is regarded as the interaction of the 'Inner Child', which replaces the child, and the 'Adult Self', which replaces the parent. The Inner Child represents the non-optimal or damaged emotional part of the person, whereas the 'Adult Self' represents her/his thinking and logical part. The notions of inner child and adult self have their roots partly in Eric Berne's transactional analysis [Berne, 2010], a popularisation of Freud's psychoanalysis, and have been introduced in psychotherapy by various clinicians in the past decades (e.g. [Whitfield, 1991] and [Farmer, 1990]). We consider the Inner Child and the Adult Self as two independent agents each with its own "rational" behaviour.

The parent strategy of 'Attending' is mapped to 'Mentalizing', which is the capacity to understand the mental state (including the emotional condition) of oneself and others. Mentalizing has been introduced in Attachment Theory and psycho-dynamic therapy by Peter Fonagy and his colleagues in the past decade and is considered to be an overarching foundational technique in different types of psychotherapy, including Cognitive Behaviour Therapy (CBT) [Allen *et al.*, 2008]. Finally, the child's strategy of 'Go' to the parent is mapped to 'Seek Support' from the Adult Self, which, in the case that the Adult Self chooses Mentalizing, leads to an increased ability by the inner child to regulate its emotions, an ability that is considered as the key goal in psychotherapy [Schore, 2012]. The psychotherapy game then starts with the Adult Self/Inner Child dyad in an insecure (e.g. avoidant) attachment type and, using reinforcement from a psychotherapist or a self-help therapeutic technique, gradually converges to inner secure attachment. This iterative game thus provides a game-theoretic model of psychotherapy and can be considered as 're-parenting' the inner child, a concept used in particular in Schema Therapy [Young *et al.*, 2006], which is an extension of CBT that is heavily influenced by Attachment Theory.

4 Social Game

In the previous section we considered an iterated game in which a single, external force reinforces the parent's payoff corresponding to the observed outcome following each round. The payoff reinforcements are governed by some predefined rule, such that only specific round outcomes are encouraged.

In a realistic situation, parents will have relationships with other parents such as friends and family members. Belsky has done extensive research into the relative effects of these 'social networks', and has concluded that the way a parent behaves towards their child can be influenced by factors such as "praise for their skill in handling children", the provision of information and advice and the natural emergence of social expectations [Belsky, 1984]. Indeed, Green et al have highlighted the theory that attachment style is malleable with regards to environmental circumstances and feedback from these social networks [Green *et al.*, 2007]. As such, we now consider 'social games' represented as mathematical graphs, where each node represents an iterated parent-child attachment game, such that the social game is actually a collection of individual iterated games that are interconnected.

Framework

Whereas before the static reinforcement rule governed which round outcomes were to be reinforced, we now model the case whereby society (i.e. the neighbourhood of games) determines which parental actions are desirable. Within this context, reinforcement on the parent's payoff matrix represents a sort of social validation of the action choices made by them in each round of the iterated game they play with their child.

We define the social game to be an undirected graph $G = (V, E)$ where V is a set of vertices (nodes) and E is a set of edges. Each node $v \in V$ represents an individual parent-child iterated game, therefore the total number of individual games is equal to the order of the graph $|V|$, and we label the nodes from 1 to $|V|$. Each game is initialised to either a secure game of Type IIB2a (Figure 2) and represented as a green node, or an avoidant game of Type IIA (Figure 1) represented by a red node. For the purposes of the analysis below we only consider connected graphs such that each node has at least one connection to another node.

As before, the child picks their action in each round according to their game theoretic strategy (BRTLM), and their payoffs do not change. For simplicity we assume that the parent simply picks their action according to the Nash equilibrium, without defining any explicit underlying learning process. In the case of the single iterated game, reinforcements on the parent's payoff matrix occurred after each round according to some predefined reinforcement rule which ensured that only the payoffs for certain 'desirable' outcomes were reinforced. For the social game we relax this constraint and allow reinforcement on every element in the parent's payoff matrix.

Any node connected to some particular node v is a neighbour of v, and we label the set of these direct neighbours of v as the set of nodes C_v, so that v has $|C_v| \geq 1$ neighbours. We define the *neighbourhood* for v as $N_v = C_v \cup \{v\}$, i.e. the set containing v and its direct neighbours, such that $|N_v| = |C_v| + 1$.

