Reviewer 4 The reviewer makes a pertinent point regarding the assumptions made in the model chosen for the analysis. Indeed, we remark in the conclusion that, as EGT models are improved to capture such complex games as you mention (such as with active environments or heterogeneous agents), we must analyse those models to understand their implications for stability. This is a point for future work.

Regarding the novelty of the work, it should be noted that the references (Tuyls et al, Bloembergen et al) focus on developing models for learning algorithms. We, on the other hand, focus on understanding what these models tell us about the behaviour of Q-Learning. The closest analogue to our own work is that of Leonardos and Piliouras, who show that the behaviour of Q-Learning can vary depending on the exploration-exploitation parameter. For this specific case, they find, as we do, that the behaviour of Q-Learning can become rather complex. We, however, aim to analyse also the influence that the number of players and number of actions has, which we show through both our analysis and our experimentation. There has not been a statistical study, such as ours, which aims at comprehensively classifying how stability can be affected by increasing the number of players as well as learning parameters in the Q-Learning literature.

Reviewer 6 We thank the reviewer for their analysis and comments. Specifically, we wish to address the reviewer's concerns regarding the averaging over payoff matrices. We should start by pointing out that we do not average the effect over all possible games of interaction. Our aim instead is to understand how the 'type' of game (e.g. zero-sum, coordination etc) affects the stability of the system. We make the point that, for example, there are an infinite number of zero-sum games, and so it would not be instructive if we were to just take one zero-sum game and claim that the stability in this case is indicative of all zero-sum games. Rather, we choose a parameter Γ which gives an insight into the correlations between the payoff matrices. This is a completely standard method in the literature (c.f. (Sanders et al 2018, Galla 2013)). For example, for a two-player game, $\Gamma = -1$ corresponds to a zero-sum game, while $\Gamma = 1$ is a coordination game.

Now, with Γ defined as such, if we are to analyse its effect on stability, we must average over all games which share the same Γ. We must stress that to only analyse one zero-sum game would not be indicative of all such games. Therefore, we find the 'average' dynamics for a particular choice of Γ (again, this is an established technique in the literature). Similarly, in our experiments, if we are to analyse how the 'type' of game affects stability, we must average over multiple games of the same type. A critical finding is that, for Q-Learning, the type of game has almost no influence on the stability of the learning dynamics..

Reviewer 37 The reviewer makes some very interesting points regarding the domain considered. It is true that we focus on the restricted domain for Γ, as we were able to ensure that our theoretical analysis was sound. Unfortunately, the analysis does not extend to positive Γ. You are correct in stating that the behaviour in this regime can be analysed through a numerical study. This is a point of work that we considered immediately upon conclusion of our present work. However, we felt that it was important to focus on the region in which the stability of the system could be analytically determined. We shall be sure to clarify your point in revision.

We thank the reviewer for making their point regarding the onset of chaos. Indeed it is true that our results focus on the instability of the system. However, we also numerically estimated Lypapunov exponents using the Kantz algorithm and found them to be positive in games of large players/actions. We will be sure to make this point clear in our revision also. Regarding the existence of periodic orbits, we conducted experiments which look for recurrent behaviour, but were unable to find any, even for small classes of 2x2 games. A point of work which we are looking at is to consider the possibility that Q-Learning may not admit any periodic solutions at all.

Reviewer 65 We thank the reviewer for pointing out the clarification needed in the payoff matrices. We must clarify that the payoffs themselves are not Gaussian. Each agent receives their reward from a payoff matrix. Our analysis requires that we understand how the structure of the 'type' of game affects stability. As such, we parameterise the correlations between the payoff elements by Γ. Now to truly understand the effect of this parameter, we must analyse the 'average' behaviour over all games which share the same Γ. To do this, we assert that the payoffs are drawn from a multivariate Gaussian, allowing us to leverage Gaussian identities. Therefore, we are not restricting to a particular class of game, but rather generalising over all possible payoff matrices and averaging the effect of those who share the same payoff correlations.

Regarding the analysis of other reinforcement learning algorithms, we definitely believe that this is an important and interesting line of inquiry, but far beyond the scope of the current contribution. We mention in our conclusion that, as dynamical models are developed for the algorithms you mention, analyses such as the present work will help provide insight into what they predict regarding stability.

Reviewer 90 The reviewer makes quite useful suggestions regarding using some standard references for game theoretic concepts. Indeed, our choice of references was focused on those which give the most relevant insights into the results of the present work. We will, however, add these references in our revision.

Regarding the behaviour of τ, which partially encodes exploration-exploitation, we chose not to focus on this result as it was analysed in (Leonardos and Piliouras 2020). Furthermore, the analytic result shows that the dependence of stability on τ shares the same relationship as α. We shall be sure to clarify this, and we thank the reviewer for pointing it out.

Finally regarding the number of iterations for the numerical analysis and number of payoff matrices analysed, we ensured that our analysis covered a number of initial conditions for each payoff matrix, and that a small enough criterion for convergence was applied by visually assessing the games to make sure they have converged. Furthermore, we can see that the numerical results confidently confirm the predictions 1 - 5 made in the Discussion (Sec 3.4).