

# Two Rational Players Walk into an Undecidable Game (An Informal Note)

Pedro Gardete  
(Nova School of Business and Economics)\*  
June 2025

## Abstract

This note considers a repeated game with a prisoner's dilemma stage game, where Player A observes the output of a computational process  $P_0$ , which embeds a program  $P$  with undecidable halting behavior. Player B does not observe A's actions or  $P_0$ 's output, and observes noisy realizations of its own payoffs. While Nash equilibrium remains well-defined, verifying whether a given strategy profile constitutes an equilibrium is undecidable, as it requires evaluating deviations contingent on  $P_0$ 's noncomputable output. Consequently, interpreting Nash equilibrium as a prediction of play requires players to resolve undecidable problems.

---

\*Gardete: [pedro.gardete@novasbe.pt](mailto:pedro.gardete@novasbe.pt). The author is not an expert in this topic and finds it difficult to adequately apologize for this note. All rights granted by tenure, freedom of speech, and any other applicable legislation are hereby invoked to legitimate its existence. GPT and Perplexity.ai provided substantial assistance, particularly in navigating the near-infinite density of concepts in the existing literature.

# 1 The Argument

**Computational Device.** Let  $P$  be a program whose halting behavior is undecidable. Embed  $P$  into a wrapper  $P_0$  that, at each period  $t$ , checks whether  $P$  has halted; if so, it outputs one and restarts  $P$ ; otherwise, it simply outputs zero and allows  $P$  to keep running.

**Game.** Let two players A and B compete in a repeated game with a prisoner's dilemma stage game. Player A observes the computational output of  $P_0$  at each time  $t$ , and may condition its action (e.g., cooperate or defect) on that output. For example, A might cooperate if the output is zero, and defect if the output is one. Given its cyclic nature,  $P_0$  never halts, and so its output is available to player A throughout the game's infinite duration. Player B does not observe A's actions, payoffs, or the output of  $P_0$ , which are Player A's private information. Player B knows the programs  $P$  and  $P_0$ , and that A has access to the output of  $P_0$  in each period. Let player B observe a noisy version of its own payoff.

Consider some strategy profile in which B always cooperates. Then, with some probability  $p$ , B observes the payoff corresponding to the  $(C, C)$  profile; with probability  $1 - p$ , it observes the payoff from  $(D, C)$ . Non-Markovian strategies are allowed. For example, B may punish A forever if it believes a deviation has occurred with sufficiently high probability or if it observes a certain proportion of low payoffs, possibly induced by A's deviations.

**Implications.** The minimal description above roughly defines a family of games with well-defined Nash equilibria whose characterization is undecidable. Suppose we consider a strategy profile  $\sigma$  in which A ignores the output of  $P_0$  altogether. Checking whether  $\sigma$  constitutes a Nash equilibrium requires determining whether A would benefit from deviating to a strategy that conditions on  $P_0$ 's output. This procedure is infeasible, since the result by Turing et al. (1936) makes it impossible to compute the payoff of the deviation. Moreover, the payoff distribution induced by A's potential deviations is noncomputable, making it impossible for B to form well-defined beliefs over  $P_0$ 's output or A's future behavior. Nash equilibrium and its existence remains well-defined in the standard sense, since the stage-game actions and payoffs are themselves well defined.

Given the description above, the reader may think that it is only equilibrium analysis that is compromised. However, depending on whether one interprets Nash equilibrium as a prediction of actual outcomes, play itself may be compromised. For example, select  $P$  to be a program whose halting status is undecidable. If one interprets Nash equilibrium as a prediction of play, it becomes unclear what fully rational players would play in such a setting, or even whether fully rational players could meaningfully play such a game at all. One would, in effect, be required to assume that rational players are capable of resolving undecidable problems to fully implement Nash equilibrium.

## References

TURING, A. M., ET AL. (1936): “On computable numbers, with an application to the Entscheidungsproblem,” *J. of Math*, 58(345-363), 5.