Now we define the *individual reinforcement parameter* r_v for node v, which is set according to the following rule:

$$r_v = \frac{2}{|N_v|}$$

Since we have stated the condition $|C_v| \geq 1$ (i.e. each node v must have at least 1 direct neighbour) we can deduce that $0 < r_v \leq 1$. We use this reinforcement parameter r_v to define the *reinforcements* at node v as follows. For each round of the iterated game at this arbitrary node v, the child chooses an action $a_c \in A_c = \{Go, Don't\,Go\}$ and the parent chooses an action $a_p \in A_p = \{Attend, Ignore\}$, resulting in an action-combination (a_c, a_p). Following the round, the element corresponding to this action-combination (a_c, a_p) in the parent's payoff matrix is *reinforced* by $r_v\, m_v$ where $0 < m_v \leq |N_v|$. Here, m_v is the number of nodes in N_v whose parents also played the action a_p (i.e. the same action that the parent at node v played). For example, if in the round just gone only the parent at node v played action a_p then $m_v = 1$, whereas if one of v's neighbours also played a_p then $m_v = 2$.

At this arbitrary node v, reinforcements on the parent's payoff matrix are therefore determined by the actions of neighbouring parents rather than the outcomes of neighbouring games. For the case where $m_v > \frac{|N_v|}{2}$ the reinforcement $r_v\, m_v > 1$ and the corresponding payoff will increase. When $m_v < \frac{|N_v|}{2}$ the reinforcement $0 < r_v\, m_v < 1$ and the corresponding payoff decreases. Finally, if $m_v = \frac{|N_v|}{2}$ then $r_v\, m_v = 1$ and the corresponding payoff remains unchanged.

Network Malleability

As a sufficient condition for reinforcement we require that the majority action choice chosen by the parents of nodes in N_v be different to the action chosen by the parent at node v. If this is not the case then $r_v\, m_v = 1$ and the payoff does not change. If $r_v\, m_v \neq 1$ then payoff reinforcement will occur, with two possibilities. Either this reinforcement (or a sequence of these reinforcements) will result in new Nash equilibria, and thus new attachment styles for v, or it will not.

We refer to network configurations where either the payoffs are not reinforced, or they are reinforced such that the existing Nash equilibria cannot be broken, as 'unmalleable' networks. In unmalleable networks, the attachment styles of the nodes will not change. Networks in which both reinforcements occur, and where a single (or successive) reinforcement(s) can result in the emergence of new attachment styles for one (or more) of the nodes, are referred to as 'malleable' networks. The attachment styles of one or more of the nodes within malleable networks can change. It is these 'malleable' configurations that we are interested in. Figure 5 shows all such malleable networks of 3 nodes, and Figure 6 shows all malleable networks of 4 nodes, determined as the result of simulations run over 1000 rounds.

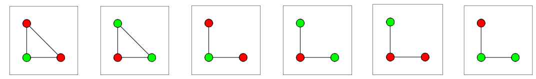

Figure 5: Malleable Networks of 3 nodes

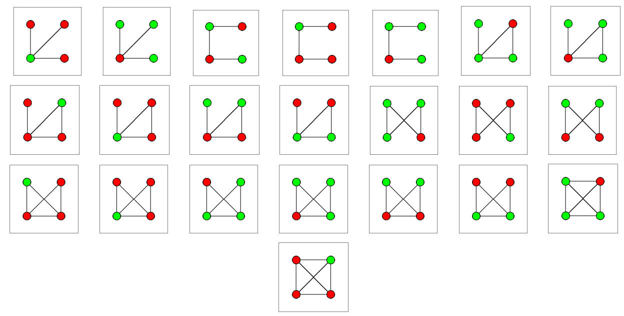

Figure 6: Malleable Networks of 4 nodes

5 Summary

In the single parent-child dyadic model, we assumed that there was some external force such as a psychotherapist encouraging and praising only certain outcomes to rounds of the iterated game. Based on a number of assumptions on the part of the parent (such as a fixed exploration parameter and a decreasing learning rate), we have demonstrated how a combination of controlled outcome reinforcement, plus a model of learning and re-appraisal (along with an initial willingness to deviate from embedded behaviour) on the part of the parent can result in the emergence of a stable, secure attachment style. We have also shown how a parent who places too much relative importance on future rewards (with a large discount factor δ) may actually hinder rather than help this evolution. Although we have only considered an evolution from avoidant to secure attachment, our method can in principle also be applied to dyads initially playing ambivalent or disorganised games, and indeed other non-attachment games in which we want to model the evolution of a new, stable Nash equilibrium at some particular outcome. In fact, we have highlighted how our method also maps into an internal dyad within an individual to provide a game-theoretic model of psychotherapy, which to our knowledge is new.

