8 Ethics Statement

General-sum games are an important model for a wide range of real-world scenarios, but have been studied less than zero-sum games. Our work contributes to the foundational study of learning algorithms for these games. It is important to develop learning algorithms that can prevent defect-defect outcomes in situations modeled by the prisoner's dilemma, such as public goods problems. Opponent-shaping methods, as studied in this paper, can achieve this. However, cooperation between the players can also be undesirable, for instance, in markets, where it leads to collusion. Moreover, in the long-term, opponent-shaping could allow agents to manipulate others, including humans. It could thus make sense to also study such negative consequences of opponent-shaping and the limits of their responsible use in the future. However, we believe that at this stage the benefits of furthering our understanding and of developing algorithms that can achieve more cooperative learning outcomes outweigh their downsides.

References

Stefano V. Albrecht and Peter Stone. Reasoning about hypothetical agent behaviours and their parameters, 2019.

Robert Axelrod. *The Evolution of Cooperation*. Basic, New York, 1984.

Waïss Azizian, Ioannis Mitliagkas, Simon Lacoste-Julien, and Gauthier Gidel. A tight and unified analysis of gradient-based methods for a whole spectrum of games, 2020.

David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. The Mechanics of n-Player Differentiable Games. *arXiv:1802.05642 [cs]*, June 2018. URL `http://arxiv.org/abs/1802.05642`. arXiv: 1802.05642.

Andrew Barto and Sridhar Mahadevan. Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems: Theory and Applications*, 13, 12 2002. doi: 10.1023/A:1025696116075.

Jakob N. Foerster, Richard Y. Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. Learning with Opponent-Learning Awareness. *arXiv:1709.04326 [cs]*, September 2018. URL `http://arxiv.org/abs/1709.04326`. arXiv: 1709.04326.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL `https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf`.

Werner Güth, Rolf Schmittberger, and Bernd Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4):367–388, December 1982. ISSN 0167-2681. doi: 10.1016/0167-2681(82)90011-7. URL `https://www.sciencedirect.com/science/article/pii/0167268182900117`.

Marc Harper, Vincent Knight, Martin Jones, Georgios Koutsovoulos, Nikoleta E. Glynatsi, and Owen Campbell. Reinforcement learning produces dominant strategies for the iterated prisoner's dilemma. *PLOS ONE*, 12(12):e0188046, Dec 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0188046. URL `http://dx.doi.org/10.1371/journal.pone.0188046`.

He He, Jordan Boyd-Graber, Kevin Kwok, and Hal Daumé III au2. Opponent modeling in deep reinforcement learning, 2016.

Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and Herbert Gintis. Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence From Fifteen Small-Scale Societies. In *American Anthropologist - AMER ANTHROPOL*, volume 108. January 2006. ISBN 978-0-19-926205-2. doi: 10.1093/0199262055.001.0001. Journal Abbreviation: American Anthropologist - AMER ANTHROPOL.

Adrian Hutter. Learning in two-player games between transparent opponents, 2020.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

King Lee and Louis K. *The Application of Decision Theory and Dynamic Programming to Adaptive Control Systems*. Thesis, September 1967. URL https://macsphere.mcmaster.ca/handle/11375/18669. Accepted: 2015-12-21T20:09:58Z.

Alistair Letcher. Stability and exploitation in differentiable games. Master's thesis, University of Oxford, 2018.

Alistair Letcher, David Balduzzi, Sebastien Racaniere, James Martens, Jakob Foerster, Karl Tuyls, and Thore Graepel. Differentiable Game Mechanics. *arXiv:1905.04926 [cs, stat]*, May 2019a. URL http://arxiv.org/abs/1905.04926. arXiv: 1905.04926.

Alistair Letcher, Jakob Foerster, David Balduzzi, Tim RocktÃ€schel, and Shimon Whiteson. Stable Opponent Shaping in Differentiable Games. *arXiv:1811.08469 [cs]*, January 2019b. URL http://arxiv.org/abs/1811.08469. arXiv: 1811.08469.

Eric V. Mazumdar, Michael I. Jordan, and S. Shankar Sastry. On Finding Local Nash Equilibria (and Only Local Nash Equilibria) in Zero-Sum Games. *arXiv:1901.00838 [cs, math, stat]*, January 2019. URL http://arxiv.org/abs/1901.00838. arXiv: 1901.00838.

Richard Mealing and Jonathan L. Shapiro. Opponent modeling by expectation–maximization and sequence prediction in simplified poker. *IEEE Transactions on Computational Intelligence and AI in Games*, 9(1):11–24, 2017. doi: 10.1109/TCIAIG.2015.2491611.

Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. The Numerics of GANs. *arXiv:1705.10461 [cs]*, June 2018. URL http://arxiv.org/abs/1705.10461. arXiv: 1705.10461.

Hessel Oosterbeek, Randolph Sloof, and Gijs van de Kuilen. Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Experimental Economics*, 7(2):171–188, June 2004. ISSN 1573-6938. doi: 10.1023/B:EXEC.0000026978.14316.74. URL https://doi.org/10.1023/B:EXEC.0000026978.14316.74.

Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. The MIT Press, 1994. ISBN 0262150417.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Sébastien Racanière, Theophane Weber, David Reichert, Lars Buesing, Arthur Guez, Danilo Jimenez Rezende, Adrià Puigdomènech Badia, Oriol Vinyals, Nicolas Heess, Yujia Li, Razvan Pascanu, Peter Battaglia, Demis Hassabis, David Silver, and Daan Wierstra. Imagination-augmented agents for deep reinforcement learning. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/9e82757e9a1c12cb710ad680db11f6f1-Paper.pdf.

Alan G. Sanfey, James K. Rilling, Jessica A. Aronson, Leigh E. Nystrom, and Jonathan D. Cohen. The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science*, 300(5626): 1755–1758, June 2003. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1082976. URL https://science.sciencemag.org/content/300/5626/1755. Publisher: American Association for the Advancement of Science Section: Report.

Florian Schäfer and Anima Anandkumar. Competitive Gradient Descent. *arXiv:1905.12103 [cs, math]*, June 2020. URL `http://arxiv.org/abs/1905.12103`. arXiv: 1905.12103.

Florian Schäfer, Anima Anandkumar, and Houman Owhadi. Competitive mirror descent, 2020.

J. Schmidhuber. A possibility for implementing curiosity and boredom in model-building neural controllers. In J. A. Meyer and S. W. Wilson (eds.), *Proc. of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, pp. 222–227. MIT Press/Bradford Books, 1991.

Gabriel Synnaeve and Pierre Bessiere. A Bayesian Model for Opening Prediction in RTS Games with Application to StarCraft. In *Computational Intelligence and Games*, pp. 000, Seoul, South Korea, August 2011. URL `https://hal.archives-ouvertes.fr/hal-00607277`.

Ben G. Weber and Michael Mateas. A data mining approach to strategy prediction. In *2009 IEEE Symposium on Computational Intelligence and Games*, pp. 140–147, 2009. doi: 10.1109/CIG.2009.5286483.

Ying Wen, Yaodong Yang, Rui Luo, Jun Wang, and Wei Pan. Probabilistic recursive reasoning for multi-agent reinforcement learning, 2019.

C. Zhang and V. Lesser. Multi-Agent Learning with Policy Prediction. *AAAI Conference on Artificial Intelligence*, 2010.

A NONCONVERGENCE OF HOLA IN THE TANDEM GAME

In the following, we show that for the choice of look-ahead rate $\alpha = 1$, HOLA does not converge in the Tandem game. This shows that given a large enough look-ahead rate, even in a simple quadratic game, HOLA need not converge.

Proposition 6. *Let L^1, L^2 be the two players' loss functions in the Tandem game as defined in Section 5 and let h_i^n denote the n-th order exact LOLA update for player i (where $n = 0$ denotes naive learning). Consider the look-ahead rate $\alpha := 1$. Then the functions $(h_i^n)_{n\in\mathbb{N}}$ for $i = 1, 2$ do not converge pointwise.*

Proof. We will prove the auxiliary statement that

$$h_i^n(x, y) = 2^{n+2} - 2(1 + x + y)$$

for $i = 1, 2$. It then follows trivially that the h_i^n cannot converge.

