# Parametric Open Source Games

**Aleksandar Todorov**⋆ A.TODOROV.4@STUDENT.RUG.NL
**Jesse ten Napel** J.TEN.NAPEL@STUDENT.RUG.NL
**Alexander Müller** A.K.MULLER.2@STUDENT.RUG.NL
*University of Groningen, Groningen, the Netherlands*
*Safe AI Netherlands*

## Abstract

Open-source game theory studies agents whose behavior may depend on one another's decision procedures, but most existing models use discrete or symbolic programs. We introduce *parametric open-source games*, a continuous analogue of program equilibria in which players choose parameter vectors and semantics maps convert the full parameter profile into mixed actions in an underlying finite game. We establish equilibrium existence results, derive an exact coupling threshold at which selfish gradient ascent in symmetric $2 \times 2$ games switches from defection toward cooperation, and give a one-dimensional boundary test for parametric program Nash equilibria. We further extend the framework to a neural semantics class whose first-order cooperation condition is governed by the ratio of cross-player to self-player sensitivity. Across canonical games, the framework shows how access to internal parameterizations can qualitatively reshape learning dynamics and equilibrium structure, and how sufficiently strong open-source coupling can steer selfish optimization toward cooperative outcomes.

## 1. Introduction

Open-source game theory studies settings where agents are able to inspect one another's internal decision procedures. In the classical program equilibrium framework, players submit programs that may inspect the opponent's program before outputting an action in an underlying base game [7, 16]. This possibility can qualitatively change the strategic reasoning of agents and consequently cooperation, punishment, and coordination may become contingent on what an agent proves, simulates, or infers about its opponent's internal policy, yielding outcomes that differ sharply from the Nash equilibria of the underlying normal form game [5, 6, 10]. Most existing open-source models are nevertheless symbolic or proof-theoretic [1, 14], while modern learning systems are usually continuous, differentiable, and parameterized by high-dimensional vectors [13, 15, 17]. Conversely, differentiable games and multi-agent optimization study smooth parameterized objectives, but usually treat each player's policy as depending only on its own parameters [2, 4, 8, 11, 12].

In this work, we develop a parametric model that preserves the central open-source feature, namely that a player's behavior may depend on an opponent's internal description, while enabling standard tools from continuous optimization. In a *parametric open-source game*, each player chooses a parameter vector, and a continuous semantics map converts the full parameter profile into mixed actions in an underlying finite game. The induced game is ordinary and continuous over parameter space, but it retains the open-source feature that behavior can depend on opponents' internal descriptions, providing a connection between program-style transparency and gradient-based learning.

© A. Todorov⋆, J.t. Napel & A. Müller.

Our contributions are threefold. We formalize parametric open-source games and give basic equilibrium existence results for the induced parametric game. We then analyze a sigmoid semantics in symmetric $2 \times 2$ games, deriving a first-order cooperation threshold and a boundary PPNE test, both supported by empirical verification. Finally, we introduce a neural semantics class that preserves the same first-order criterion through the ratio of cross-player to self-player sensitivity. The framework suggests that open-source strategic reasoning need not be tied to symbolic representations, as once encoded by a continuous semantics map, it can reshape both learning dynamics and equilibrium structure in ways that are analyzable with standard tools from continuous optimization.

## 2. Parametric Open-Source Games

Proofs of formal statements are deferred to Appendix A. Let $G = (N, (S_i)_{i\in N}, (u_i)_{i\in N})$ be a finite normal form game, where $N = \{1, \dots, n\}$ denotes a set of $n$ agents with each agent $i \in N$ having a finite action set $S_i$ and utility $u_i : S \to \mathbb{R}$ with $S = \prod_i S_i$. For a mixed profile $\sigma \in \prod_i \Delta(S_i)$, we write the (continuous) multilinear extension of $u_i$ as $u_i(\sigma) = \sum_{s\in S}\Big(\prod_{j\in N} \sigma_j(s_j)\Big)u_i(s)$.

Each player $i$ chooses a parameter vector in a nonempty compact convex set $\Theta_i \subset \mathbb{R}^{d_i}$, and we denote by $\Theta = \prod_i \Theta_i$ and $\theta = (\theta_i)_{i\in N}$. Next, we state the two central definitions of our model.

**Definition 1 (Parametric open-source game)** *A semantics for player $i$ is a continuous map $\phi_i : \Theta \to \Delta(S_i)$. Given semantics $\phi = (\phi_i)_i$, the induced payoff is*

$$U_i(\theta) = u_i(\phi_1(\theta), \dots, \phi_n(\theta)).$$

*The induced parametric game is then $\bar{G} = (N, (\Theta_i)_{i\in N}, (U_i)_{i\in N})$. It is closed-source if every $\phi_i$ depends only on $\theta_i$, and open-source otherwise.*

**Definition 2 (Parametric program Nash equilibrium)** *A profile $\theta^\star \in \Theta$ is a parametric program Nash equilibrium (PPNE) if, for every $i$ and every $\theta_i \in \Theta_i$,*

$$U_i(\theta_i^\star, \theta_{-i}^\star) \geq U_i(\theta_i, \theta_{-i}^\star).$$

Since each $U_i$ is continuous on compact $\Theta$, Glicksberg's theorem gives a mixed-strategy Nash equilibrium over $\Theta$. If $\theta_i \mapsto U_i(\theta_i, \theta_{-i})$ is quasiconcave on $\Theta_i$ for each fixed $\theta_{-i}$, Kakutani's theorem also gives a pure PPNE. We prove this in Appendix A.1. We next specialize the framework to two-action games, where the effect of open-source dependence can be analyzed explicitly.

## 3. Semantics Families and Game Dynamics

We now instantiate the framework with semantics maps that are simple enough to analyze explicitly while still exhibiting open-source effects. Experimental details are deferred to Appendix B.

### 3.1. Sigmoid Semantics

We adopt a two-action setting $S_i = \{C, D\}$ and use a one-dimensional parameter $\theta_i \in [-B, B] \subseteq \mathbb{R}$. The sigmoid semantics determines the probability of playing $C$ (cooperation) as

$$p_i(\theta) = \phi_i(\theta)(C) = \sigma(\theta_i + \gamma\theta_{-i}), \qquad \sigma(x) = (1 + e^{-x})^{-1}.$$

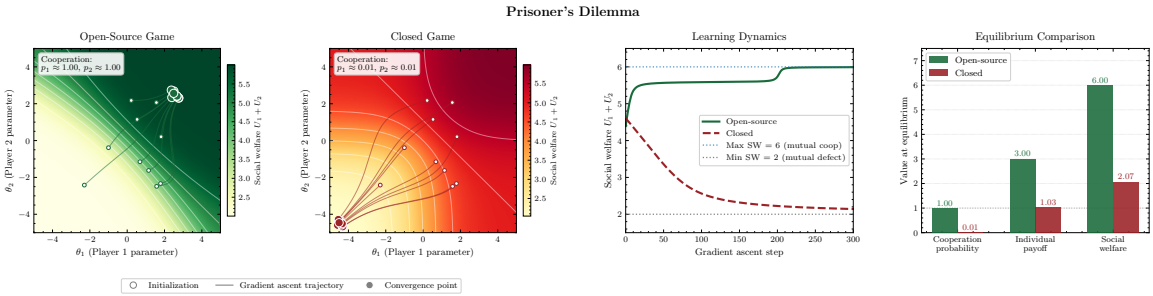


Figure 1: Open-source versus closed-source sigmoid semantics in the Prisoner's Dilemma. Under open-source semantics, projected gradient-ascent trajectories move toward the cooperative corner and high welfare, whereas closed-source dynamics move toward defection.

