



HAL
open science

Balancing rationality and social influence: Alpha-rational Nash equilibrium in games with herding

Khushboo Agarwal, Konstantin Avrachenkov, Veeraruna Kavitha, Raghupati Vyas

► **To cite this version:**

Khushboo Agarwal, Konstantin Avrachenkov, Veeraruna Kavitha, Raghupati Vyas. Balancing rationality and social influence: Alpha-rational Nash equilibrium in games with herding. GameNets 2025 - 13th EAI International Conference on Game Theory for Networks, Mar 2025, Cambridge, United Kingdom. hal-04955252

HAL Id: hal-04955252

<https://inria.hal.science/hal-04955252v1>

Submitted on 18 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Balancing rationality and social influence: Alpha-rational Nash equilibrium in games with herding^{*}

Khushboo Agarwal¹, Konstantin Avrachenkov¹, Veeraruna Kavitha², and
Raghupati Vyas²

¹ Inria Sophia Antipolis, 2004 Route des Lucioles, Valbonne 06902, France
{khushboo.agarwal,k.avrachenkov}@inria.fr

² IEOR, IIT Bombay, Powai, Mumbai, 400076, India
{vkavitha, raghupati.vyas}@iitb.ac.in

Abstract. The classical game theory considers rational players and proposes Nash equilibrium (NE) as the solution. However, real-world scenarios rarely feature rational players; instead, players make inconsistent and irrational decisions. Often, irrational players exhibit herding behaviour by simply following the majority.

In this paper, we consider a mean-field game with α -fraction of rational players and the rest being herding-irrational players. For such a game, we introduce a novel concept of equilibrium named α -Rational NE (in short, α -RNE). We extensively analyze the α -RNEs and their implications in games with two actions. Due to herding-irrational players, new equilibria may arise, and some classical NEs may be deleted.

We establish that the rational players are not harmed but benefit from the presence of irrational players. More interestingly, in some examples, the rational players attain higher utility (under α -RNE) than even the social optimal utility (in the classical setting), by leveraging upon the herding behaviour of irrational players.

Surprisingly, the irrational players may also benefit by not being rational. We observe that irrational players do not lose compared to some classical NEs for participation and bandwidth-sharing games. Importantly, in bandwidth-sharing game, the irrational players also receive utility near social optimal utility. Such examples indicate that it may sometimes be ‘rational’ to be irrational.

Keywords: bounded rationality · game theory · mean-field games

1 Introduction

Classical game theory explores the interactions between rational and intelligent players. In [14], a player is defined as rational if it consistently makes deci-

* The work is supported by: the French National Agency for Research (ANR) via the project n°ANR-22-PEFT-0010 of the France 2030 program PEPR réseaux du futur; DST-INRIA Cefipra project ‘Learning In Operations and Networks’ (LION); and CSIR, India. This work has been accepted at EAI GameNets 2025 and will be published in the Springer LNICST series.

sions aligned with its objectives, striving to maximize its utility. Additionally, an intelligent player possesses complete knowledge of the game and can perform computations to identify its optimal strategy. This line of thought is widely acknowledged and serves as a benchmark for analyzing an ideal world.

Nevertheless, contemporary perspectives challenge the strong assumptions regarding rationality and intelligence due to human irrationality and computational limitations. This has sparked interest in understanding actual human behavior, leading to the emergence of fields like behavioural game theory, behavioural economics, and neuro-economics (see [8, 18, 19] respectively). Social experiments play a major role in driving research in these domains.

Several behavioral traits like behavioral probability weighting, cognitive hierarchy model, time-inconsistent planning, impact of extrinsic vs intrinsic motivation, among the players have been studied in [1, 6, 7, 17] recently. In these models, the players are partially aligned towards their utilities. Another strand of literature considers players who do not have the full potential to evaluate their utility functions accurately. More realistic notions of equilibrium considering such bounded-rational players are proposed, for example, quantal response, action-sampling, payoff-sampling, impulse-balance and k -fault tolerant equilibrium (see [10, 15, 20, 21]). A common thread in all these models is that the players exhibit some form of rationality or the other. We differ precisely at this point.

