LAB TO ALGORITHM: PREDICTING AIS WITH HUMANS, AND VICE VERSA

MARTA BOCZOŃ, EMANUEL VESPA, AND ALISTAIR J. WILSON

Abstract. A now mature literature on repeated prisoner's dilemma has outlined a number of regularities in how human subjects behave. In this literature a core task is to predict when the participants will collude on the jointly cooperative action, and when they will coordinate on the myopic solution: joint defection. Orthogonal to this, a new literature in industrial organization has begun to look at when Artificial Intelligence (AI) pricing agents collude in repeated settings. In this paper we begin to explore the extent to which the regularities that show-up in human subject behavior also manifest in the behavior of pricing agents. While there are similarities, that we document, there are also points of divergence. Moving forwards, the aim is to connect both literatures: Theoretical rules developed for human subjects can be predictive for AI agents, and thereby a useful tool for theoretic exercises in predicting AI in counterfactual settings. Conversely, AI agents can be used to develop insightful experiments to further refine and test our understanding of human behavior through experiments. As such, the tasks of predicting and understanding both human and AI behavior can be symbiotic.

1. Introduction

The *basin of attraction for always defect* (Blonski and Spagnolo, 2001, 2015) has been shown to serve as a clear line-in-the-sand for predicting regions where one may expect collusive outcomes in infinitely repeated games. This ordinal property of the basin has been first documented by (Dal Bó and Fréchette, 2018) in a two-player Repeated Prisoner Dilemma (RPD) and recently, has been extended to a multiple-player setting by Boczoń, Weidman, Vespa, and Wilson (2024).

Thus far, data necessary to evaluate ordinal properties of selection criteria, such as the basin of attraction, were bound to be extracted from observed behavior of human subjects in the laboratory, with treatment parameters often selected to study other hypotheses. While meta-studies have brought some of this together, the ability to study a wider set of parameters across many other environments is necessary to pin down which measures are most predictive and the domains they can cover. However, experimental methods are often best- placed to examine relatively coarse hypotheses, across a sparse set of parameters. As such, it is particularly useful to find empirically driven methods that might supplement and target experiments for maximum inference.

In this paper, we propose a simulation method for evaluating ordinal properties of selection criteria outside of the laboratory using Artificial Intelligence (AI). Our approach is based on an emerging literature in industrial organizations that examines the pricing behavior of Artificial Intelligence Agents (AIAs). Given the growing interest in AIAs as

Date: February, 2024.
We would like to thank: David Cooper and Guillaume Fréchette.

pricing agents, and their potential for collusion, there is a natural connection between AIAs and measuring cooperation behavior in repeated games. Specifically, we examine how the steady-state behavior of AIAs is related to the behavior of lab subjects in N-player RPD. In particular, we will demonstrate a strong parallelism between laboratory results for two- and N-player RPD and the results from experimental simulations using AIAs, with extensive variation across the number of players and game primitives.

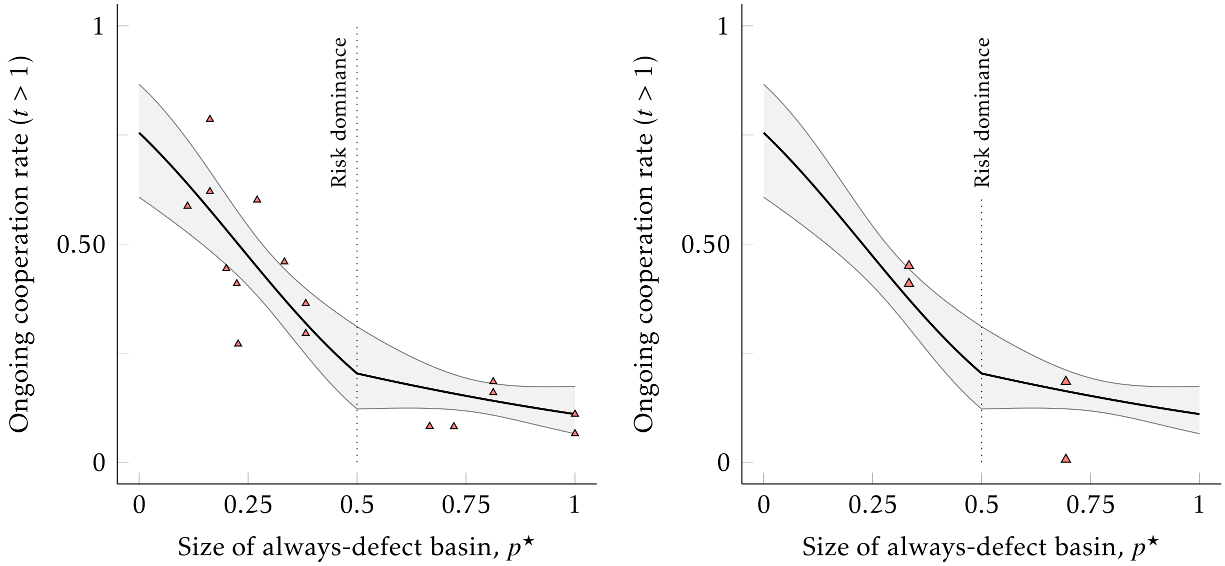
1.1. **Literature.** Until now, economic studies on AI have predominantly focused on the risks associated with companies increasingly relying on algorithms for pricing decisions. As demonstrated by Calvano, Calzolari, Denicolo, and Pastorello (2020), the widely-used AI-learning algorithm Q-learning (Watkins, 1989) has the capability to sustain supracompetitive prices in a standard dynamic Bertrand environment with implicit coordination. This finding has prompted concerns among many that the growing dependence on AI may undermine market competitiveness and potentially lead to collusive behavior. See, for example, Brown and MacKay (2023); Chassang and Ortner (2023).