The auxiliary result can be proven by induction. The base case $n = 0$ follows from

$$\nabla_i L^i(x, y) = 2 - 2(x + y) = 2^2 - 2(1 + x + y)$$

for $i = 1, 2$. Next, for the inductive step, we have to show that

$$h_1^n(x, y) = -\nabla_1(L^i(x, y + h_2^{n-1}(x, y))) \tag{9}$$

$$h_2^n(x, y) = -\nabla_2(L^i(x + h_1^{n-1}(x, y), y)) \tag{10}$$

for any $n > 0$. This is also a simple calculation of a derivative, which we have verified using Mathematica. The relevant notebook can be found at `https://www.wolframcloud.com/obj/jonny0/Published/HOLA-nonconvergence-tandem.nb` □

B PROOF OF PROPOSITION 1

To begin, recall that some differentiable game with continuously differentiable loss functions L^1, L^2 is given, and that $h^n = (h_1^n, h_2^n)$ denotes the n-th order exact LOLA update function. We assume that the iLOLA update function h exists, defined via

$$h_i(\theta) := \lim_{n\to\infty} h_i^n(\theta),$$

for all $\theta \in \mathbb{R}^d$.

To prove Proposition 1, we need to show that h_1, h_2 are consistent, i.e., satisfy Definition 3, under the assumption that

$$\lim_{n\to\infty} \nabla_i h_{-i}^n(\theta) = \nabla_i h_{-i}(\theta)$$

for $i = 1, 2$ and any θ.

To that end, define the *(exact) LOLA operator* Ψ as the function mapping a pair of update functions $f := (f_1, f_2)$ to the RHS of Equations 2 and 3,

$$\Psi_1(f)(\theta) := -\alpha\nabla_1(L^1(\theta_1, \theta_2 + f_2(\theta_1, \theta_2))) \tag{11}$$

$$\Psi_2(f)(\theta) := -\alpha\nabla_2(L^2(\theta_1 + f_1(\theta_1, \theta_2), \theta_2)) \tag{12}$$

for any θ. Note that then we have $h_i^{n+1} = \Psi_i(h^n)$, i.e., Ψ maps n-th order LOLA to $n + 1$-order LOLA.

In the following, we show that iLOLA is a fixed point of the LOLA operator, i.e., $\Psi(h) = h$. It follows from the definition of Ψ that then h is consistent. We denote by $\|\cdot\|$ the Euclidean norm or the induced operator norm for matrices. We focus on showing $\Psi_1(h) = h_1$. The case $i = 2$ is exactly analogous.

For arbitrary θ and n, define $\hat{\theta}_2 := \theta_2 + h_2(\theta)$ and $\hat{\theta}_2^n := \theta_2 + h_n^2(\theta)$ as the updated parameter of player 2. First, it is helpful to show that $\Psi_1(h^n)(\theta)$ converges to $\Psi_1(h)(\theta)$:

$$
\begin{aligned}
0 &\leq \|\Psi_1(h)(\theta) - \Psi_1(h^n)(\theta)\| && (13)\\
&= \alpha\|\nabla_1(L^1(\theta_1, \theta_2 + h_2(\theta))) - \nabla_1(L^1(\theta_1, \theta_2 + h_2^n(\theta))\| && (14)\\
&= \alpha\|(\nabla_1 h_2(\theta))^\top \nabla_2 L^1(\theta_1, \hat{\theta}_2) - (\nabla_1 h_2^n(\theta))^\top \nabla_2 L^1(\theta_1, \hat{\theta}_2^n) + \nabla_1 L^1(\theta_1, \hat{\theta}_2) - \nabla_1 L^1(\theta_1, \hat{\theta}_2^n)\| && (15)\\
&\leq \alpha\|(\nabla_1 h_2(\theta))^\top \nabla_2 L^1(\theta_1, \hat{\theta}_2) - (\nabla_1 h_2^n(\theta))^\top \nabla_2 L^1(\theta_1, \hat{\theta}_2^n)\| + \alpha\|\nabla_1 L^1(\theta_1, \hat{\theta}_2) - \nabla_1 L^1(\theta_1, \hat{\theta}_2^n)\| && (16)\\
&= \alpha\|(\nabla_1 h_2(\theta))^\top (\nabla_2 L^1(\theta_1, \hat{\theta}_2) - \nabla_2 L^1(\theta_1, \hat{\theta}_2^n)) && (17)\\
&\quad + (\nabla_1 h_2(\theta) - \nabla_1 h_2^n(\theta))^\top \nabla_2 L^1(\theta_1, \hat{\theta}_2^n)\| + \alpha\|\nabla_1 L^1(\theta_1, \hat{\theta}_2) - \nabla_1 L^1(\theta_1, \hat{\theta}_2^n)\|\\
&\leq \alpha\|(\nabla_1 h_2(\theta))^\top\|\|\nabla_2 L^1(\theta_1, \hat{\theta}_2) - \nabla_2 L^1(\theta_1, \hat{\theta}_2^n)\|\\
&\quad + \alpha\|(\nabla_1 h_2(\theta) - \nabla_1 h_2^n(\theta))^\top\|\|\nabla_2 L^1(\theta_1, \hat{\theta}_2^n)\| + \alpha\|\nabla_1 L^1(\theta_1, \hat{\theta}_2) - \nabla_1 L^1(\theta_1, \hat{\theta}_2^n)\| && (18)\\
&\overset{n\to\infty}{\longrightarrow} 0. && (19)
\end{aligned}
$$

In the last step, we used the following two facts. First, since $\nabla_i L^1(\theta)$ is assumed to be continuous in θ_2, and $\lim_{n\to\infty} \hat{\theta}_2^n = \theta_2 + \lim_{n\to\infty} h_2^n(\theta) = \theta_2 + h_2(\theta) = \hat{\theta}_2$ by assumption, it follows that $\lim_{n\to\infty} \nabla_2 L^1(\theta_1, \hat{\theta}_2^n) = \nabla_2 L^1(\theta_1, \hat{\theta}_2)$ and $\lim_{n\to\infty} \nabla_1 L^1(\theta_1, \hat{\theta}_2^n) = \nabla_1 L^1(\theta_1, \hat{\theta}_2)$. Second, by assumption, $\lim_{n\to\infty} \nabla h_2^n(\theta) = \nabla h_2(\theta)$. In particular, both $\|\nabla_2 L^1(\theta_1, \hat{\theta}_2^n)\|$ and $\|\nabla_1(h_2(\theta))^\top - \nabla_1(h_2^n(\theta))^\top\|$ must be bounded, and thus the three terms in (18) must all converge to 0 as $n \to \infty$. It follows by the sandwich theorem that $\lim_{n\to\infty} \Psi_1(h^n)(\theta) = \Psi_1(h)(\theta)$.

Now we can directly prove that $\Psi_1(h)(\theta) = h_1(\theta)$. It is

$$
\begin{aligned}
0 &\leq \|\Psi_1(h)(\theta) - h_1(\theta)\| && (20)\\
&= \|\Psi_1(h)(\theta) - \Psi_1(h^n)(\theta) + \Psi_1(h^n)(\theta) - h_1^n(\theta) + h_1^n(\theta) - h_1(\theta)\| && (21)\\
&\leq \|\Psi_1(h)(\theta) - \Psi_1(h^n)(\theta)\| + \|\Psi_1(h^n)(\theta) - h_1^n(\theta)\| + \|h_1^n(\theta) - h_1(\theta)\| && (22)\\
&= \|\Psi_1(h)(\theta) - \Psi_1(h^n)(\theta)\| + \|h_1^{n+1}(\theta) - h_1^n(\theta)\| + \|h_1^n(\theta) - h_1(\theta)\| && (23)\\
&\overset{n\to\infty}{\longrightarrow} 0, && (24)
\end{aligned}
$$

where in the last step we have used the above result, as well as the assumption that $h_1^n(\theta)$ converges pointwise, and thus must also be a Cauchy sequence, so the last and the middle term both converge to zero as well.

It follows by the sandwich theorem that $\Psi_1(h)(\theta) = h_1(\theta)$. Since θ was arbitrary, this concludes the proof. □

C INFINITE-ORDER TAYLOR LOLA

In this Section, we repeat the analysis of iLOLA from Section 4.1 for infinite-order Taylor LOLA (Taylor iLOLA). I.e., we define Taylor consistency, and show that Taylor iLOLA satisfies this consistency equation under certain assumptions. This result will be needed for our proof of Proposition 2.

To begin, assume that some differentiable game with continuously differentiable loss functions L^1, L^2 is given. Define the Taylor LOLA operator Φ that maps pairs of update functions (f_1, f_2) to the associated Taylor LOLA update

$$\Phi_i(f) := -\alpha\nabla_i(L^i + (\nabla_{-i}L^i)^\top f_{-i}) \tag{25}$$

for $i = 1, 2$.

We then have the following definition.