The coupling $\gamma \geq 0$ measures how strongly player $i$'s behavior responds to the opponent's parameter, while $\gamma = 0$ recovers the closed-source parameterization.

As a base experiment, we study selfish projected gradient ascent in the induced parametric game,

$$\theta_i^{t+1} = \Pi_{[-B,B]}\left(\theta_i^t + \eta\, \partial_{\theta_i} U_i(\theta^t)\right), \qquad i \in \{1,2\},\ \eta > 0,$$

using finite-difference gradients of $U_i$ with respect to each player's own parameter. The underlying interaction remains a one-shot symmetric $2 \times 2$ game.

The trajectories in Figure 1 visualize how the semantics map changes and which regions of parameter space are attractive under this learning rule. In the Prisoner's Dilemma, closed-source dynamics move toward defection, while sufficiently coupled open-source dynamics move toward the cooperative boundary. Nonetheless, the cooperation results and basin of attraction strongly depend on the choice of $\gamma$. The following theorem provides the exact conditions under which, for a general symmetric $2 \times 2$ game with payoffs $R$ for mutual cooperation, $S$ for being exploited, $T$ for exploiting, and $P$ for mutual defection and an arbitrary $\gamma \geq 0$, the cooperative regime first emerges under gradient ascent as a local property.

**Theorem 3 (Phase transition)** *Assume $R + T - P - S > 0$ and define*

$$\gamma^\star = \frac{T + P - R - S}{R + T - P - S}.$$

*For initializations near $\theta = (0,0)$, selfish gradient ascent points toward lower cooperation when $\gamma < \gamma^\star$ and toward higher cooperation when $\gamma > \gamma^\star$.*

Thus $\gamma^\star$ is the local open-source coupling needed to reverse the initial incentive of gradient ascent. For Stag Hunt, $\gamma^\star \leq 0$, so cooperation is already locally attractive. We verify Theorem 3 empirically across four canonical games, and Figure 2 shows that the empirical transition in mean cooperation matches the analytical value of $\gamma^\star$ closely.

### 3.2. Boundary Equilibria

The result from Theorem 3 describes only the local direction of gradient ascent near the symmetric midpoint. It therefore explains when learning initially moves toward cooperation, but not whether

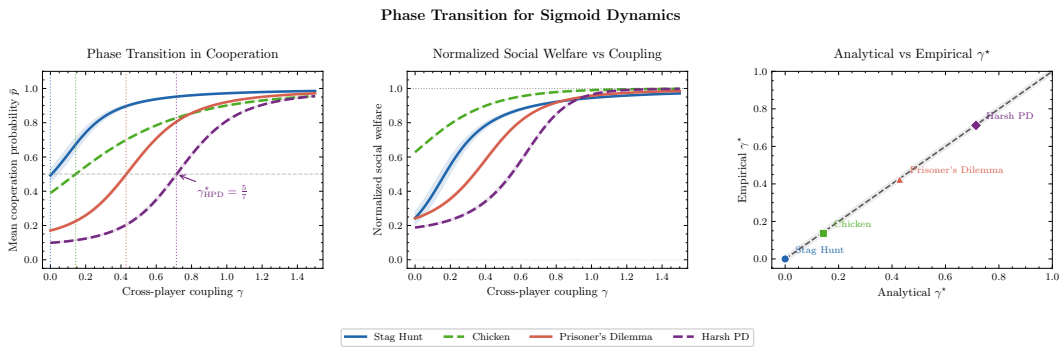


Figure 2: Phase transition in $2\times2$ games. Panel (a) shows mean terminal cooperation as a function of $\gamma$, with dotted lines marking $\gamma^\star$ and the dashed line marking $\bar{p} = 0.5$. Panel (b) shows normalized social welfare, and Panel (c) compares analytical and empirical transition points. All results are averaged over 20 seeds.

the limiting boundary point is stable against unilateral deviations. This distinction matters because projected gradient ascent in the sigmoid model often converges to the boundary of $\Theta = [-B, B]^2$. Hence, the natural equilibrium candidates are boundary points such as $(B, B)$ and $(-B, -B)$.

**Theorem 4 (Boundary PPNE characterization)** *For fixed $\gamma \geq 0$, suppose that $\theta_i \mapsto U_i(\theta_i, B)$ is differentiable on $(-B, B)$ and satisfies*

$$\frac{d}{d\theta_i} U_i(\theta_i, B) \geq 0 \qquad \textit{for all } \theta_i \in (-B, B) \textit{ and } i \in \{1, 2\}.$$

*Then $(B, B)$ is a PPNE.*

Theorem 4 shows that equilibrium verification does not require analyzing the full two-dimensional landscape. For each player, we freeze the opponent at the candidate boundary value and maximize over the player's own parameter interval. Figure 3 applies this test in the Prisoner's Dilemma. Under open-source coupling, the unilateral payoff is maximized at the cooperative boundary; under closed-source semantics, it is maximized at the defective boundary. This verifies the limit points observed in Figure 1 as PPNEs. A weaker-coupling example in which open-source semantics does not yield a PPNE is given in Appendix B.3 and Figure 5.