To address this ongoing debate, Asker, Fershtman, and Pakes (2021) and Asker, Fershtman, and Pakes (2022) delve into how AI decisions are influenced by the type of algorithm used. They illustrate that supracompetitive prices hinge on the extent to which AI algorithms can learn counterfactually from alternative choices (referred to as synchronous learning), rather than relying solely on learning from on-the-path experiences (termed asynchronous learning). In differentiating between asynchronous and synchronous updating, the former only requires knowledge of the profits received from the actually played price, while the latter's information requirement depends on how profits from counterfactual prices are calculated.

In a distinct context, AI-driven pricing algorithms have also been applied by Johnson, Rhodes, and Wildenbeest (2023) to assess a platform's ability to structure its marketplace in a way that promotes competition, enhances consumer surplus, and maximizes its own payoff. Here, we use AIs to explore parallels in cooperative behavior between pricing algorithms and humans.

This paper also aligns with the experimental literature exploring the evolution of cooperation in infinitely repeated games. It is connected to Dal Bó (2005), who contrasts cooperation rates in finite and infinite horizon PD games; Aoyagi and Fréchette (2009), who demonstrate that in RPD games with imperfect public monitoring, the degree of cooperation increases with the quality of the public signal; and Duffy and Ochs (2009) who observe a rise in cooperation as subjects gain experience in RPD games with high continuation probability. In this study, we employ AI agents to evaluate ongoing levels of cooperation across an even broader range of canonical and multi-player RPD games.

Finally, this paper contributes to a broader literature aiming to comprehend and document patterns in equilibrium selection, particularly those suitable for theoretical modeling. The initial strides in this direction were taken in the laboratory by Dal Bó and Fréchette (2011) and more recently by Boczoń, Weidman, Vespa, and Wilson (2024). In alignment with this literature, we extensively test the basin of attraction for always defect as a measure of strategic uncertainty using simulations of both two- and multi-player RPD games.



(A) Ongoing cooperation, $N = 2$

(B) Ongoing cooperation, $N > 2$

Figure 1. Strategic uncertainty and cooperation rates in RPD

Note: In both figures the solid line indicates the meta-study predictions for cooperation rate at each $p^\star$ from the piecewise-linear probit estimates; the shaded region represents the 95 percent confidence intervals for the prediction. In the left figure, each data point indicates a separate treatment in Dal Bó and Fréchette (2018). In the right figure, each data points indicates a separate treatment in Boczoń, Weidman, Vespa, and Wilson (2024).

2. Basin of Attraction

In this section, we first characterize the basin of attraction for always defect in a two-player RPD and then, extend the basic framework to an environment with multiple players and independent beliefs.

The *basin for always defect* is defined as the set of beliefs p for which player i receives a higher expected payment from defection than cooperation. For a two-player RPD the basin of attraction is the interval $[0, p^\star(x,\delta)]$ with the critical-point given by:

$$p^\star(x,\delta) \equiv \frac{(1-\delta)x}{\delta}, \qquad (1)$$

where $\delta \in (0,1)$ is the discount rate, and $x > 0$ captures the relative temptation-/sucker-payoff.