Definition 4 (Taylor consistency)**.** Two update functions h_1, h_2 are called Taylor consistent if for any $i = 1, 2$, it is

$$\Phi(h_1, h_2) = (h_1, h_2).$$

Next, let h_i^n denote i's n-th order Taylor LOLA update. I.e., $h_i^n := \Phi_i(h^{n-1})$ for $n \geq 0$, where we let $h_i^{-1} := 0$. Then we define

Definition 5 (Taylor iLOLA). If (h_1^n, h_2^n) converges pointwise as $n \to \infty$, define Taylor iLOLA as the limiting update

$$h := \lim_{n\to\infty} \begin{pmatrix} h_1^n \\ h_2^n . \end{pmatrix}$$

Finally, we provide a proof that Taylor iLOLA is Taylor consistent; i.e., we give a Taylor version of Proposition 1.

Proposition 7. *Let h_i^n denote player i's n-th order Taylor LOLA update. Assume that $\lim_{n\to\infty} h_i^n(\theta) = h_i(\theta)$ and $\lim_{n\to\infty} \nabla_i h_{-i}^n(\theta) = \nabla_i h_{-i}(\theta)$ for all θ and $i \in \{1, 2\}$. Then Taylor iLOLA is Taylor consistent.*

Proof. The proof is exactly analogous to that of Proposition 1, but easier. We show $\Phi(h) = h$. It follows from the definition of Φ in Equation 25 that then h is Taylor consistent. We focus on showing $\Phi_1(h) = h_1$, and the case $i = 2$ is exactly analogous.

First, we show that $\Phi_1(h^n)(\theta)$ converges to $\Phi_1(h)(\theta)$ for all θ. Letting n be arbitrary and omitting θ in the following for clarity, it is

$$\begin{aligned}
0 &\leq \|\Phi_1(h) - \Phi_1(h^n)\| && (26)\\
&= \| -\alpha\nabla_1(L^1 + (\nabla_2 L^1)^\top h_2) + \alpha\nabla_1(L^1 + (\nabla_2 L^1)^\top h_2^n)\| && (27)\\
&= \alpha\| -\nabla_{12}L^1 h_2 - (\nabla_2 L^1)^\top(\nabla_1 h_2)^\top + \nabla_{12}L^1 h_2^n + (\nabla_2 L^1)^\top(\nabla_1 h_2^n)^\top\| && (28)\\
&\leq \alpha\|\nabla_{12}L^1(h_2^n - h_2)\| + \alpha\|(\nabla_2 L^1)^\top(\nabla_1 h_2^n - \nabla_1 h_2)^\top\| && (29)\\
&\leq \alpha\|\nabla_{12}L^1\|\|h_2^n - h_2\| + \alpha\|\nabla_2 L^1\|\|\nabla_1 h_2^n - \nabla_1 h_2\| && (30)\\
&\xrightarrow{n\to\infty} 0. && (31)
\end{aligned}$$

In the last step, we used the assumptions that $\lim_{n\to\infty} h_2^n = h_2$ and $\lim_{n\to\infty} \nabla h_2^n = \nabla h_2$. It follows by the sandwich theorem that $\lim_{n\to\infty} \Phi_1(h^n)(\theta) = \Phi_1(h)(\theta)$.

It follows from the above that $\Phi_1(h)(\theta) = h_1(\theta)$, using exactly the same argument as in Equations 20-24 with Φ instead of Ψ. Since θ was arbitrary, this concludes the proof.

□

D Proof of Proposition 2

We begin by proving that LCGD and CGD do not coincide with LOLA and iLOLA. It is sufficient to manifest a single counter-example: we consider the Tandem game given by $L^1 = (x+y)^2 - 2x$ and $L^2 = (x+y)^2 - 2y$ (using x, y instead of θ_1, θ_2 for simplicity). Throughout this proof we use the notation introduced by Balduzzi et al. (2018) and Letcher et al. (2019b) including the *simultaneous gradient*, the *off-diagonal Hessian* and the *shaping term* of the game as

$$\xi = \begin{pmatrix} \nabla_1 L^1 \\ \nabla_2 L^2 \end{pmatrix} \qquad \text{and} \qquad H_o = \begin{pmatrix} 0 & \nabla_{12}L^2 \\ \nabla_{21}L^2 & 0 \end{pmatrix} \qquad \text{and} \qquad \chi = \operatorname{diag}\left(H_o^T \nabla L\right)$$

respectively. Note that in two-player games, LOLA's shaping term reduces to

$$\chi = \begin{pmatrix} \nabla_{12}L^2 \nabla_2 L^1 \\ \nabla_{21}L^1 \nabla_1 L^2 \end{pmatrix} .$$

LCGD ≠ LOLA. Following Schäfer & Anandkumar (2020), LCGD is given by

$$\text{LCGD} = -\alpha \begin{pmatrix} \nabla_x f - \alpha D_{xy}^2 f \nabla_y g \\ \nabla_y g - \alpha D_{yx}^2 g \nabla_x f \end{pmatrix} = -\alpha \begin{pmatrix} I & -\alpha D_{xy}^2 f \\ -\alpha D_{yx}^2 g & I \end{pmatrix} \begin{pmatrix} \nabla_x f \\ \nabla_y g \end{pmatrix} = -\alpha(I - \alpha H_o)\xi$$

while LOLA is given (Letcher et al., 2019b) by

$$\text{LOLA} = -\alpha(I - \alpha H_o)\xi + \alpha^2 \chi .$$

Any game with $\chi \neq 0$ will yield a difference between LCGD and LOLA; in particular,

$$\chi = 4(x+y)\begin{pmatrix}1\\1\end{pmatrix}$$

in the Tandem game implies that LCGD $\neq$ LOLA whenever parameters lie outside the measure-zero set $\{x+y=0\} \subset \mathbb{R}^2$.

CGD does not recover HOLA. Since CGD is obtained through a bilinear approximation (Taylor expansion) of the loss functions, one would expect that the authors' claim of recovering HOLA is with regards to Taylor (not exact) HOLA. For completeness, and to avoid any doubts for the reader, we prove that CGD neither corresponds to exact nor Taylor HOLA.

Following Schäfer & Anandkumar (2020), the series-expansion of CGD is given by

$$\text{CGD}n = -\alpha \sum_{i=0}^{n} \begin{pmatrix} 0 & -\alpha D^2_{xy} f \\ -\alpha D^2_{yx} g & 0 \end{pmatrix}^i \begin{pmatrix} D_x f \\ D_y g \end{pmatrix} = -\alpha \sum_{i=0}^{n} (-\alpha H_o)^i \xi$$

and converges to CGD whenever $\alpha < 1/\|H_o\|$ (where $\|\cdot\|$ denotes the operator norm induced by the Euclidean norm on the space). Assume for contradiction that the series-expansion of CGD recovers HOLA, i.e. that CGDn = HOLAn for all n. In particular, we must have

$$\text{CGD} = \lim_{n\to\infty} \text{CGD}n = \lim_{n\to\infty} \text{HOLA}n = \text{iLOLA}$$

whenever $\alpha < 1/\|H_o\|$. In the tandem game, we have

$$H_o = 2\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

with $\|H_o\| = 2$, so CGD = iLOLA whenever $\alpha < 1/2$. Moreover, H_o being constant implies that

$$\nabla \text{HOLA}n = \nabla \text{CGD}n = -\alpha \sum_{i=0}^{n} (-\alpha H_o)^i \nabla \xi \,,$$

so gradients of HOLA also converge pointwise for all $\alpha < 1/2$. In particular, CGD = iLOLA must satisfy the (exact or Taylor) consistency equations by Proposition 1 or Proposition 7. However, the update for CGD is given by

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = -\alpha(I+\alpha H_o)^{-1}\xi = -2\alpha(x+y-1)\begin{pmatrix} 1 & 2\alpha \\ 2\alpha & 1 \end{pmatrix}^{-1}\begin{pmatrix}1\\1\end{pmatrix} = \frac{-2\alpha(x+y-1)}{1+2\alpha}\begin{pmatrix}1\\1\end{pmatrix}.$$

For the exact case, the RHS of the first consistency equation is

$$\begin{aligned} -\alpha\nabla_x\left((x+y+h_2)^2 - 2x\right) &= -2\alpha\left((1+\nabla_x h_2)(x+y+h_2) - 1\right) \\ &= \frac{-2\alpha}{1+2\alpha}\left(x+y+\frac{-2\alpha(x+y-1)}{1+2\alpha} - 1 - 2\alpha\right) \\ &= h_1 + \frac{4\alpha^2(x+y+2\alpha)}{(1+2\alpha)^2} \end{aligned}$$

which does not coincide with the LHS of the consistency equation ($= h_1$) whenever parameters lie outside the measure-zero set $\{x+y+2\alpha=0\} \subset \mathbb{R}^2$. Similarly for Taylor iLOLA, the RHS of the first consistency equation is

$$\begin{aligned} -\alpha\nabla_x\left((x+y)^2 - 2x + 2(x+y)h_2\right) &= -2\alpha\left(x+y-1+\frac{-2\alpha(x+y-1)}{1+2\alpha} + \frac{-2\alpha(x+y)}{1+2\alpha}\right) \\ &= h_1 + \frac{4\alpha^2(x+y)}{(1+2\alpha)^2} \end{aligned}$$

which does not coincide with the LHS of the consistency equation ($= h_1$) whenever parameters lie outside the measure-zero set $\{x+y=0\} \subset \mathbb{R}^2$. This is a contradiction to consistency; we are done.