We attempt to model the players who do not even acknowledge the utility function while making choices. In particular, for some $\alpha \in (0, 1]$, we consider two types of players: (i) rational players, constituting an α -fraction of the population, and (ii) *irrational players, who exhibit herding behavior*. The latter group does not optimize the utility function—they simply choose the action taken by the majority. Such behavior is discussed at length in papers like [5, 11, 13, 22] and is also evident in career or fashion choices driven by the trends, and consumers favoring famous brands, etc. In the mentioned papers, either the focus is only on the herding players or they do not consider game theoretic aspects, while we study the resultant of the strategic and ‘herding’ interactions in a mixture of the population.

Assuming a large number of players, we analyze the game within mean-field framework. Drawing inspiration from the Nash equilibrium (NE) in the classical mean-field games, for a given value of α , *we introduce a novel equilibrium concept termed ‘ α -Rational NE’ (in short, ‘ α -RNE’) for our game, which encompasses both the rational and irrational choices of the players.*

In this paper, we consider games with two actions to obtain initial insights into the new concept of α -RNE. There are numerous examples of important games with two choices (for instance, participation games with participate or not-participate choices, minority games with two restaurants, vaccination games, etc.), which highlight the relevance of the games of interest. Further, we believe the insights derived here can prove to be a stepping stone for extending the notions of this paper to general games (with more actions) in the future.

Notably, we show that the set of α -RNEs does not always coincide with the set of NEs. If rational players dominate the system ($\alpha > 1/2$), the two sets are

the same. However, when $\alpha \leq 1/2$, some classical NEs may disappear, and two new equilibria (α or $1 - \alpha$) can emerge. We also provide simple conditions for identifying α -RNEs based on the utility difference function.

After characterizing the α -RNEs, we compare the utilities of rational and herding players at a given α -RNE and assess whether either type is at an advantage or disadvantage relative to the NE. We show that rational players benefit from the presence of herding players, rather than being harmed. This aligns with the observation in [2], where behavioral probability weighting is used to model player behavior, but here we deduce this in the context of herding players.

Interestingly, under α -RNE, *the rational players can receive more utility than the social optimal utility obtained under $\alpha = 1$ case*. Even though the incompetent behaviour of the irrational players never allows them to surpass the social optimal utility, nonetheless, in some cases, the irrational players get utility at par with the rational players. Remarkably, in an example, we observe that *all the players (including irrational players) receive higher utility at α -RNE than they could have received if everyone were rational*. Such instances encourage us to claim that *'it may be rational to be irrational sometimes'*.

Another motivation for α -rational NE comes from the behavioral game dynamics. In fact, we showed in [3] that any limit of turn-by-turn dynamics, involving both myopic rational and herding players, is one of the α -RNEs.

2 New notion: α -RNE with herding

In classical game theory, it is assumed that all players are perfectly rational, and then, the widely-known and accepted Nash Equilibrium (NE) is provided as the solution of the game. However, in reality, we rarely encounter such perfectly rational players. Instead, more often than not, players take decisions based on some simple rules. The most common of such rules is the one where players exhibit herding behaviour; for example, in a stock market, players tend to buy the derivative that they believe the majority of the players will purchase, or on a traffic signal, people cross the road when they see others crossing the road, etc.

Our aim in this paper is to propose an appropriate notion of equilibrium that caters to such a mix of rational and irrational (to be more specific, the ones with the herding behaviour) players. Towards this, consider a large population and assume that there are α fraction of rational players, while the remaining population is composed of irrational players, for some $\alpha \in (0, 1]$. Each player has to choose an action from the set of actions, denoted by $\mathcal{A} := \{1, 2, \dots, n\}$, where $n < \infty$. For each $a \in \mathcal{A}$, let $\mu(a)$ be the fraction of players who choose action a ; define $\mu := (\mu(a))_{\{a \in \mathcal{A}\}}$. Similarly, let $\mu^R := (\mu^R(a))_{\{a \in \mathcal{A}\}}$ be the empirical distribution corresponding to (only) rational players.