### 3.3. Neural Open-Source Semantics

The sigmoid semantics isolates open-source dependence through a single scalar $\gamma$, which makes the threshold analysis transparent. However, it also hard-codes a linear dependence on the opponent's parameter. To allow richer semantics while preserving the same local interpretation, we consider

$$p_i(\theta) = \sigma\big(\alpha\theta_i + \beta\theta_{-i} + h(\theta_i^2, \theta_{-i}^2, \theta_i\theta_{-i}; W)\big)\,, \qquad h(x; W) = W_2^\top \tanh(W_1^\top x + b_1),$$

where $K \in \mathbb{N}$ is the network hidden width and $W = (W_1, b_1, W_2)$, where $W_1 \in \mathbb{R}^{3\times K}$, $b_1 \in \mathbb{R}^K$, $W_2 \in \mathbb{R}^K$. The (learnable) scalars $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ have a direct interpretation, namely that $\alpha$ measures first-order sensitivity to the player's own parameter, while $\beta$ measures first-order

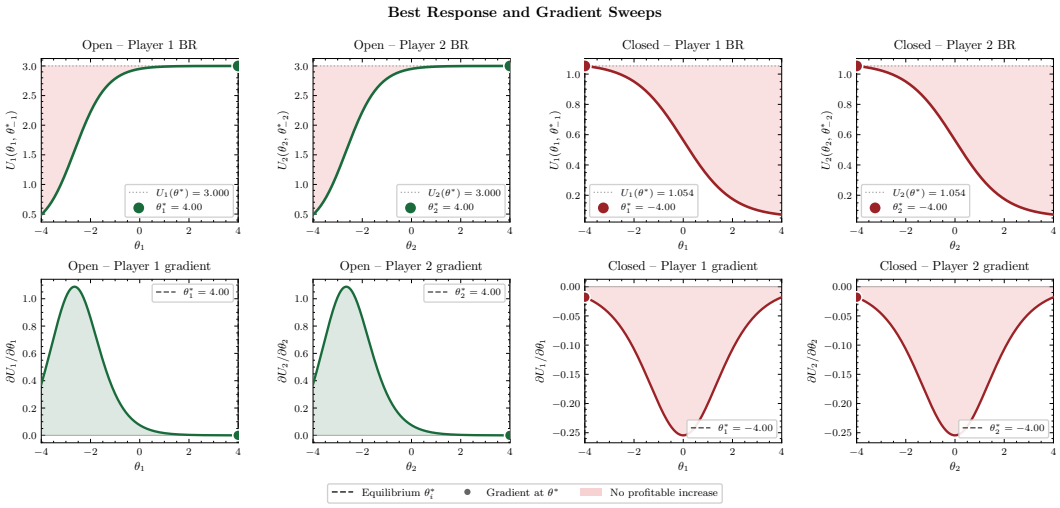


Figure 3: Boundary PPNE verification in the Prisoner's Dilemma. For each candidate equilibrium, the opponent is fixed at its boundary value, and the remaining player's payoff is evaluated over $\theta_i \in [-B, B]$. Open-source semantics maximizes payoff at the cooperative boundary, while closed-source semantics maximizes payoff at the defective boundary.

sensitivity to the opponent's parameter. The residual network $h$ adds nonlinear dependence on the full parameter profile, but it receives only quadratic features. Therefore, its first-order derivative vanishes at $\theta = (0, 0)$, so it cannot change the local incentive condition at the midpoint. This constraint ensures the model is more expressive away from the midpoint, but its first-order open-source coupling remains identifiable.

Under this constraint, the sigmoid threshold from Theorem 3 extends cleanly. If $\alpha > 0$, then near $\theta = (0, 0)$ the relevant coupling is no longer $\gamma$ itself, but the sensitivity ratio $\beta/\alpha$. In particular, selfish gradient ascent initially points toward cooperation exactly when $\frac{\beta}{\alpha} > \gamma^\star$ with $\gamma^\star$ as in Theorem 3. The formal statement and proof are given in Appendix A.4.

Figure 4 tests whether this first-order condition continues to organize learning once the semantics are neural. We compare sigmoid closed-source semantics, sigmoid open-source semantics, fixed neural semantics with $\beta = 0$, fixed neural semantics with $\beta/\alpha = 1.5$, and two jointly learned neural semantics. In the learned conditions, both the player parameters and the neural semantics parameters are optimized. The warm-start condition initializes $(\alpha, \beta) = (1, 1.5)$, already above the cooperative threshold in the relevant games, whereas the cold-start condition initializes $(\alpha, \beta) = (0, 0)$.

The results show that fixed neural semantics reproduce the sigmoid baselines when their first-order ratios match. Neural-fixed-closed tracks sigmoid-closed, and neural-fixed-open tracks sigmoid-open. This supports the interpretation of $\beta/\alpha$ as the effective local open-source coupling. The learned conditions show a stronger point: warm-started neural semantics reach the same high-welfare regime as the fixed open-source models, but cold-started neural semantics do not reliably discover it. Thus, the threshold is not merely a representational condition, but also acts as an optimization barrier: the model class can represent cooperative open-source dependence, but gradient-based learning may fail to find it unless the first-order coupling is initialized in the right regime.

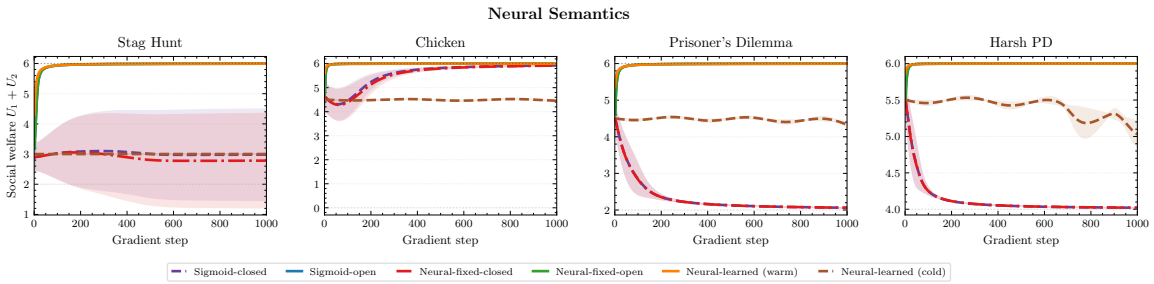


Figure 4: Learning curves under neural open-source semantics across canonical $2 \times 2$ games. Fixed neural semantics match the corresponding sigmoid baselines when the first-order ratio $\beta/\alpha$ is matched. Jointly learned neural semantics reach high welfare from a warm initialization above the threshold, but cold starts do not reliably discover cooperative coupling. Shaded regions indicate one standard deviation across seeds.

## 4. Discussion and Conclusion

We introduced parametric open-source games as a continuous model of strategic interaction in which each agent chooses a parameter vector and a semantics map converts the full parameter profile into mixed actions in an underlying finite game. The model preserves the central mechanism of open-source reasoning [16], namely that strategic behavior may be conditioned on another agent's internal description, while making the resulting game amenable to tools from continuous optimization and differentiable games.

Within this framework, our results show that open-source dependence can reshape both learning dynamics and equilibrium structure, analogous to the program equilibrium setting [7, 10, 16]. The sigmoidal and neural families of open-source semantics studied within this work both exhibit the same organizing principle: cooperation becomes locally attractive when cross-player sensitivity is sufficiently large relative to self-sensitivity. For the sigmoid model, this yields an explicit critical coupling, while for the neural model, the same condition is preserved through the first-order ratio between cross-player and self-player sensitivity. We then separate this local dynamical analysis from global equilibrium verification by using boundary best-response sweeps, which test whether a candidate limit point is stable against unilateral deviations.