LCGD = LookAhead. We have already shown that LCGD is given by $-\alpha(I-\alpha H_o)\xi$ in the proof that LCGD $\neq$ LOLA. This coincides exactly with LookAhead following Letcher et al. (2019b).

CGD recovers high-order LookAhead. The series expansion of CGD is given by

$$\text{CGD}n = -\alpha \sum_{i=0}^{n} (-\alpha H_o)^i \xi \,.$$

On the other hand, high-order (Taylor) LookAhead is defined recursively by expanding

$$h_1^n = -\alpha \nabla_1 \left(L^1(\theta^1, \theta^2 + \perp h_2^{n-1})\right) \approx -\alpha \left(\nabla_1 L^1 + \nabla_{21} L^1 h_2^{n-1}\right)$$
$$h_2^n = -\alpha \nabla_2 \left(L^2(\theta^1 + \perp h_1^{n-1}, \theta^2)\right) \approx -\alpha \left(\nabla_2 L^2 + \nabla_{12} L^2 h_1^{n-1}\right) \,,$$

where $\perp$ is the stop-gradient operator (see (Balduzzi et al., 2018) for details on this operator) and $h_1^{-1} = h_2^{-1} = 0$. This can be written more succinctly as

$$\begin{pmatrix} h_1^n \\ h_2^n \end{pmatrix} = -\alpha \begin{pmatrix} \nabla_1 L^1 + \nabla_{21} L^1 h_2^{n-1} \\ \nabla_2 L^2 + \nabla_{12} L^2 h_1^{n-1} \end{pmatrix} = -\alpha\xi - \alpha H_o \begin{pmatrix} h_1^{n-1} \\ h_2^{n-1} \end{pmatrix} .$$

We prove by induction that

$$\begin{pmatrix} h_1^n \\ h_2^n \end{pmatrix} = -\alpha \sum_{i=0}^{n} (-\alpha H_o)^i \xi$$

for all $n \geq 0$. The base case is trivial; assume the statement holds for any fixed $n \geq 0$. Then

$$\begin{pmatrix} h_1^{n+1} \\ h_2^{n+1} \end{pmatrix} = -\alpha\xi - \alpha H_o \left(-\alpha \sum_{i=0}^{n} (-\alpha H_o)^i \xi\right) = -\alpha\xi - \alpha \sum_{i=1}^{n+1} (-\alpha H_o)^i \xi = -\alpha \sum_{i=0}^{n+1} (-\alpha H_o)^i \xi$$

as required. Finally we conclude

$$\text{LookAhead}n = \begin{pmatrix} h_1^n \\ h_2^n \end{pmatrix} = -\alpha \sum_{i=0}^{n} (-\alpha H_o)^i \xi = \text{CGD}n$$

as required. $\square$

E Proof of Proposition 3

We prove that the two pairs of linear functions

$$h_1 = h_2 = -2(x + y + 1)$$

and

$$h_1 = h_2 = -\frac{1}{2}(x + y - 2)$$

are solutions to the consistency equations in the Tandem game with $\alpha = 1$. (See below for a generalization to any $\alpha > 0$.) For the first pair of functions, we have

$$-\nabla_x \left(f(x, y + h_2)\right) = -\nabla_x \left((x + y + 2)^2 - 2x\right) = -2(x + y + 1) = h_1$$

for the first consistency equation and similarly

$$-\nabla_y \left(g(x + h_1, y)\right) = -\nabla_x \left((x + y + 2)^2 - 2y\right) = -2(x + y + 1) = h_2$$

for the second. For the second pair of functions we similarly obtain

$$-\nabla_x \left(f(x, y + h_2)\right) = -\nabla_x \left(\frac{1}{4}(x + y + 2)^2 - 2x\right) = -\frac{1}{2}(x + y - 2) = h_1$$

for the first consistency equation and

$$-\nabla_y \left(g(x + h_1, y)\right) = -\nabla_x \left(\frac{1}{4}(x + y + 2)^2 - 2y\right) = -\frac{1}{2}(x + y - 2) = h_2$$

for the second. This shows that both functions are solutions to the consistency equations. For general $\alpha > 0$, we can similarly show that $h_1 = h_2 = ax + by + c$ with

$$a = \frac{\pm\sqrt{1 + 8\alpha} - 1 - 4\alpha}{4\alpha} \quad ; \quad b = \frac{-2\alpha(1 + a)}{1 + 2\alpha(1 + a)} \quad ; \quad c = \frac{2\alpha}{1 + 2\alpha(1 + a)}$$

are two distinct solutions (depending on $\pm$) to the consistency equations in the Tandem game, noting that the denominators cannot be 0 for $\alpha > 0$ (otherwise leading to a contradiction in the expression for a). This is left to the reader, noting that the proof for $\alpha = 1$ is sufficient to establish that consistent solutions are not always unique. $\square$

F PROOF OF PROPOSITION 4

Recall from the proof of Proposition 3 that the linear functions

$$h_1 = h_2 = -2(x + y + 1)$$

are consistent solutions to the Tandem game with $\alpha = 1$. The SFPs of the Tandem game are $(x, 1-x)$ for each $x \in \mathbb{R}$, but none of these are preserved by the consistent solutions above since

$$h_1(x, 1-x) = h_2(x, 1-x) = -4 \neq 0 \,.$$

We conclude that consistency does not imply preservation of SFPs.

Moreover, we prove that any (non-zero) linear solution to the consistency equations cannot preserve more than one SFP in the Tandem game, for any opponent shaping rate α. Assuming it did, we must have linear functions

$$h_1 = ax + by + c$$
$$h_2 = a'x + b'y + c'$$

satisfying

$$h_1(x, 1-x) = 0 = h_1(x', 1-x')$$

for some $x \neq x' \in \mathbb{R}$. Subtracting RHS from LHS we obtain $(x - x')(a - b) = 0$ hence $a = b$, which substituted again into the LHS yields $b = -c$. Applying the same method for h_2 we obtain $a' = b' = -c'$ and so h_1, h_2 take the form

$$h_1 = a(x + y - 1)$$
$$h_2 = a'(x + y - 1).$$

Note that since h_1, h_2 were assumed to be nonzero, it follows that $a, a' \neq 0$. Plugging these into the first consistency equation, we obtain

$$a(x + y - 1) = -2\alpha \left[(1 + a')\left((x + y)(1 + a') - a'\right) - 1\right] \,.$$

Comparing x terms and constant terms yields

$$a = -2\alpha(1 + a')^2 \qquad \text{and} \qquad a = -2\alpha\left(1 + a' + {a'}^2\right)$$

which concludes the contradiction $a' = 0$. □

G PROOF OF PROPOSITION 5

LOLA and SOS diverge. Assume $(x_0, y_0) \neq 0$ and $\alpha > 1$. We prove the more general claim that p-LOLA diverges for any $0 \leq p \leq 1$ (where p may take a different value at each learning step), recalling that LOLA and SOS are both special cases of p-LOLA (Letcher et al., 2019b). Indeed, the p-LOLA gradient update is given by

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = -\alpha(I - \alpha H_o)\xi + p\alpha^2\chi = -\alpha \begin{pmatrix} y + \alpha x(1 + p) \\ -x + \alpha y(1 + p) \end{pmatrix}$$

and we show that each update leads to increasing distance from the origin as follows:

$$\begin{aligned}
\left\|(x + h_1, y + h_2)\right\|^2 &= x^2 - 2x\alpha(y + \alpha x(1 + p)) + \alpha^2\left(y^2 + \alpha^2 x^2(1 + p)^2 + 2\alpha xy(1 + p)\right) + \\
&\quad y^2 - 2y\alpha(-x + \alpha y(1 + p)) + \alpha^2\left(x^2 + \alpha^2 y^2(1 + p)^2 - 2\alpha xy(1 + p)\right) \\
&= \left(x^2 + y^2\right)\left(1 - \alpha^2(2p + 1) + \alpha^4(1 + p)^2\right) \\
&\geq \left(x^2 + y^2\right)\left(1 - \alpha^2 + \alpha^4\right) \coloneqq \left\|(x, y)\right\|^2 \lambda
\end{aligned}$$

where the inequality follows because the final expression in p has positive derivative for $\alpha > 1$, hence minimized at $p = 0$. Now $\lambda > 1$ for any $\alpha > 1$, so we conclude by induction that

$$\left\|(x_n, y_n)\right\|^2 \geq \lambda^n \left\|(x_0, y_0)\right\|^2 \to \infty$$

as $n \to \infty$, provided $(x_0, y_0) \neq 0$, as required.