Define the function $u : \mathcal{A} \times [0, 1]^{|\mathcal{A}|} \rightarrow \mathbb{R}$ to represent the utility of players. Thus, each player receives the utility $u(a, \mu)$ if it chooses an action a and the empirical distribution of the actions by the rest of the population is μ . Note that the utility function is the same for all the players and depends upon μ like in mean-field games (see, for example, [9]). As in classical theory, the rational

players are capable of performing extensive computations and thus, choose an action that maximizes their utility. Hence, if μ were the empirical distribution of the actions chosen by the entire population³, the best response (μ^R) of any rational player against μ would satisfy the following:

$$\text{support}(\mu^R) \subseteq \text{Arg max}_{a \in \mathcal{A}} u(a, \mu), \quad (1)$$

where $\text{support}(\mu) := \{a \in \mathcal{A} : \mu(a) > 0\}$.

On the other hand, irrational players exhibit herding behaviour—they blindly follow others and do not optimize like rational players. One simple way to model the herding behaviour of the irrational players is to assume that such players choose an action that is played the most by rational players, i.e., they choose the action a which satisfies $\mu^R(a) \geq \mu^R(a')$ for all $a' \neq a$. However, since typically irrational players cannot distinguish rational from irrational players, we consider a more realistic way of capturing the herding behavior of the irrational players. We assume that each irrational player chooses an action played by the majority among all other players, including irrational players. To be precise, we assume that each irrational player chooses the following action, against μ :

$$f(\mu) := \min \left\{ i : i \in \text{Arg max}_{a \in \mathcal{A}} \mu(a) \right\}. \quad (2)$$

In the above, for simplicity and tractability of the analysis, we assume that the action with the smallest index in the set $\text{Arg max}_{a \in \mathcal{A}} \mu(a)$ is preferred in case of a tie.

Before we proceed further, we provide few important remarks regarding the choices made by herding (irrational) players:

(i) The action chosen by the irrational players as per the rule (2) may not be the best response to μ ; it is just a response driven by the herding behaviour of the players. Thus, it is not a rational choice where the players attempt to leverage upon the efforts (or the decisions) of others, as in free-riding (see [12]).

(ii) The second remark is about the possibility of herding players following (only) themselves. In reality, one can have many more variants of irrational behaviors, for example, some players can choose randomly or can have some blind preferences, or can avoid the crowd, etc., and these players might be insignificant in the bigger picture (or in a large population) but can be significant enough to lead (or mis-lead) the herding crowd. Basically, some initial players (whose proportion is negligible at the limit) make some choices and the herding crowd starts following them, eventually leading to all herding players choosing an action different from the choices of rational players. Alternatively, the rational players might estimate an action to be beneficial in the beginning (like in game dynamics, e.g., [16]), irrationals can follow such rational players and themselves. But, later

³ Note that the game is described in the mean-field framework, therefore, the action chosen by a single player does not affect the outcome of the game (see [9]). Given this, it is appropriate to view μ as the empirical distribution corresponding to the ‘entire’ population.

the rational players can find an alternate action to be beneficial in view of the new empirical measure μ . In all such cases, at the limit, it appears that herding players are following themselves. Such possibilities are formalized in [3] for a special game dynamics among (myopic) rational and herding players (see, for example, [3, Theorems 2, 4(ii.b)]).

Finally, the proportion of players choosing different actions in \mathcal{A} is given by:

$$\mu(a) = \alpha\mu^R(a) + (1 - \alpha)1_{\{a=f(\mu)\}}, \text{ for each } a \in \mathcal{A}. \quad (3)$$

The above relationship is obtained as $\alpha\mu^R(a)$ -fraction of rational players choose the action a , and all irrational players choose the same action a only if $a = f(\mu)$ (see (2)).

At this point, one should note that rational players choose an action anticipating the response of the irrational players (as μ depends on $f(\mu)$, see (1) and (3)). Thus, we define a pair (μ, μ^R) to be an equilibrium if it satisfies:

- (i) $\mu^R \in \text{Best Response}(\mu)$;
- (ii) the empirical measure (μ) of the population is given by (3), when the corresponding counterpart for rational players is given by μ^R ; and
- (iii) $f(\mu)$ denotes the majority action as in (2).