Using experimental results from the meta-study on the two-player RPD (Dal Bó and Fréchette, 2018) in Figure 1(A) we illustrate the relationships between the scalar basin-size measure of strategic uncertainty and ongoing cooperation rates (in rounds two and beyond). The solid line indicates the predicted cooperation rate at each $p^\star$ from the piecewise-linear probit estimates; the shaded region represents the 95 percent confidence interval for the prediction (clustering by treatment).[1] From the figure we observe consistently low levels of cooperation when always-defect is risk dominant ($p^\star > 1/2$); and a significantly decreasing relationship with $p^\star$ when collusion is risk dominant ($p^\star < 1/2$).

To generalize the notion of basin of attraction for always defect to an N-player environment we follow Boczoń, Weidman, Vespa, and Wilson (2024). We assume that beliefs of all N players are fully independent and that players' payoffs are determined solely by their own action and a deterministic binary signal of the others' actions.[2] Under independence the generalized version of the basin of attraction is then given by

$$p^\star(x, \delta, N) = \left(\frac{1-\delta}{\delta} x\right)^{1/N-1}. \tag{2}$$

In Figure 1(B) we show the relationship between the scalar basin-size measure and ongoing cooperation rates in multi-player RPD studied in Boczoń, Weidman, Vespa, and Wilson (2024). As for $N = 2$ we observe unambiguous directional predictions in cooperation for any counterfactual change in the primitives.

3. Q-Learning algorithms

In order to illustrate parallels between AIAs an human subjects, we simulate over 1.8 million games with an identical structure to an N-player RPD environment, but whre two-state AIAs act as the decision makers.[3] In our simulations we vary: (i) the number of players, $N = \{2, 3, \ldots, 10\}$; (ii) the discount factor, $\delta = \{0.75, 0.90, 0.95, 0.99\}$; the always-defect basin size ($p^\star(x, N, \delta) = \{0, 0.01, 0.03, 0.05, \ldots, 0.99\}$, chosen by varying x; and (iv) the algorithm learning mode. In the asynchronous learning mode AIAs learn solely from the payoffs observed from their chosen decisions, whereas in the synchronous mode we allow AIAs to learn both from the path and the counterfactual.[4] For each treatment environment/algorithm we simulate 1,000 distinct repeated games, where each simulated game runs for 10,000 rounds (this was a sufficient length to obtain convergent behavior for all treatments/algorithm modes). Our final measures from each simulation are the

[1] We estimate the probit regression using meta-study data from 996 participants across 18 experimental treatments, where we focus on late-session cooperation (supergames 16-20). The individual-level cooperation decisions serve as the left-hand side variable, and the basin size is included on the right-hand side in a piecewise-linear fashion around the risk-dominance dividing point. Our econometric specification is inspired by Dal Bó and Fréchette (2018, Table 4). However, to maintain a continuous relationship, we modify their specification by eliminating a degree of freedom that allowed for a discontinuity at $p^\star = 1/2$.

[2] The authors also study a correlated extension and find that it cannot well explain the observed behavior of subjects in the laboratory.

[3] With two internal states the AIA decision makers have access to a conditioning variable that could be used to construct a history-dependent strategy such as the grim trigger. However, the way the algorithm makes use of this state variable is entirely endogenous, determined by the particular learning path.

[4] We thank John Asker for sharing MATLAB code, which we re-implemented in *Python*.

ongoing cooperation rates among AIAs, where initial behavior is entirely random, driven by an initially diffuse uniform distribution over the action choice weights for each state.[5]

4. Basin Parallelism with AI behavior

We present the results of our AIA simulations in Figure 2, with the asynchronous results in Panel (A) and the synchronous results in Panel (B). Each triangular data point represents the average long-run cooperation rate across AIAs at a given value of $p^{\star}$, pooling treatments across N and δ.[6] As such, each point represents an average across 18,000 AIA supergames. In each figure we superimpose the fitted relationship between the basin and ongoing cooperation from the RPD meta-study using human subjects, cf. Figure 1(B). Figure 2(A) makes clear a top-level observation that the results from the asynchronous algorithm display behavior that is highly consistent with the predictions of the independent extension: collusion decreases as $p^{\star}$ increases and essentially disappears once $p^{\star} > 0.5$. In the region with $p^{\star} < 0.5$, asynchronous AIAs broadly mirror the behavior of subjects in the laboratory.[7] For values of $p^{\star}$>0.5, asynchronous AIAs cooperate less than humans, although the difference is not large. In contrast, for the more sophisticated synchronous algorithm shown in Panel (B), we observe much larger differences in behavior between AIAs and humans. Mirroring the results from Asker, Fershtman, and Pakes (2022), the synchronous algorithm is much less successful at colluding, only doing so at very low values of the basin.