The presented models are intentionally idealized. They assume full access to opponent parameters, focus on symmetric two-player examples, and study one-shot base games. Nevertheless, they provide an interpretable starting point for studying how transparency over internal representations can affect incentives, learning dynamics, and equilibrium structure in parametric multi-agent systems. Future work should relax these assumptions by considering partial or noisy transparency, certifiable properties instead of full parameter disclosure, asymmetric semantics, larger populations, and sequential environments such as Markov games. These extensions would connect the framework more directly to realistic multi-agent systems, where agents may observe only incomplete but strategically relevant information about one another's internal structure. Overall, parametric open-source games suggest that internal parameter access is not a superficial modeling detail, but it can change the incentives created by gradient-based learning, while introducing new robustness questions about how such access is represented, optimized, and verified.

## Acknowledgement

We thank Davide Grossi at the University of Groningen for introducing us to the field of computational game theory and for his encouragement to pursue a publication.

## References

[1] Mihaly Barasz, Paul Christiano, Benja Fallenstein, Marcello Herreshoff, Patrick LaVictoire, and Eliezer Yudkowsky. Robust Cooperation in the Prisoner's Dilemma: Program Equilibrium via Provability Logic, April 2021. URL http://arxiv.org/abs/1401.5577. arXiv:1401.5577 [cs].

[2] Tamer Basar and Geert Jan Olsder. *Dynamic Noncooperative Game Theory*. Society for Industrial and Applied Mathematics, 1999. ISBN 978-0-89871-429-6. Google-Books-ID: k1oF5AxmJlYC.

[3] Claude Berge. *Topological Spaces: Including a Treatment of Multi-valued Functions, Vector Spaces and Convexity*. Oliver & Boyd, 1877. Google-Books-ID: jtAYAQAAIAAJ.

[4] Benjamin Chasnov, Lillian Ratliff, Eric Mazumdar, and Samuel Burden. Convergence Analysis of Gradient-Based Learning in Continuous Games. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, pages 935–944. PMLR, August 2020. URL https://proceedings.mlr.press/v115/chasnov20a.html.

[5] Andrew Critch. A PARAMETRIC, RESOURCE-BOUNDED GENERALIZATION OF LÖB'S THEOREM, AND A ROBUST COOPERATION CRITERION FOR OPEN-SOURCE GAME THEORY. *The Journal of Symbolic Logic*, 84(4):1368–1381, December 2019. ISSN 0022-4812, 1943-5886. doi: 10.1017/jsl.2017.42.

[6] Andrew Critch, Michael Dennis, and Stuart Russell. Cooperative and uncooperative institution designs: Surprises and problems in open-source game theory, August 2022. URL http://arxiv.org/abs/2208.07006. arXiv:2208.07006 [cs].

[7] Lance Fortnow. Program equilibria and discounted computation time. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 128–133, California, July 2009. ACM. ISBN 978-1-60558-560-4. doi: 10.1145/1562814.1562833. URL https://dl.acm.org/doi/10.1145/1562814.1562833.

[8] Drew Fudenberg and David K. Levine. *The Theory of Learning in Games*. MIT Press, 1998. ISBN 978-0-262-06194-0. Google-Books-ID: G6vTQFluxuEC.

[9] I. L. Glicksberg. A Further Generalization of the Kakutani Fixed Point Theorem, with Application to Nash Equilibrium Points. *Proceedings of the American Mathematical Society*, 3(1): 170–174, 1952. ISSN 0002-9939. doi: 10.2307/2032478. URL https://www.jstor.org/stable/2032478.

[10] Patrick LaVictoire, Benja Fallenstein, Eliezer Yudkowsky, Mihaly Barasz, Paul Christiano, and Marcello Herreshoff. Program equilibrium in the prisoner's dilemma via löb's theorem. In *AAAI workshop on multiagent interaction without prior coordination*, 2014.

[11] Tianyi Lin, Zhengyuan Zhou, Panayotis Mertikopoulos, and Michael Jordan. Finite-Time Last-Iterate Convergence for Multi-Agent Learning in Games. In *Proceedings of the 37th International Conference on Machine Learning*, pages 6161–6171. PMLR, November 2020. URL https://proceedings.mlr.press/v119/lin20h.html.

[12] Eric Mazumdar, Lillian J. Ratliff, and S. Shankar Sastry. On Gradient-Based Learning in Continuous Games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, January 2020. ISSN 2577-0187. doi: 10.1137/18M1231298. URL http://arxiv.org/abs/1804.05464. arXiv:1804.05464 [cs].

[13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015. ISSN 1476-4687. doi: 10.1038/nature14236. URL https://www.nature.com/articles/nature14236.

[14] Caspar Oesterheld. Robust program equilibrium. *Theory and Decision*, 86(1):143–159, February 2019. ISSN 1573-7187. doi: 10.1007/s11238-018-9679-3. URL https://doi.org/10.1007/s11238-018-9679-3.

[15] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, October 2017. ISSN 1476-4687. doi: 10.1038/nature24270. URL https://www.nature.com/articles/nature24270.

[16] Moshe Tennenholtz. Program equilibrium. *Games and Economic Behavior*, 49(2):363–373, November 2004. ISSN 0899-8256. doi: 10.1016/j.geb.2004.02.002. URL https://www.sciencedirect.com/science/article/pii/S0899825604000314.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

## Appendix A. Proofs of Statements

### A.1. Existence of PPNE

**Theorem 5 (Existence of PPNE)** *For any parametric open-source game $G$, a mixed-strategy Nash equilibrium over $\Theta$ exists. If additionally each map $\theta_i \mapsto U_i(\theta_i, \theta_{-i})$ is quasiconcave on $\Theta_i$ for every fixed $\theta_{-i}$, then a pure-strategy PPNE exists.*

We prove the two existence claims of Theorem 5 separately.

- For the existence of a mixed-strategy equilibrium, observe first that each $\Theta_i$ is a nonempty compact subset of a Euclidean space, hence a compact metric space. Moreover, by construction of the induced game, each $U_i$ is continuous on the product space $\Theta = \prod_{j \in N} \Theta_j$. Therefore, the induced parametric game is an $n$-player game with compact metric strategy spaces and continuous payoffs. By Glicksberg's generalization of Kakutani's fixed point theorem [9], such a game admits a mixed-strategy Nash equilibrium. Hence $G$ has a mixed equilibrium over $\Theta$.