In the social game we considered a society of interconnected dyads (nodes), and the long term effect of social reinforcement on the actions chosen by the parents at each of these nodes in determining individual changes in attachment styles. Although we have not considered different levels of social conformity amongst the dyads, we have classified all 'malleable' social structures of 3 and 4 nodes. In future, we can apply this technique to "Small World Networks" [Watts and Strogatz, 1998].

References

[Ainsworth *et al.*, 1978] M.D.S. Ainsworth, M.C. Blehar, E. Waters, and S. Wall. *Patterns of attachment: A psychological study of the strange situation*. Lawrence Erlbaum, Hillsdale NJ, 1978.

[Allen *et al.*, 2008] J.G. Allen, P. Fonagy, and A.W. Bateman. *Mentalizing in Clinical Practice*. American psychiatric Publishing, 2008.

[Bauer, 2006] P.J. Bauer. Constructing a past in infancy: a neuro-developmental account. *Trends in Cognitive Sciences*, 10(4):175–181, 2006.

[Belsky, 1984] J. Belsky. The determinants of parenting: a process model. *Child Development*, 55 (1):83–96, 1984.

[Berne, 2010] E. Berne. *Games People Play: The Psychology of Human Relationships*. Pengiun, 2010.

[Bowlby and Ainsworth, 1953] J. Bowlby and M.D.S Ainsworth. *Child Care and the Growth of Love*. Pelican, 1953.

[Buono *et al.*,] P.L. Buono, R. Chau, G.Lewis, N. Madras, M. Pugh, L. Rossi, and T. Witelski. Mathematical models of mother/child attachment. Problem proposed by L. Atkinson, J. Hunter and B. Lancee at the Fields-MITACS Industrial Problem Solving Workshop August 2006.

[Farmer, 1990] S. Farmer. *Adult Chilren of Abusive Parents*. Ballantine Books, 1990.

[Green *et al.*, 2007] B.L. Green, C. Furrer, and McAllister C. How do relationships support parenting? effects of attachment style and social support on parenting behaviour in an at-risk population. *American Journal of Community Psychology*, 40 (1-2):96–108, 2007.

[Howard, 1960] R.A. Howard. *Dynamic Programming and Markov Processes*. MIT Press, 1960.

[Kaelbling *et al.*, 1996] L.P. Kaelbling, M.L. Littman, and A.W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[Kushnir *et al.*, 2010] T. Kushnir, F. Xu, and H.M. Wellman. Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, 22(7):1–7, 2010.

[Mehrabian, 1995] A. Mehrabian. Framework for a comprehensive description and measurement of emotional states. *Genet Soc Gen Psychol Monogr*, 1995.

[Petters *et al.*, 2010] D. Petters, E. Waters, and F. Schnbrodt. Strange carers: robots as attachment figures and aids to parenting. *Interaction studies: Social behaviour and communication in biological and artificial systems*, 2010.

[Petters, 2005] D. Petters. Building agents to understand infant attachment behaviour. *International Joint Conference on Artificial Intelligence*, 2005.

[Scherer and Roesch, 2010] K.R. Scherer and T. Bnzigerand E. Roesch. *A blueprint for affective computing: A sourcebook and manual*. Oxford University press, 2010.

[Schore, 2012] A.N. Schore. *The Science of the Art of Psychotherapy*. W. W. Norton and Co., 2012.

[Shapley, 1953] L.S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences of the United States of America*, 1953.

[Sutton and Barto, 1998] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.

[Waters *et al.*, 2000] E. Waters, S. Merrick, D. Treboux, J. Crowell, and L. Albersheim. Attachment security in infancy and early adulthood: A twenty-year longitudinal study. *Child Development*, 71, 2000.

[Watkins and Dayan, 1992] C.J.C.H Watkins and P. Dayan. Q-learning. *Machine Learning*, 8:279–292, 1992.

[Watkins, 1989] C.J.C.H. Watkins. *Learning from delayed rewards*. PhD thesis, University of Cambridge, 1989.

[Watts and Strogatz, 1998] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature* 393:440–442, 1998.

[Whitfield, 1991] C. Whitfield. *Healing the Child Within*. Health Communications, 1991.

[Young *et al.*, 2006] J.E. Young, J.S. Klosko, and M.E. Weishaar. *Schema Therapy: A Practitioner's Guide*. Guildford Press, 2006.