Consistent solution converges. We begin by showing that the following linear functions satisfy the consistency equations for the Hamiltonian game:

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \frac{-\alpha}{1+2\alpha^2} \begin{pmatrix} y + 2\alpha x \\ -x + 2\alpha y \end{pmatrix} .$$

Indeed, the RHS of the first consistency equation is

$$\begin{aligned} -\alpha \nabla_x \left(x \left(y - \alpha \frac{-x + 2\alpha y}{1+2\alpha^2} \right) \right) &= \frac{-\alpha}{1+2\alpha^2} \Big(y(1+2\alpha^2) - \alpha(-x+2\alpha y) + \alpha x \Big) \\ &= \frac{-\alpha}{1+2\alpha^2} \Big(y + 2\alpha x \Big) = h_1 \end{aligned}$$

and similarly for the second equation.

To prove uniqueness, assume there is a second pair of linear update functions $\hat{h}_1, \hat{h_2}$ also satisfying consistency. Let $a, b, c \in \mathbb{R}$ such that $\hat{h}_1(x, y) = ax + by + c$. Note that substituting the second equation into the first yields

$$\begin{aligned} \hat{h}_1(x, y) &= -\alpha \nabla_x \left(L^1(x, y - \alpha \nabla_y \left(L^2(x + \hat{h}_1(x, y), y) \right)) \right) \\ &= -\alpha \left(y + \alpha \left(2x + \hat{h}_1(x, y) + x \nabla_x \hat{h}_1(x, y) + y \nabla_y \hat{h}_1(x, y) + xy \nabla_{xy} \hat{h}_1(x, y) \right) \right) \end{aligned}$$

Expanding the above and substituting the equation for $\hat{h}_1$, we obtain

$$ax + by + c = -2\alpha^2 x(1 + a) - \alpha y(1 + 2\alpha b) - \alpha^2 c$$

for all $x, y \in \mathbb{R}$, which yields (by comparing coefficients)

$$a(1 + 2\alpha^2) = -2\alpha^2 \quad ; \quad b(1 + 2\alpha^2) = -\alpha \quad ; \quad c(1 + \alpha^2) = 0.$$

It follows that

$$\hat{h}_1(x, y) = \frac{-\alpha}{1+2\alpha^2} \Big(y + 2\alpha x \Big) = h_1(x, y),$$

proving the uniqueness of h_1. Since h_2 is directly determined by h_1 via the second consistency equation, this concludes the proof.

Finally we prove that this linear update leads to decreasing distance from the origin as follows:

$$\begin{aligned} \|(x + h_1, y + h_2)\|^2 &= x^2 - \frac{2x\alpha}{1+2\alpha^2}(y + 2\alpha x) + \frac{\alpha^2}{(1+2\alpha^2)^2} \left(y^2 + 4\alpha^2 x^2 + 4\alpha xy \right) + \\ &\qquad y^2 - \frac{2y\alpha}{1+2\alpha^2} \alpha(-x + 2\alpha y) + \frac{\alpha^2}{(1+2\alpha^2)^2} \left(x^2 + 4\alpha^2 y^2 - 4\alpha xy \right) \\ &= \left(x^2 + y^2 \right) \left(1 - \frac{\alpha^2(3 + 4\alpha^2)}{(1+2\alpha^2)^2} \right) := \|(x, y)\|^2 \lambda \, . \end{aligned}$$

Notice that the derivative of λ is strictly negative in α while its limit as $\alpha \to \infty$ is 0, with value 1 at $\alpha = 0$, hence $|\lambda| = \lambda < 1$ for any $\alpha > 0$. We conclude by induction that

$$\|(x_n, y_n)\|^2 = \lambda^n \|(x_0, y_0)\|^2 \to 0$$

as $n \to \infty$, with λ decreasing (hence the speed of convergence increasing) as α increases. □

H Training Details COLA

All code was implemented using Python. The code relies on the PyTorch library for autodifferentiability (Paszke et al., 2019).

H.1 Quadratic and bilinear games

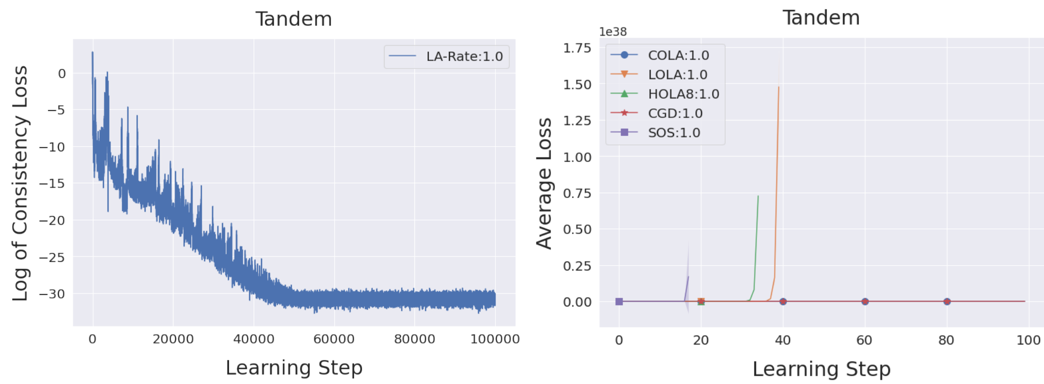
For the quadratic and bilinear games, COLA uses a neural network with 1 non-linear layer for both $h_1(\theta^1, \theta^2)$ and $h_2(\theta^1, \theta^2)$. The non-linearity is a ReLU function. The layer has 8 nodes. For training, we randomly sample pairs of parameters on a [-1, 1] parameter region. In general, the size of the region is a hyperparameter. We use a batch size of 8. We found that training is improved with a learning rate scheduler. For the learning rate scheduling with a γ of 0.9. We train the neural network for 120'000 steps. To compute the consistency loss we use the squared distance measure. The optimizer used is Adam (Kingma & Ba, 2017).

H.2 NON-QUADRATIC GAMES

For the non-quadratic games, we deploy a neural network with 3 non-linear layers using Tanh activation functions. Each layer has 16 nodes. For this type of game, the parameter region is set to [-7, 7], because the parameters will be squished into probability space, allowing us to explore the full probability space. During training, we used a batch size of 64. The optimizer used is Adam (Kingma & Ba, 2017).

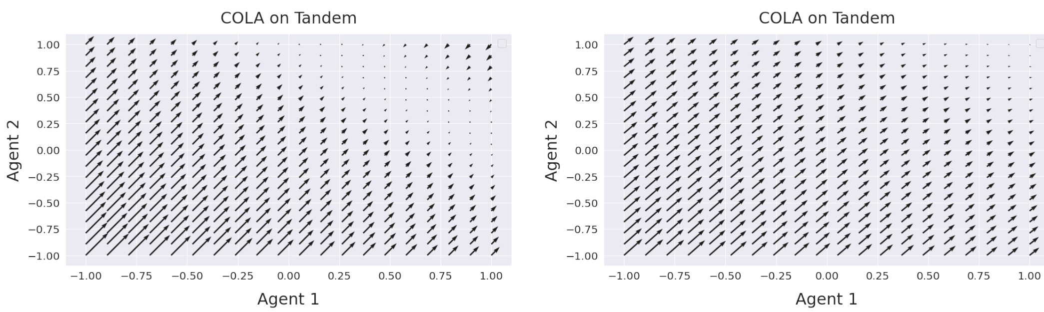
I Further experimental results

I.1 Tandem game

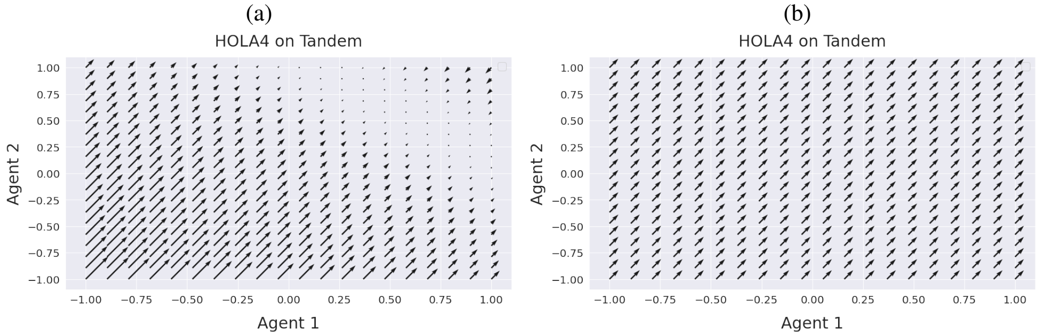


(a) (b)

Figure 4: Tandem game with a look-ahead rate of 1.0. The standard deviation for the initialization of parameters used here is 0.1, which is standard in the literature (Letcher et al., 2019b).