We conclude that:

Result 1 (Exploration of AIAs Behavior Relative to Humans)**.** *Asynchronous AIAs that learn only from past experiences on the path display collusion behavior that is consistent with the prediction of the independent extension of the size of always-defect basin and track the behavior of human subjects quite closely. On the contrary, there are large differences between the behavior generated by sophisticated synchronous algorithm that also learns counterfactually and the behavior of humans.*

5. Conclusion

Our results here suggest that Q-learning algorithms can be predictive of human behavior in these repeated settings. Future research can explore and leverage this link, where we now outline some of the possible ways this can be accomplished. First, the exercise suggests ways in which AIAs can complement the laboratory. For example, in a standard

[5] In general, we follow Asker, Fershtman, and Pakes (2022) in this setup, with the only substantial change being the switch from a dynamic Bertrand environment they study with many price actions, to the two-action environment studied in our laboratory treatments.

[6] AIAs we study require a substantial degree of training to converge. For this reason, we examine the long-run, convergent behavior of Q-learning AIAs within our simple N-player social dilemma environment. The ongoing cooperation rates that we report correspond to the convergent behavior that AIAs achieve for a given parameterization.

[7] For higher N, and lower $p^{\star}$ the data exhibit a non-monotonicity for $\delta = 0.75$ in regions close to a zero basin. The reason for this is that at very low values of x (<< $10^{-}5$), asynchronous AIAs have difficulties learning the relevant punishment strategies to support cooperation, and intead, serially alternate between cooperation and non-cooperation.

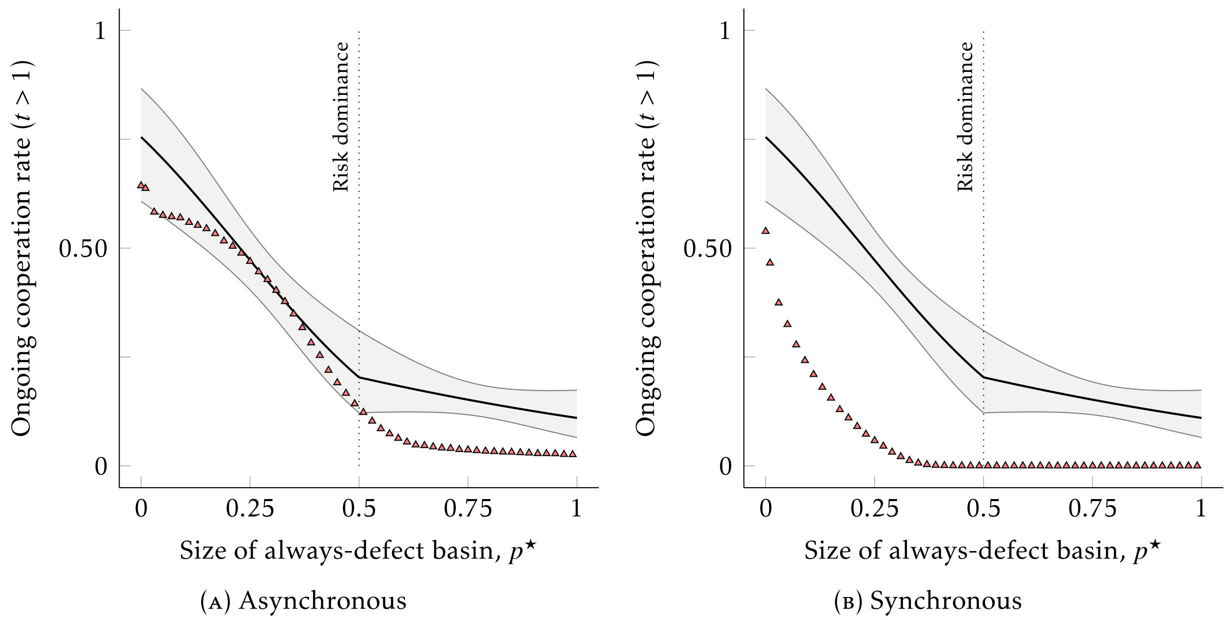


FIGURE 2. Cooperation behavior of two different AIAs

RPD environment, the asynchronous algorithm can be used to predict behavior for sets of parameters for which there is no/very little experimental data. So long as one can show some parallels between AIAs and human-subject behavior, AIAs might be used for thought experiments or exploring the extremes of the parameter space.