The above discussion is formally summarized below:

Definition 1. For $\alpha \in (0, 1]$, a pair of empirical measures (μ, μ^R) is called an α -Rational Nash Equilibrium, or in short, α -RNE, if it satisfies (1), (2) and (3).

Observe that the above definition is a natural extension of the NE defined in the classical mean-field games (MFGs), where ‘all’ players are rational and optimize the utility function $u(\cdot; \mu)$ (see [9]). Thus, (1) is satisfied by μ in MFGs at the NE (not by μ^R), i.e., $\text{support}(\mu) \subseteq \text{Arg max}_{a \in \mathcal{A}} u(a, \mu)$. Due to the presence of herding-players, we will see that an α -RNE need not be a classical NE.

In the coming, we delve deep into the new notion for the game with two actions, where we derive several interesting insights. Further, we also derive the simple conditions to identify the α -RNEs.

3 Game with two actions

Let us consider that players can choose either action 1 or 2, i.e., $\mathcal{A} = \{1, 2\}$. For this setting, we propose simpler notations as follows: (i) let z be the proportion of players who choose $a = 1$ and thus, $\mu = (z, 1 - z)$ and (ii) let y be the proportion of rational players (among rational players) who choose $a = 1$ and thus, $\mu^R = (y, 1 - y)$. As a result, we write the utility function $u(\cdot, \mu)$ as $u(\cdot, z)$.

Finally, the α -RNE is given by $(\mu, \mu^R) \equiv (z, y)$ which satisfy the following:

$$z = \alpha y + (1 - \alpha)1_{\{z \geq \frac{1}{2}\}}, \quad (4)$$

$$\text{support}(\mu^R) \subseteq \text{Arg max}_{a \in \mathcal{A}} u(a, z). \quad (5)$$

Note that in the underlying case with two actions, $f(\mu) = 1$ only if $z \geq 1/2$. Therefore, (2) and (3) together lead to (4) given above.

Now, by (4), α -RNE can be represented only in terms of the proportion z . Denote the set of α -RNEs by \mathcal{N}_α . Then, one can easily verify that the set \mathcal{N}_α has the following structure (recall $\alpha \in (0, 1]$):

$$\mathcal{N}_\alpha = \{z : (z, y^*(z)) \text{ is an } \alpha\text{-RNE}\}, \text{ for } y^*(z) := \begin{cases} \frac{z}{\alpha}, & \text{if } z < \frac{1}{2}, \\ 1 - \frac{1-z}{\alpha}, & \text{if } z \geq \frac{1}{2}. \end{cases} \quad (6)$$

In the above, $y^*(z)$ is provided by solving (4). Thus, by virtue of the above structure, it is sufficient to solve for (5) alone, instead of solving for (4) and (5) simultaneously. Now, observe that $\text{Arg max}_{a \in \mathcal{A}} u(a, z)$ can be simply recognized by comparing the utilities $u(1, z)$ and $u(2, z)$ in the underlying case. This motivates us to define the following *utility difference function*:

$$h(z) := u(1, z) - u(2, z). \quad (7)$$

The idea is to identify the α -RNEs using the zeros of h , i.e., from the set:

$$\mathcal{Z}^h := \{z^* \in [0, 1] : h(z^*) = 0\}. \quad (8)$$

3.1 Identification of equilibria

For the sake of reference, we first provide characterization of the set of classical NEs of MFGs with two actions. The proofs of the next result and the upcoming ones are provided in Appendix A.