(a) (b)



(c) (d)

Figure 5: Gradients field of the Tandem game at two different look-ahead rates, 0.1 and 1.0.

I.2 BALDUZZI AND HAMILTONIAN GAME

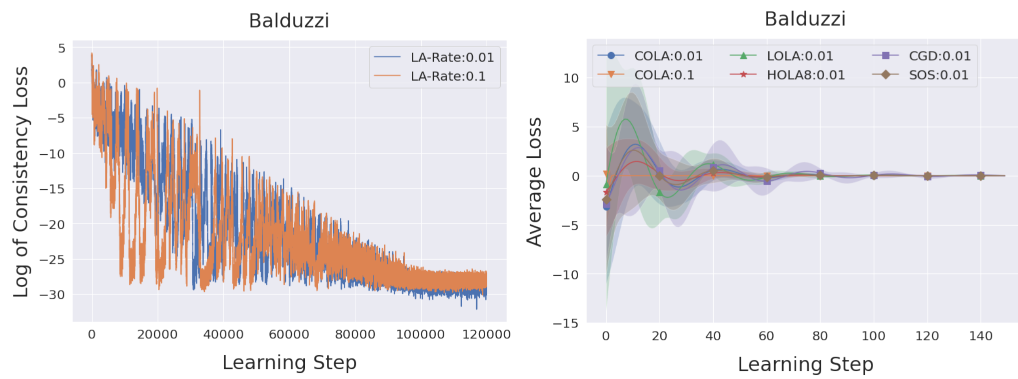
The Hamiltonian game was originally introduced in (Balduzzi et al., 2018) as a minimal example of Hamiltonian dynamics. Recall that its loss function is

$$L^1(x, y) = xy \quad \text{and} \quad L^2(x, y) = -xy \tag{32}$$

The Balduzzi game was introduced to investigate the behaviour of differentiable game algorithms when a weak attractor is coupled with strong rotational forces in the Hamiltonian dynamics (Balduzzi et al., 2018), captured by the losses

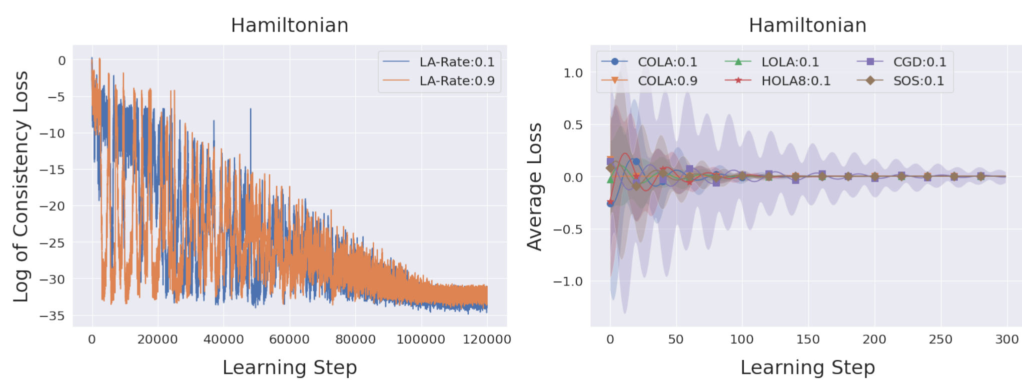
$$L^1(x, y) = \frac{1}{2}x^2 + 10xy \quad \text{and} \quad L^2(x, y) = \frac{1}{2}y^2 - 10xy \tag{33}$$

Results for both games are displayed in Figures 12 and 7.



(a) (b)

Figure 6: Balduzzi game with a look-ahead rate of 0.01 and 0.1. The standard deviation for the initialization of parameters used here is 1.0



(a) (b)

Figure 7: Hamiltonian game with a look-ahead rate of 0.1 and 0.9. The standard deviation for the initialization of parameters used here is 1.0

Table 4: (a) Comparison of consistency losses over multiple look-ahead rates on the Hamiltonian game. (b) Cosine similarity between COLA and LOLA, HOLA3 and HOLA6 over different look-ahead rates on the Hamiltonian game.

(a)

α	LOLA	HOLA3	HOLA6	COLA
0.9	6.99	4.63	13.68	9.08e-15
0.5	12.67	13.77	13.00	3.06e-15
0.4	5.78	7.14	6.38	2.20e-15
0.1	0.08	0.01	1.74e-6	1.32-16
0.05	1.63e-5	3.98e-10	3.08e-15	3.97e-17

(b)

α	LOLA	HOLA3	HOLA6
0.9	1.0	-1.0	0.4852
0.5	1.0	1.0	0.996
0.4	1.0	1.0	0.999
0.1	1.0	1.0	1.0
0.05	1.0	1.0	1.0

Table 5: Comparison of consistency losses over multiple look-ahead rates on the Balduzzi game.

α	LOLA	HOLA3	HOLA6	COLA
0.9	1.78e+6	4.65e+10	3.88e+17	5.09e-13
0.1	2.71e+2	1.09e+3	1.73e+4	6.04e-15
0.05	1.61e+1	4.01	1.03	4.13e-15
0.03	2.16	0.07	0.01	1.08e-15
0.01	0.03	1.08e-5	1.69e-10	1.32e-16

I.3 MATCHING PENNIES

Matching Pennies (MP) is a single-shot, zero-sum game, where two players, A and B, each flip a biased coin (Lee & K, 1967). Player A wins if the outcomes of both flips are the same and player B wins if they are different.

Table 6: Payoff Matrix for the Matching Pennies game.

	Head	Tail
Head	(+1, -1)	(-1, +1)
Tail	(-1, +1)	(+1, -1)

Table 7: Cosine similarities over multiple COLA training runs on the MP and Tandem game for different look-ahead rates.

Game@LR	Cosine Sim
MP@10	0.971 +/- 0.006
MP@0.5	0.998 ± 0.002
Tandem@0.1	1.0
Tandem@1.0	0.976 ±0.008

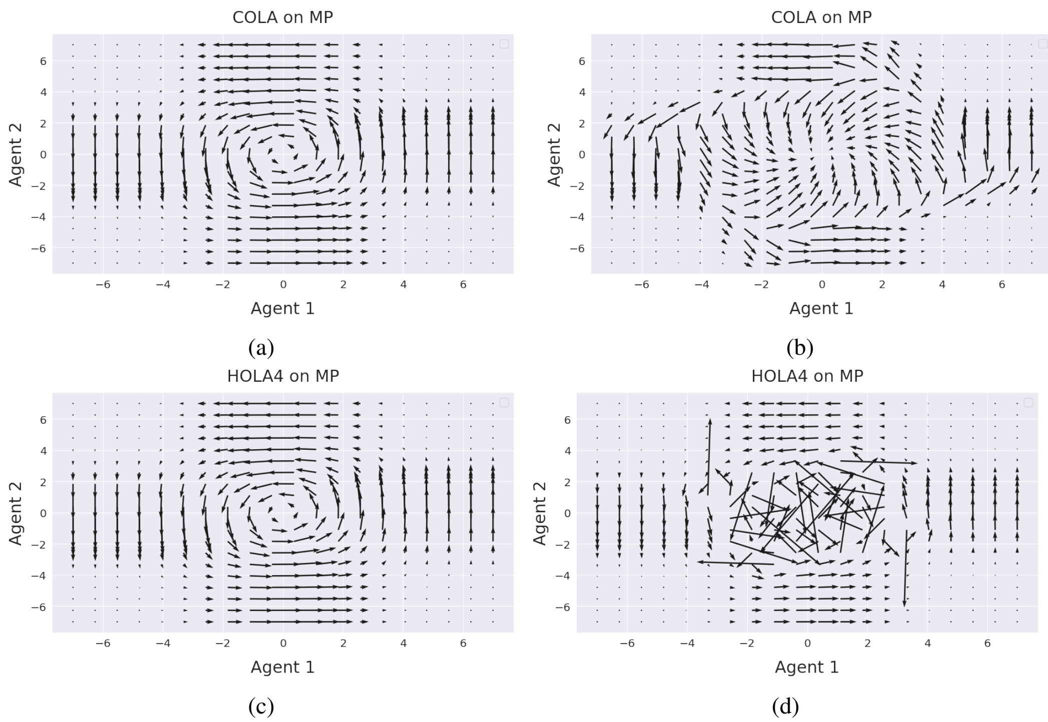


Figure 8: Gradients field of the MP game at two different look-ahead rates: 0.5 and 10. COLA is on the upper row, HOLA4 is on the lower row.