- We now prove the pure-strategy claim under the additional quasiconcavity assumption. For each player $i$, define the best-response correspondence

$$B_i(\theta_{-i}) = \arg\max_{\theta_i \in \Theta_i} U_i(\theta_i, \theta_{-i}).$$

  Since $\Theta_i$ is compact and $U_i(\cdot, \theta_{-i})$ is continuous for each fixed $\theta_{-i}$, the maximum is attained. Thus $B_i(\theta_{-i})$ is nonempty for every $\theta_{-i}$. Because $\Theta_i$ is compact and $U_i$ is continuous, Berge's maximum theorem [3] implies that $B_i$ has closed graph and is upper hemicontinuous. Since $U_i(\cdot, \theta_{-i})$ is quasiconcave on the convex set $\Theta_i$, each upper contour set $\{\theta_i \in \Theta_i : U_i(\theta_i, \theta_{-i}) \geq c\}$ is convex, and therefore the argmax set $B_i(\theta_{-i})$ is convex. Hence, for every $i$, the operator $B_i : \Theta_{-i} \rightrightarrows \Theta_i$ has nonempty, convex, compact values and is upper hemicontinuous.

  Define the joint best-response correspondence as $B(\theta) = \prod_{i \in N} B_i(\theta_{-i})$ for $\theta \in \Theta$. Since $\Theta = \prod_{i \in N} \Theta_i$ is nonempty, compact, and convex, and since each $B_i$ is upper hemicontinuous with nonempty, convex, compact values, the $B$ has the same properties. Kakutani's fixed point theorem therefore yields some $\theta^\star \in \Theta$ such that $\theta^\star \in B(\theta^\star)$. Equivalently, for every player $i$, $\theta_i^\star \in B_i(\theta_{-i}^\star)$, which means $U_i(\theta_i^\star, \theta_{-i}^\star) \geq U_i(\theta_i, \theta_{-i}^\star)$ for all $\theta_i \in \Theta_i$. Thus $\theta^\star$ is a pure-strategy Nash equilibrium of the induced parametric game.

### A.2. Phase Transition for Sigmoidal Dynamics

We prove a slightly more general version as presented in Theorem 3. For the symmetric $2 \times 2$ base game with payoffs $R, S, T, P$, player 1's expected payoff as a function of cooperation probabilities $p_1, p_2$ is

$$U_1(p_1, p_2) = p_1 p_2 R + p_1(1 - p_2)S + (1 - p_1)p_2 T + (1 - p_1)(1 - p_2)P.$$

Hence

$$\frac{\partial U_1}{\partial p_1} = p_2(R-T) + (1-p_2)(S-P),$$
$$\frac{\partial U_1}{\partial p_2} = p_1(R-S) + (1-p_1)(T-P).$$

Under open-source sigmoid semantics

$$p_i(\theta) = \sigma(\theta_i + \gamma\theta_{-i}), \qquad \sigma'(x) = \sigma(x)(1-\sigma(x)),$$

we have

$$\frac{\partial p_1}{\partial \theta_1} = p_1(1-p_1),$$
$$\frac{\partial p_2}{\partial \theta_1} = \gamma\, p_2(1-p_2).$$

Applying the chain rule gives

$$\frac{\partial U_1}{\partial \theta_1} = \big[p_2(R-T)+(1-p_2)(S-P)\big]\, p_1(1-p_1) + \big[p_1(R-S)+(1-p_1)(T-P)\big]\, \gamma\, p_2(1-p_2). \quad (1)$$

This expression separates the direct closed-source incentive from the open-source cross-player incentive. Interior critical points satisfy $\partial_{\theta_1} U_1 = 0$, which yields the threshold coupling

$$\gamma = \frac{p_1(1-p_1)}{p_2(1-p_2)} \cdot \frac{p_2(T-R) + (1-p_2)(P-S)}{p_1(R-S) + (1-p_1)(T-P)}.$$

In the symmetric case $p_1 = p_2 = p$, this reduces to

$$\gamma(p) = \frac{p(T-R) + (1-p)(P-S)}{p(R-S) + (1-p)(T-P)}. \quad (2)$$

The threshold in Theorem 3 is the special case corresponding to the symmetric midpoint $\theta = (0,0)$, for which $p_1 = p_2 = \frac{1}{2}$. Substituting $p = \frac{1}{2}$ into (2) gives

$$\gamma^\star = \gamma\left(\tfrac{1}{2}\right) = \frac{T+P-R-S}{R+T-P-S}.$$

Equivalently, evaluating (1) at $p_1 = p_2 = \frac{1}{2}$ yields

$$\frac{\partial U_1}{\partial \theta_1}\Big|_{\theta=(0,0)} = \frac{1}{8}\Big[(R+S-T-P) + \gamma(R+T-S-P)\Big],$$

so under the assumption $R+T-P-S > 0$,

$$\operatorname{sgn}\left(\frac{\partial U_1}{\partial \theta_1}(0,0)\right) = \operatorname{sgn}\big(\gamma - \gamma^\star\big).$$

For completeness, the limiting threshold values near full defection and full cooperation are

$$\gamma_{\text{def}} = \lim_{p\to 0} \gamma(p) = \frac{P-S}{T-P},$$
$$\gamma_{\text{coop}} = \lim_{p\to 1} \gamma(p) = \frac{T-R}{R-S}.$$

Moreover,

$$\frac{d\gamma}{dp} = \frac{(T+S-R-P)(T-S)}{\big[p(R-S)+(1-p)(T-P)\big]^2},$$

so the monotonicity of $\gamma(p)$ is determined by the sign of $T+S-R-P$. The learning dynamics claim follows from continuity of the gradient field, as near $\theta = (0,0)$, the gradient of each player's payoff with respect to their own parameter has the same sign as $\gamma - \gamma^\star$ (by symmetry), so simultaneous gradient ascent either pushes both parameters toward $+B$ or toward $-B$.

In other words, under open-source sigmoid semantics with coupling $\gamma \geq 0$, if $\gamma < \gamma^\star$ and the initialization lies in a neighborhood of $\theta = (0,0)$, each player has a local incentive to move in the direction of less cooperation, and learning dynamics starting near this midpoint drift toward defection. Conversely, if $\gamma > \gamma^\star$, each player has a local incentive to move in the direction of higher-cooperation parameter values, and learning dynamics starting near this midpoint drift toward cooperation.

### A.3. Boundary Characterizations of PPNEs

The full statement underlying Theorem 4 is that for a fixed $\gamma \geq 0$, the following are equivalent:

(i) $(B,B)$ is a PPNE.

(ii) For each $i \in \{1,2\}$, $B \in \arg\max_{\theta_i \in [-B,B]} U_i(\theta_i, B)$.

(iii) For each $i \in \{1,2\}$ and all $\theta_i \in [-B,B]$, $U_i(B,B) \geq U_i(\theta_i, B)$.