Greater exploration of the parameter space may then help fine-tune empirical selection criteria, even in settings for which there exists substantial data. For example, some facets of AIAs behavior may not be not fully captured by the summary basin $p^\star$. As an example, in two-player PD games with high temptations and low sucker payoffs, AIAs begin to exhibit serial alternation across the (Cooperate, Defect) and (Defect, Cooperate) actions well before this behavior becomes efficient. This prediction from AIAs can then be examined in the laboratory. The data from such experiments could clear up whether the predicted discrepancies were exclusive to AIAs, or whether they are shared by humans, suggesting a need for a correction to the selection criterion at these regions of the parameter space.

Finally, AIAs can be used to explore behavior and shape selection theory in extension environments that differ from the RPD. For example, with AIAs it is relatively simple: (i) to expand the action set (as in the Bertrand/Cournot setting); (ii) to allow for state variables that evolve with the game (stochastic/dynamic games); (iii) to allow for imperfect monitoring (á la Green and Porter, 1984); or (iv) to study features that reduce strategic uncertainty such as sequential moves or explicit communication between AIAs. Naturally, studying whether empirical selection criteria such as the basin of attraction for always defect work in these other settings are outside of the scope of this paper. However, we suspect that AIAs will be a key aide for future explorations of these selection questions within experimental contexts. Moreover, the increasing interest in AIAs will mean that studying their behavior will have increasing external validity.

References

Aoyagi, M. and G. Fréchette (2009), "Collusion as public monitoring becomes noisy: Experimental evidence." *Journal of Economic Theory*, 144, 1135–1165.

Asker, John, Chaim Fershtman, and Ariel Pakes (2021), "Artificial intelligence and pricing: The impact of algorithm design." Working Paper 28535, National Bureau of Economic Research, URL http://www.nber.org/papers/w28535.

Asker, John, Chaim Fershtman, and Ariel Pakes (2022), "Artificial intelligence, algorithm design and pricing." *AEA Papers and Proceedings*, 112, 452–56.

Blonski, Matthias and Giancarlo Spagnolo (2015), "Prisoners' other dilemma." *International Journal of Game Theory*, 44, 61–81.

Boczoń, Marta, Taylor Weidman, Emanuel Vespa, and Alistair J Wilson (2024), "Testing models of strategic uncertainty: Equilibrium selection in repeated games." *In preparation*.

Brown, Zach Y. and Alexander MacKay (2023), "Competition in pricing algorithms." *American Economic Journal: Microecnomics*, 15, 109–156.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolo, and Sergio Pastorello (2020), "Artificial intelligence, algorithmic pricing, and collusion." *American Economic Review*, 110, 3267–3297.

Chassang, Sylvain and Juan Ortner (2023), "Regulating collusion." *Annual Review of Economics*, 15, 177–204.

Dal Bó, Pedro (2005), "Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games." *American Economic Review*, 95, 1591–1604.

Dal Bó, Pedro and Guillaume R Fréchette (2011), "The evolution of cooperation in infinitely repeated games: Experimental evidence." *American Economic Review*, 101, 411–429.

Dal Bó, Pedro and Guillaume R Fréchette (2018), "On the determinants of cooperation in infinitely repeated games: A survey." *Journal of Economic Literature*, 56, 60–114.

Duffy, John and Jack Ochs (2009), "Cooperative behavior and the frequency of social interaction." *Games & Economic Behavior*, 66, 785–812.

Green, E.J. and R.H. Porter (1984), "Noncooperative collusion under imperfect price information." *Econometrica*, 87–100.

Johnson, Justin P., Andrew Rhodes, and Matthijs Wildenbeest (2023), "Platform design when sellers use pricing algorithms." *Econometrica*, 91, 1841–1879.

Spagnolo, Giancarlo and Matthias Blonski (2001), "Prisoners' other dilemma." *Working Paper*.

Watkins, Christopher (1989), *Learning from delayed rewards*. Ph.D. thesis, University of Cambridge, King's College, Cambridge, United Kingdom.