I.4 ULTIMATUM GAME

Table 8: (a) Comparison of consistency losses over multiple look-ahead rates on the Ultimatum game. (b) Cosine similarity between COLA and LOLA, HOLA2 and HOLA4 over different look-ahead rates on the Ultimatum game.

(a)

α	LOLA	HOLA2	HOLA4	COLA
1.1	2.16e-3	4.77e-3	0.01	6.23e-5
0.7	4.29e-4	1.91e-4	1.53e-4	3.12e-6
0.3	2.50e-5	1.97e-7	4.85e-8	1.41e-7
0.1	2.36e-7	1.29e-11	5.59e-13	1.98e-7
0.001	3.12e-15	4.20e-17	9.19e-17	2.58e-8

(b)

α	LOLA	HOLA2	HOLA4
1.1	0.941	0.943	0.92
0.7	0.99	0.99	0.99
0.3	0.99	0.99	0.99
0.1	0.99	0.99	0.99
0.001	0.99	0.99	0.99

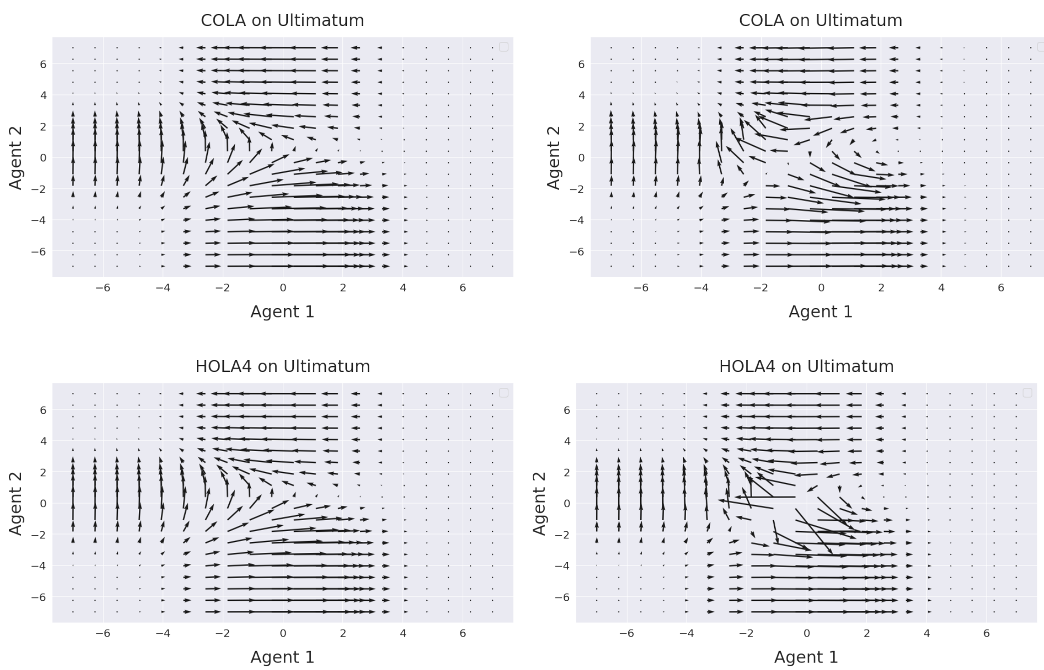


Figure 9: Gradients field of the ultimatum game at two different look-ahead rates: 0.2 and 1.1. COLA is on the upper row, HOLA4 is on the lower row.