If additionally $\theta_i \mapsto U_i(\theta_i, B)$ is differentiable on $(-B,B)$, then (iii) is implied by

$$\frac{d}{d\theta_i} U_i(\theta_i, B) \geq 0 \quad \text{for all } \theta_i \in (-B,B) \text{ and } i = 1,2.$$

- To show the equivalence claim, recall that by definition, $(B,B)$ is a PPNE if and only if, for each player $i \in \{1,2\}$,

  $$U_i(B,B) \geq U_i(\theta_i, B) \qquad \text{for all } \theta_i \in [-B,B].$$

  This is exactly statement (iii). Hence (i) and (iii) are equivalent.

  Next, fix $i \in \{1,2\}$. Statement (ii) says that

  $$B \in \arg\max_{\theta_i \in [-B,B]} U_i(\theta_i, B),$$

  which means precisely that

  $$U_i(B,B) \geq U_i(\theta_i, B) \qquad \text{for all } \theta_i \in [-B,B].$$

  This is exactly statement (iii). Thus (ii) and (iii) are also equivalent.

- For the differentiability claim, fix $i \in \{1, 2\}$ and suppose $f_i(\theta_i) = U_i(\theta_i, B)$ is differentiable on $(-B, B)$ and satisfies $f_i'(\theta_i) \geq 0$ for all $\theta_i \in (-B, B)$. Then $f_i$ is nondecreasing on the interval $[-B, B]$. Therefore, for every $\theta_i \in [-B, B]$,

$$U_i(\theta_i, B) = f_i(\theta_i) \leq f_i(B) = U_i(B, B),$$

which shows (iii).

### A.4. Neural Semantics

We now prove the neural analogue of the phase transition result.

**Theorem 6 (First-order threshold for neural semantics)** *Let $p_i$ follow the neural open-source semantics with parameters $(\alpha, \beta, W)$, and let $\alpha > 0$. Define*

$$\gamma^\star = \frac{T + P - R - S}{R + T - P - S}.$$

*Then, at the symmetric midpoint $\theta = (0, 0)$,*

$$\left.\frac{\partial U_1}{\partial \theta_1}\right|_{\theta=\mathbf{0}} > 0 \quad \Longleftrightarrow \quad \frac{\beta}{\alpha} > \gamma^\star.$$

*Moreover, the ratio $\beta/\alpha$ is exactly equal to the Jacobian ratio*

$$\frac{\partial p_i / \partial \theta_{-i}}{\partial p_i / \partial \theta_i}$$

*evaluated at $\theta = (0, 0)$, independently of the residual weights $W$.*

*Proof.* Recall that under neural open-source semantics,

$$p_i(\theta) = \sigma\left(\alpha\theta_i + \beta\theta_{-i} + h(\theta_i^2, \theta_{-i}^2, \theta_i\theta_{-i}; W)\right),$$

where

$$h(\theta_i^2, \theta_{-i}^2, \theta_i\theta_{-i}; W) = W_2^\top \tanh\left(W_1^\top [\theta_i^2, \theta_{-i}^2, \theta_i\theta_{-i}] + b_1\right).$$

By construction, the residual term depends only on quadratic monomials in $(\theta_i, \theta_{-i})$. Hence

$$\nabla_\theta h(0, 0; W) = 0.$$

In particular, at the symmetric midpoint $\theta = (0, 0)$, the first-order derivatives of the cooperation probabilities are determined entirely by $\alpha$ and $\beta$. Since $p_1 = p_2 = \sigma(0) = \frac{1}{2}$, we have

$$\frac{\partial p_1}{\partial \theta_1}(0, 0) = \sigma'(0)\,\alpha = \frac{\alpha}{4}, \qquad \frac{\partial p_2}{\partial \theta_1}(0, 0) = \sigma'(0)\,\beta = \frac{\beta}{4}.$$

Using the payoff derivatives from Appendix A.2, evaluated at $p_1 = p_2 = \frac{1}{2}$,

$$\left.\frac{\partial U_1}{\partial p_1}\right|_{p=\frac{1}{2}} = \frac{R + S - T - P}{2}, \qquad \left.\frac{\partial U_1}{\partial p_2}\right|_{p=\frac{1}{2}} = \frac{R + T - S - P}{2},$$

the chain rule gives

$$\frac{\partial U_1}{\partial \theta_1}(0,0) = \frac{\partial U_1}{\partial p_1}\frac{\partial p_1}{\partial \theta_1} + \frac{\partial U_1}{\partial p_2}\frac{\partial p_2}{\partial \theta_1} = \frac{1}{8}\Big[\alpha(R+S-T-P) + \beta(R+T-S-P)\Big].$$

Assuming $\alpha > 0$ and $R+T-P-S > 0$, this quantity is positive if and only if

$$\frac{\beta}{\alpha} > \frac{T+P-R-S}{R+T-P-S} = \gamma^\star.$$

Thus, the local incentive at the symmetric midpoint points toward increased cooperation exactly when $\beta/\alpha > \gamma^\star$, which proves the neural threshold result. Moreover,

$$\left.\frac{\partial p_i/\partial \theta_{-i}}{\partial p_i/\partial \theta_i}\right|_{\theta=(0,0)} = \frac{\beta/4}{\alpha/4} = \frac{\beta}{\alpha},$$

so the ratio $\beta/\alpha$ is exactly the Jacobian ratio governing first-order cross-player sensitivity at the midpoint, independently of the residual weights $W$. Equivalently, $\beta/\alpha$ is the ratio between cross-player and self-player Jacobian sensitivity of the cooperation probability at the midpoint.

## Appendix B. Experimental Details and Hyperparameters

This appendix collects implementation details, optimization settings, and hyperparameters for all experiments reported in the paper. Unless otherwise stated, all experiments are run on the induced parametric game rather than as repeated play of the underlying base game. Thus, when we refer to learning dynamics, we mean numerical optimization trajectories in parameter space under projected gradient ascent. For each experiment, we report the base game, the semantics family, the optimization procedure, the initialization scheme, and the quantities visualized in the corresponding figures.

### B.1. Experiment 1: Open-Source versus Closed Semantics in the Prisoner's Dilemma

This experiment compares open-source and closed-source sigmoid semantics in the one-shot Prisoner's Dilemma with payoff matrix $R = 3, P = 1, T = 5, S = 0$. The cross-player coupling for the open-source sigmoidal semantics is $\gamma = 1.5$, while the closed-source baseline is recovered for $\gamma = 0$. In both cases, the parameter domain is the box $\Theta = [-B, B]^2, B = 5$. For compactness, hyperparameters are also summarized in Table 1.

