The purpose of this chapter is to further develop the ideas presented in Chapter ??. As mentioned, there are four main areas which have been identified for development in this field. These all aim to better apply the study to cases which are more relevant for real world Reinforcement Learning problems. To this end, the following studies are proposed.

- Dynamics of large agent populations with heterogenous agents.
- Dynamics of Reinforcement Learning in continuous action spaces.
- Dynamics of Learning in 'stateful' environments.
- Characterisation of complex dynamics using Reinforcement Learning Algorithms

In the above, the term 'stateful' is borrowed from Bloembergen et al. [?] and refers to the consideration of state transitions in the game. The remainder of this chapter will address current research in the above topics and avenues for attempting new results.

0.1 Large Agent Dynamics

This suggestion aims to build upon the work presented by Hu et al. [?] in which the point is made that the vast majority of the work done in this area considers games with a finite (usually two) number of agents. This limits the capability of the analysis to generalise to much larger agent populations, such as swarms. To allieviate this concern, the authors present an anaysis of the learning dynamics for a large agent population (which they approximate as containing infinite agents) where each agent is an independent Q-Learning using Boltzmann exploration. The emergent dynamics is given by a Fokker-Planck equation which is numerical shown to be a strong approximation of the true learning dynamics. However, as the authors point out, this is the first attempt at considering such a problem, and relies on heavy assumptions. Possible extensions therefore require a consideration of asymmetic games, stateful games (which ties in to Section 0.3) and heterogenous populations.

0.2 Continuous Action Spaces

Here lies the opportunity to extend the work of Tuyls' et al. [?] towards continuous action spaces. This would be particularly useful for most robotic applications, which operate in continuous spaces.

Work has commenced in this area, but is still in its infancy. Notable examples are: [?], in which Letcher et al. consider the dynamics of Differentiable Games, though their focus is on its application to gradient descent in GANs as opposed to reinforcement learning, and [?], in which the dynamics of Q-Learning, with a Boltzmann action selection is considered. To achieve the result, the author replaces the strategy vector with a probability density function (pdf) over the strategy space, which leads to an integro-differential equation which describes the evolution of the pdf with time. Due to its complexity, the dynamics cannot be analysed as easily as those in [?], and so Galstyan restricts the analysis to the steady-state solution (i.e. the fixed point) of the dynamics. Their analysis shows strong agreement with the designed experiments, under varying payoff conditions (note that the payoffs are now a function, rather than a discrete matrix).

Galstyan goes on to present avenues for further work. These are summarised below:

- An analysis of the steady state equations, considering the existence and uniqueness of solutions under varying payoff structures. How do these compare with the underlying Nash Equilibrium. Tuyls and Westra also suggest the consideration of stability analysis in this regard.
- A consideration of larger agent populations. No attempt towards this has yet been made as far as I am aware.
- A consideration of these dynamics under state transitions. Tuyls and Westra suggest that an analysis using switching dynamics (Chapter ??) may be useful here, but no attempt has yet been made as far as I am aware.

Tuyls and Westra [?] also consider continuous action spaces from the same perspective as Ruijgrok and Ruijgrok [?], namely that of the replicator dynamics. The former extend the analysis of the latter by considering the case where mutations are deterministic (e.g. through epsilon-greedy exploration) and these mutations only allow for small changes within the strategy space. From an EGT perspective, this analysis provides a strong characterisation of learning dynamics in continous action spaces and generalises the results of Galstyan to generic update rules (rather than the traditional Q-Learning approach).

0.3 Stateful environments

This is potentially the most relevant of the sections in regards to applicability to reinforcement learning as it extends the typical consideration of stateless normal-form games, in which the payoffs and strategies of each player is well defined, to stochastic games, in which games have probabilistic transitions across them. I have yet to explore this section in depth.

0.4 Characterisation of Complex Dynamics

The intention of this area of study is to consider the ideas presented in works such as [?] and [?] which consider complex behaviour, including cycles and chaos, in certain games using simple reinforcement learning algorithms. These provide a great deal of insight into whether or not the game will converge and, if so, to what equilibrium. Galla, for instance, shows that adjusting a memory parameter when learning to solve an Iterated Prisoner's Dilemma game can shift the equilibrium from one showing purely defective behaviour to one showing cycles of cooperations and defection. Sanders et al. generalise this result by showing that, for a particular learning algorithm known as Experience Weighted Attraction, learning dynamics varies dependent on the mutual effect of two parameters.

This determination of where RL algorithms will show complex dynamics is important, not only for a priori understanding of resultant behaviour, but is particularly important for algorithms in which agents aim to predict the behaviour of other agents. For this to be feasible, the learning dynamics must not exhibit chaos, otherwise it would be impossible for an agent to make any reasonable predictions about the future behaviour of its opponents.