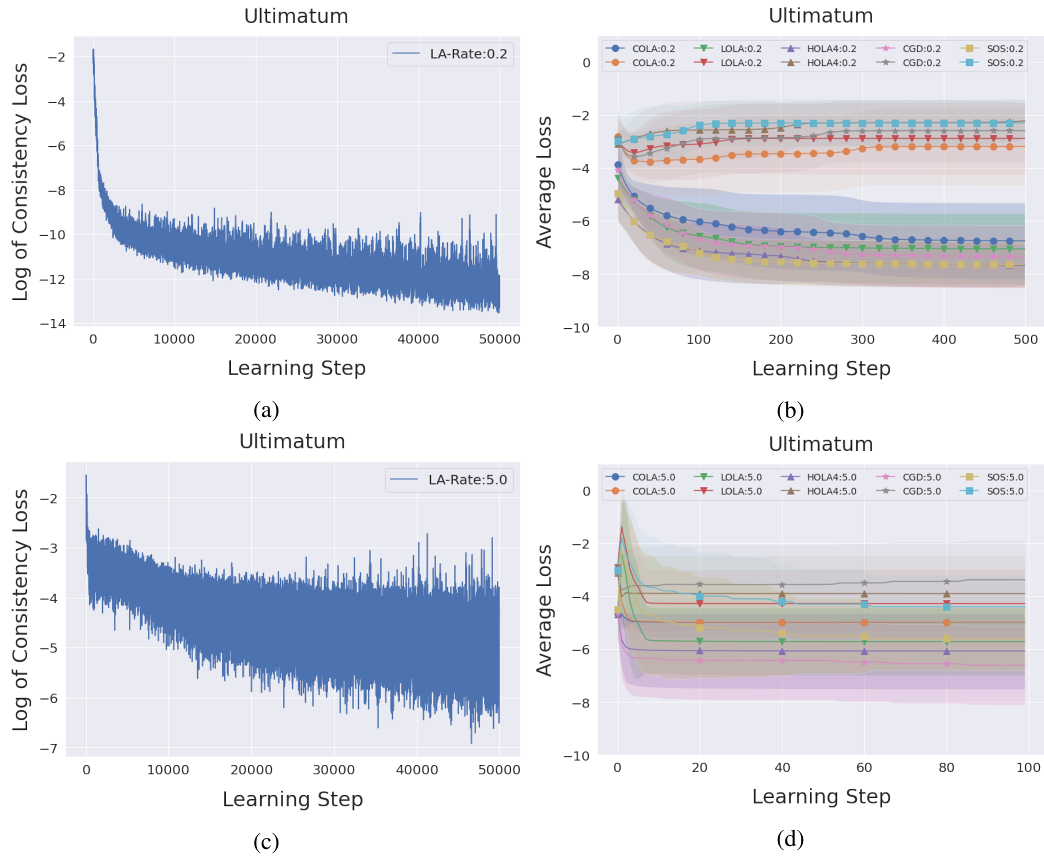


Figure 10: Ultimatum with a look-ahead rate of 0.2 and 5.0

I.5 Chicken game

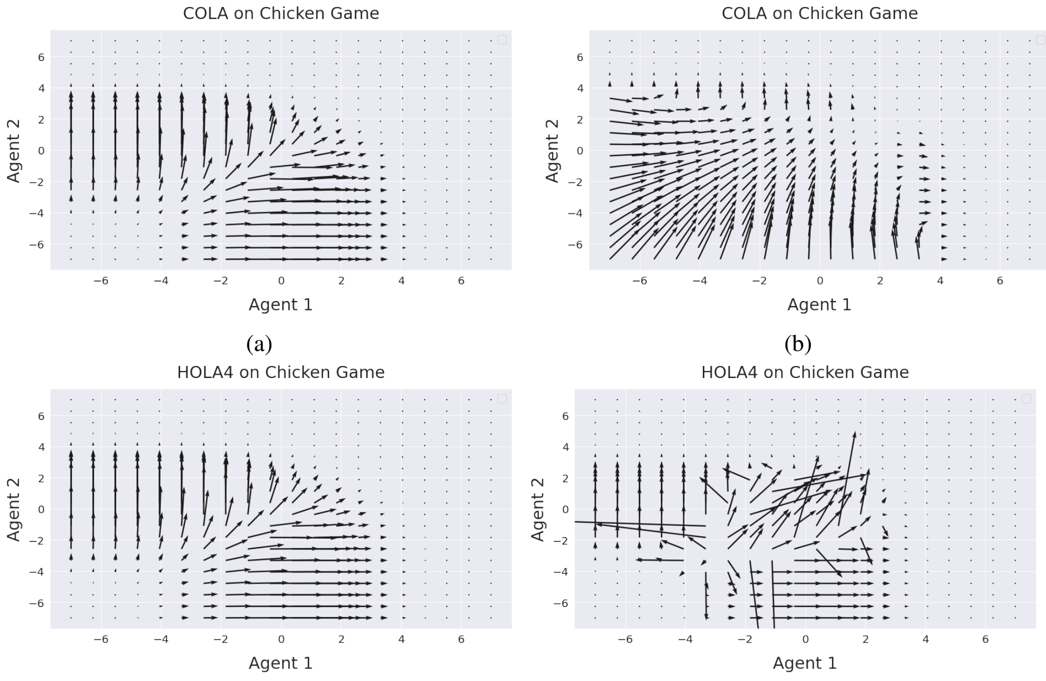
In the Chicken game, an agent can either choose to yield to avoid a catastrophic payoff but face a small punishment if they are the only agent to yield. Imagine a game where two agents drive towards each other in their cars. If both never swerve, they frontally crash into each other, an obviously catastrophic outcome. If any of the agents ”chicken out”, e.g. swerve, they do not crash but receive a small punishment for having chickened out. At the same time, the other agent is being rewarded for staying on track, as quantified in Table 9.

Table 9: Payoff Matrix for the Chicken game.

	C (swerve)	D (straight)
C (swerve)	0, 0	-1, +1
D (straight)	+1 ,-1	-100, -100

Table 10: (a) Comparison of consistency losses over multiple look-ahead rates on the chicken game.

α	LOLA	HOLA2	HOLA4	SOS	CGD	COLA
1.0	2429	3892	46637	1494	1677	5.02e-3
0.5	643	484	4320	475	2330	7.81e-3
0.1	11.99	7.69	73.28	2.73	8.46	0.75
0.05	0.84	0.17	0.47	0.37	1.31	0.06
0.01	9.97e-4	3.41e-6	5.77e-9	2.40e-4	0.04	9.83e-5



(a) (b) (c) (d)

Figure 11: Gradients field of the Chicken game at two different look-ahead rates, 0.01 and 1.0.

I.6 IPD

Table 11: Payoff Matrix for the IPD game.

	C	D
C	(-1, -1)	(0, -3)
D	(0, -3)	(-2, -2)

Table 12: Cosine similarity of LOLA, HOLA2 and HOLA4 with COLA on the IPD game.

α	LOLA	HOLA2	HOLA4
1.0	0.77	0.70	0.53
0.03	0.96	0.98	0.98

Table 13: Consistency losses of LOLA, HOLA2, HOLA4 and COLA on the IPD game at two different look-ahead rates.

α	LOLA	HOLA2	HOLA4	COLA
1.0	39.56	21.16	381.21	0.65
0.03	1.72e-3	4.72e-6	9.72e-8	0.33

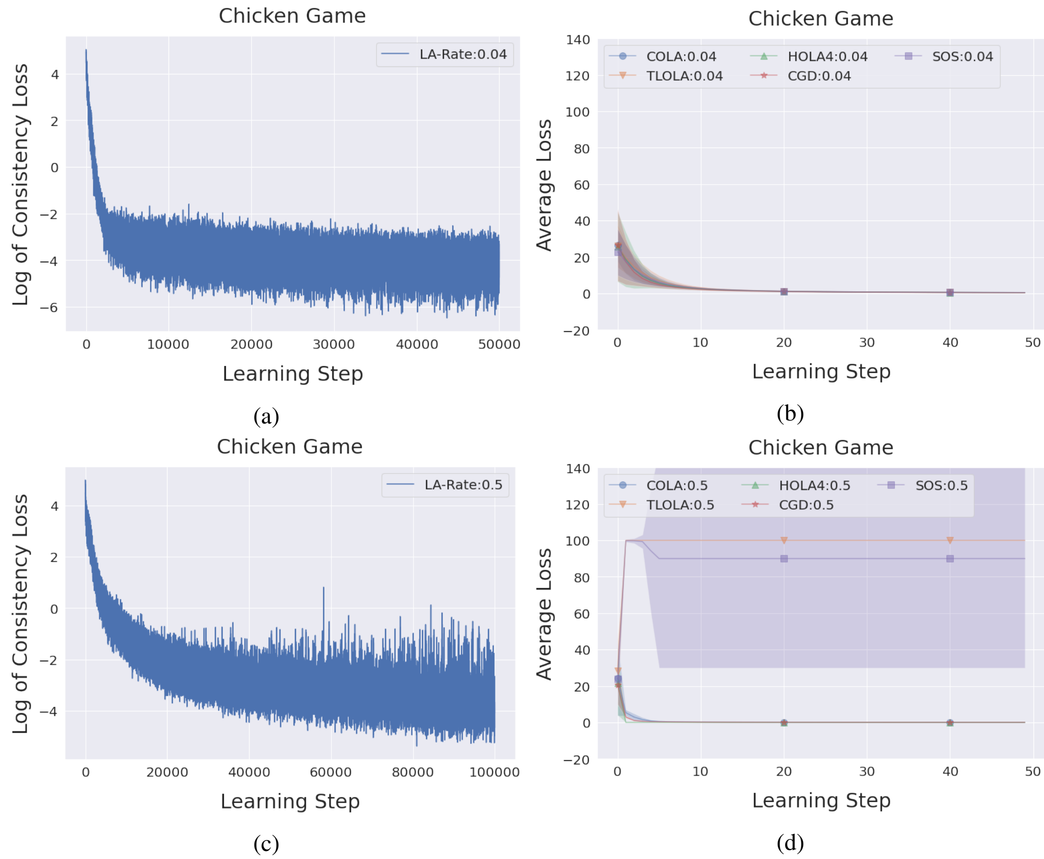


Figure 12: Chicken game with a look-ahead rate of 0.04 and 0.5, We used a standard deviation of 1.0 to initialize the parameters.

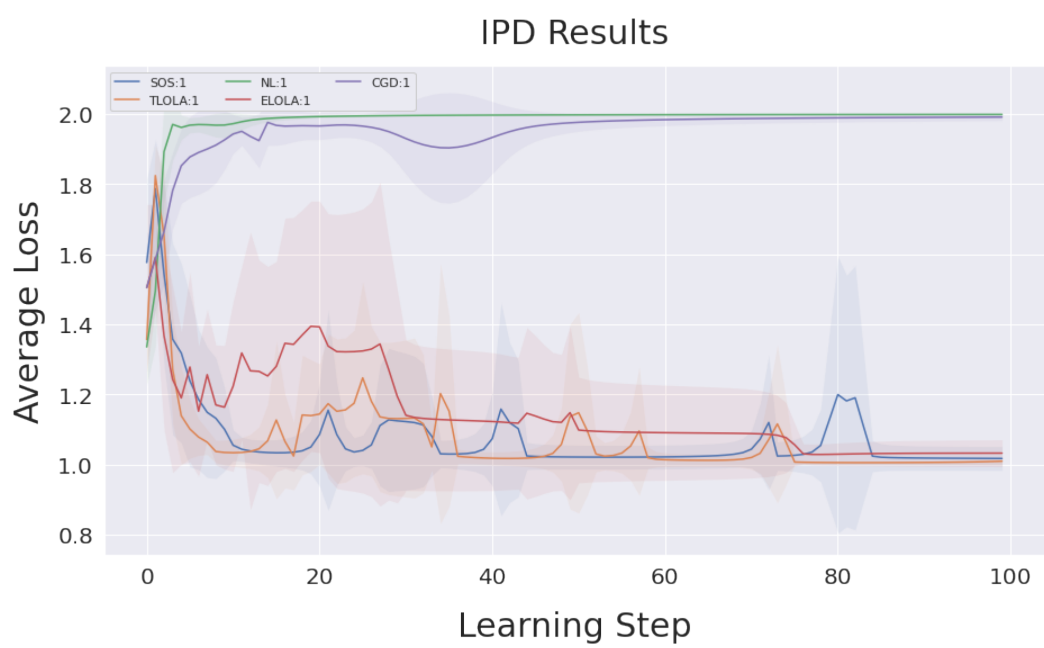


Figure 13: CGD, SOS, Taylor LOLA (TLOLA), Exact LOLA (ELOLA) and Naive Learning (NL) on the IPD at a look-ahead rate of 1.0. We used a standard deviation of 1.0 to initialize the parameters.