For each initialization $\theta^0 = (\theta_1^0, \theta_2^0)$, we run simultaneous projected gradient ascent on the induced payoffs,

$$\theta_i^{t+1} = \Pi_{[-B,B]}\big(\theta_i^t + \eta\, \partial_{\theta_i} U_i(\theta^t)\big), \qquad i \in \{1,2\},$$

with learning rate $\eta = 0.15$. In the implementation, the partial derivatives are approximated numerically by finite differences with step size $\varepsilon = 10^{-5}$. Each trajectory is run for 10,000 iterations, although only the first 300 steps are displayed in the learning-dynamics panel of Figure 1. Parameters are projected back to $[-5, 5]$ after every update. The repeated iterations in the experiment do not correspond to repeated play of the Prisoner's Dilemma itself, but to a numerical learning procedure in the induced parametric game over $\Theta = [-B, B]^2$ over timesteps $t \geq 0$ starting from an initial parameter profile $\theta^0 = (\theta_1^0, \theta_2^0)$.

Initial conditions are sampled independently and uniformly from $[-2.5, 2.5]^2$, using 10 random initializations and seed 0. To visualize the induced landscape, social welfare $U_1 + U_2$ is evaluated on a $200 \times 200$ grid over $\Theta$.

Table 1: Hyperparameters to reproduce Figure 1.

| Parameter | Value |
|---|---|
| Base game payoffs | $R = 3,\ P = 1,\ T = 5,\ S = 0$ |
| Open-source coupling | $\gamma = 1.5$ |
| Closed-source coupling | $\gamma = 0$ |
| Learning rate | $\eta = 0.15$ |
| Number of gradient steps | 600 |
| Projection bound | $B = 5$ |
| Parameter domain | $\Theta = [-5, 5]^2$ |
| Number of random initializations | 10 |
| Initialization distribution | uniform on $[-2.5, 2.5]^2$ |
| Random seed | 0 |
| Finite-difference step size | $\varepsilon = 10^{-5}$ |

### B.2. Experiment 2: Gamma Sweep and Phase Transition

This experiment largely mirrors the setup described in Appendix B.1. For each game $G$, we evaluate the open-source semantics over a uniform grid of 100 values of $\gamma$ between 0 and 1.5. The parameter domain is $\Theta = [-B, B]^2, B = 5$. For each game and each value of $\gamma$, we run simultaneous projected gradient ascent on the induced parametric game, with learning rate $\eta = 0.01$ for 600 steps. In the implementation, the partial derivatives are approximated numerically by finite differences. For every $\gamma$, we average over 20 random initializations. Initial parameters are sampled from a centered Gaussian distribution, with standard deviation 0.1 for all games except Chicken, where the standard deviation is reduced to 0.01.

For each run, we record the terminal mean cooperation probability $\bar{p} = \frac{p_1+p_2}{2}$ and the terminal social welfare $U_1 + U_2$. Across random initializations, we report the mean and standard deviation of both quantities as functions of $\gamma$. The empirical critical coupling $\gamma^*_{\text{emp}}$ is defined as the first sampled value of $\gamma$ at which the averaged cooperation curve crosses the threshold $\bar{p} = 0.5$. This is then compared against the analytical threshold $\gamma^\star = \frac{T+P-R-S}{R+T-P-S}$ derived in Theorem 3. Hyperparameters are summarized in Table 2.

### B.3. Experiment 3: PPNE verification

This experiment uses the same base game, parameter domain, and projected gradient-ascent procedure as Appendix B.1, with identical learning-rate, initialization scheme, and optimization horizon and other hyperparameters.

The goal of this experiment is not to study the learning trajectories themselves, but to verify whether a boundary candidate $\theta^\star$ is in fact a parametric program Nash equilibrium. By Theorem 4, this reduces to a one-dimensional unilateral-deviation check for each player. Concretely, after selecting a candidate boundary point $\theta^\star$, we freeze the opponent at $\theta^\star_{-i}$ and evaluate the

Table 2: Hyperparameters to reproduce Figure 2.

| Parameter | Value |
|---|---|
| Swept coupling range | $\gamma \in [0, 1.5]$ |
| Number of $\gamma$ values | 100 |
| Projection bound | $B = 5$ |
| Parameter domain | $\Theta = [-5, 5]^2$ |
| Learning rate | $\eta = 0.01$ |
| Number of gradient steps | 600 |
| Number of random initializations | 20 |
| Initialization distribution | Gaussian, centered at 0 |
| Initialization scale | 0.1 for all games, 0.01 for Chicken |
| Empirical threshold criterion | first $\gamma$ with mean $\bar{p} > 0.5$ |
| Finite-difference step size | $\varepsilon = 10^{-5}$ |
| Stag hunt | $R = 3.0,\ S = 0.0,\ T = 2.0,\ P = 1.0;\ \gamma^\star = 0$ |
| Chicken | $R = 3.0,\ S = 1.0,\ T = 5.0,\ P = 0.0;\ \gamma^\star = 1/7$ |
| Prisoner's dilemma | $R = 3.0,\ S = 0.0,\ T = 5.0,\ P = 1.0;\ \gamma^\star = 3/7$ |
| Harsh prisoner's dilemma | $R = 3.0,\ S = 0.0,\ T = 6.0,\ P = 2.0;\ \gamma^\star = 5/7$ |

best-response objective $\theta_i \mapsto U_i(\theta_i, \theta^\star_{-i})$ over the full interval $[-5, 5]$. In practice, this is implemented as a dense grid sweep with 1000 evaluation points, together with the corresponding derivative curve $\partial U_i / \partial \theta_i$. A candidate $\theta^\star$ is certified as a PPNE only if, for both players, the maximum of the swept best-response curve is attained at $\theta^\star_i$. Equivalently, the unilateral-improvement gap $\max_{\theta_i \in [-5,5]} U_i(\theta_i, \theta^\star_{-i}) - U_i(\theta^\star)$ must be numerically zero up to tolerance.

For the open-source condition, we use sigmoid semantics with coupling $\gamma = 1.5$ as in Figure 3 where the best-response payoff is maximized at the cooperative boundary $\theta^\star_i = B$, so $(B, B)$ is a PPNE. In Figure 5 we use $\gamma = 0.7$, the payoff attains a strictly larger value at an interior deviation, showing that $(B, B)$ is not a PPNE despite being a boundary candidate. Candidate boundary points are taken to be $(5, 5)$ in the open-source case and $(-5, -5)$ in the closed-source case. To compare with the gradient-ascent dynamics, we also run the same solver as in Appendix B.1 from 10 random initializations drawn uniformly from $[-2.5, 2.5]^2$, using learning rate 0.15, 600 update steps, clipping bound $B = 5$, and random seed 0, which is the same configuration as in Table 1.

The numerical verification results for $\gamma = 0.7$, intended not to be a PPNE, are shown in Figure 5. In this configuration, the open-source boundary candidate $(5, 5)$ is not a PPNE, as for each player, the best-response curve achieves a strictly larger value away from the boundary, with a unilateral-improvement gap of approximately 1, in contrast to Figure 3, where the gap is approximately 0. This illustrates the distinction emphasized in the main text, i.e., the convergence of gradient dynamics toward a boundary region does not by itself imply that the limiting boundary point is an equilibrium of the induced game. The additional figure is included to show how the same sweep procedure certifies failure of the PPNE property when a profitable unilateral deviation exists.

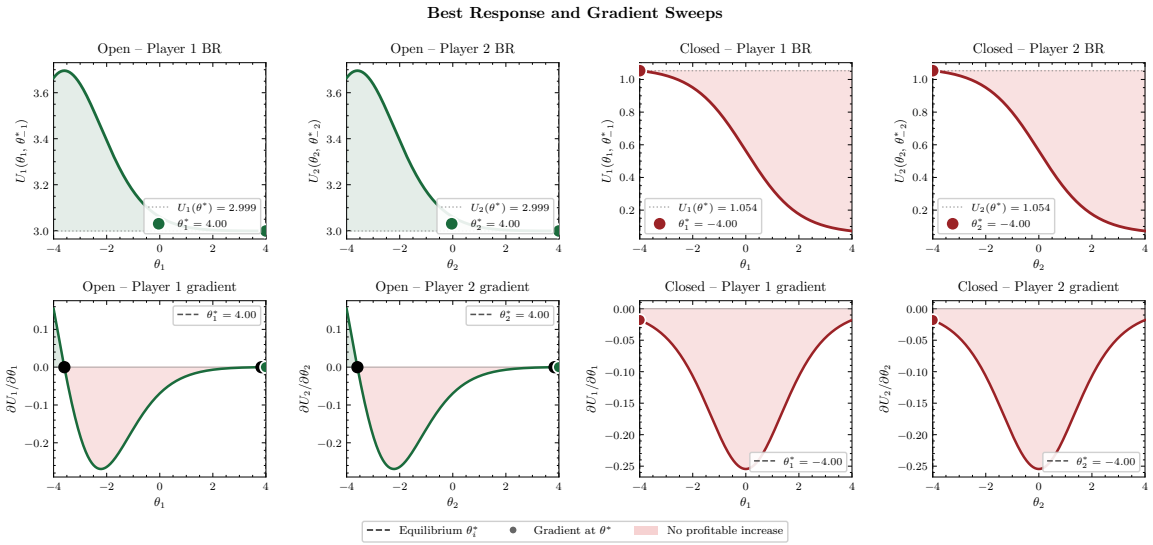


Figure 5: Boundary best-response verification for a non-PPNE open-source candidate. Using the same sweep procedure as in Figure 3, but with weaker open-source coupling $\gamma = 0.7$ rather than $\gamma = 1.5$, the candidate boundary point $(B, B)$ fails the PPNE test. The best-response curves are not maximized at $\theta_i^\star = B$, so each player has a profitable unilateral deviation when the opponent is fixed at the boundary. This contrasts with the main-text setting in Figure 3, where $\gamma = 1.5$ and the cooperative boundary is verified as a true PPNE. The solid black dot in Panels (e) and (f) indicates the best response found by the fine-grid sweep.

### B.4. Experiment 4: Neural semantics

This experiment extends the gradient-ascent setup of Appendix B.1 to the neural semantics class introduced in Section 3.3. We evaluate the four canonical symmetric $2 \times 2$ games shown in Figure 4, namely Stag Hunt, Chicken, Prisoner's Dilemma, and Harsh PD. In all conditions, we report the social welfare $U_1 + U_2$ as a function of optimization time and average across multiple random seeds.

We compare six conditions. The first two are the sigmoid baselines from the main text, namely sigmoid-closed with $\gamma = 0$ and sigmoid-open with $\gamma = 1.5$. The next two use the neural semantics of Section 3.3 with fixed residual weights. In the closed condition, we set $(\alpha, \beta) = (1, 0)$, while in the open condition we set $(\alpha, \beta) = (1, 1.5)$, so that $\beta/\alpha = 1.5$ matches the open sigmoid coupling. In both cases, the network weights $(W_1, b_1, W_2)$ are sampled once at initialization and then held fixed throughout optimization. The same sampled residual network is used for the fixed closed and fixed open conditions, so the comparison isolates the effect of the linear coefficients.

The final two conditions jointly learn both the player parameters and the semantics parameters. In the warm-start condition, the learnable parameters are initialized with $(\alpha_0, \beta_0) = (1, 1.5)$, so the initial first-order ratio already lies in the cooperative regime predicted by Appendix A.4. In the cold-start condition, they are initialized with $(\alpha_0, \beta_0) = (0, 0)$. For each player, the learnable parameter vector consists of the current scalar parameter $\theta_i$, the linear coefficients $\alpha, \beta$, and the flattened residual-network weights. Gradients are approximated by finite differences, and all parameters are updated by simultaneous gradient ascent on each player's own payoff.

Player parameters are initialized from a clipped normal distribution centered at zero, with a clipping radius of 2.0. During optimization, the player parameters are clipped to $[-10, 10]$, while the learnable semantics parameters are clipped componentwise to $[-5, 5]$. The residual network uses hidden width $K = 4$, initialization scale $0.05$, and zero initial bias vector $b_1$. The full set of hyperparameters is reported in Table 3.

The main empirical pattern is that the neural class reproduces the sigmoid baselines when the first-order ratio $\beta/\alpha$ is matched, but joint optimization succeeds reliably only under warm initialization. This is consistent with Appendix A.4, which identifies $\beta/\alpha > \gamma^\star$ as the condition under which cooperation is locally attractive at the symmetric midpoint. The fixed-open and warm-start learned conditions satisfy this requirement from the outset, whereas the cold-start learned condition must discover it through optimization and fails to do so.

Table 3: Hyperparameters for the neural-semantics experiment in Figure 4. The fixed-neural conditions use the same sampled residual network in both the closed and open variants, differing only in the linear coefficients $(\alpha, \beta)$. The learned neural conditions jointly optimize the player parameters and all semantics parameters.

| Hyperparameter | Value |
|---|---|
| Open sigmoid coupling $\gamma$ | $1.5$ |
| Hidden width $K$ | $4$ |
| Residual weight scale | $0.05$ |
| Residual-network seed | $0$ |
| Finite-difference step $\varepsilon$ | $10^{-5}$ |
| Player learning rate | $0.10$ |
| Semantics-parameter learning rate | $0.005$ |
| Player clipping bound | $10.0$ |
| Semantics-parameter clipping bound | $5.0$ |
| Player initialization clip | $2.0$ |
| Number of steps for fixed-semantics runs | $1000$ |
| Number of steps for learned-semantics runs | $1000$ |
| Number of seeds for fixed-semantics runs | $5$ |
| Number of seeds for learned-semantics runs | $5$ |

## Appendix C. Camera-Ready Edits

We improved the quality of the figures, bolded and shortened their titles, and improved the captions within the manuscript. We also added an Acknowledgment section before the references.