Theorem 1 (Identification of MFG-NEs). *Suppose $\mathcal{A} = \{1, 2\}$. Then, the set of classical NEs:*

$$\mathcal{N}_1 \subseteq \mathcal{Z}^h \cup \{0, 1\}. \quad (9)$$

For the converse, we have:

- (i) $\mathcal{Z}^h \subseteq \mathcal{N}_1$,
- (ii) $0 \in \mathcal{N}_1$ only if $h(0) \leq 0$, and
- (iii) $1 \in \mathcal{N}_1$ only if $h(1) \geq 0$. □

The above result provides simple conditions for identifying classical NEs. It states that every zero of h is a NE. Further, 0 and 1 are also NEs if $h(0) < 0$ and $h(1) > 0$, respectively. Next, we identify the set of α -RNEs in terms of \mathcal{N}_1 .

Theorem 2 (Identification of α -RNEs). *Suppose $\mathcal{A} = \{1, 2\}$. Let $O_{a,b} := \{z : a < z < b\}$ denote the open interval. If $\alpha > 1/2$, then $\mathcal{N}_\alpha = \mathcal{N}_1$. Else (i.e., if $\alpha \leq 1/2$), the following statements hold:*

- (i) $\mathcal{N}_\alpha \subseteq \mathcal{N}_1 \cup \{\alpha, 1 - \alpha\} \setminus O_{\alpha, 1-\alpha}$, and
- (ii) *conversely,*
 - (a) $\mathcal{N}_1 \setminus O_{\alpha, 1-\alpha} \subseteq \mathcal{N}_\alpha$,
 - (b) $(1 - \alpha) \in \mathcal{N}_\alpha$ if and only if $h(1 - \alpha) \leq 0$, and

(c) For $\alpha < 1/2$, $\alpha \in \mathcal{N}_\alpha$ if and only if $h(\alpha) \geq 0$. □

Interestingly, the above theorem asserts that *the presence of irrational players has no effect on the set of equilibria* ($\mathcal{N}_\alpha = \mathcal{N}_1$) *when rational players outnumber irrational players* ($\alpha > 1/2$).

More interestingly, the situation drastically differs when $\alpha \leq 1/2$. Firstly, two new equilibria can arise, namely α and $1 - \alpha$, see Theorem 2(i), under the conditions specified in (ii.b) and (ii.c), respectively. Secondly, not every zero of h can be an α -RNE—the zeroes which are only in the interval $[0, \alpha] \cup [1 - \alpha, 1]$ are allowed. Thus, *some classical NEs are deleted and new equilibria are added when the irrational players form the majority*.

Next, for any equilibrium, it is natural to ask if the equilibrium is attainable through some dynamics. In classical setups, best-response dynamics, fictitious play, etc. are known to converge to the NE (under certain conditions). In our mixed-behavioral setup, the turn-by-turn dynamics studied in [3] converges to some/all α -RNEs. In particular, it has been shown in [3, Theorems 1, 3, 4] that when myopic-rational and herding players play a game with two actions one after the other, in a random order and only once, then the empirical distribution of players choosing different actions converges to points that satisfy all the conditions of α -RNE. Thus, such dynamics provides a natural foundation for the new concept of α -RNE.

3.2 Comparison of utilities

In our framework, both rational and irrational players participate in the game. Therefore, first and foremost, one would like to know whether the utility of rational players diminishes due to the presence of irrational players. Subsequently, one might be interested in knowing if the irrational players suffer due to their herding behaviour, when compared with the utility they could have obtained if they were rational.

Further, it is known that NE often results in players gaining lesser utility than the social optimal utility. So happens because the NE provides stability only against unilateral deviations. However, if multiple players deviate from NE, higher utility can possibly be achieved. Considering this, it is plausible that players' utility may be closer to the social optimal utility at α -RNE, than at NE, since irrational players collectively deviate from NE. If this anticipation holds, it suggests a rational inclination towards irrationality. We formally investigate all these aspects below.

Denote the expected utility of a rational player at $z_\alpha^* \in \mathcal{N}_\alpha$ by $u_\alpha^R(z_\alpha^*)$, for $0 < \alpha \leq 1$, and observe:

$$u_\alpha^R(z_\alpha^*) = y^*(z_\alpha^*)u(1, z_\alpha^*) + (1 - y^*(z_\alpha^*))u(2, z_\alpha^*). \quad (10)$$

In the above, $y^*(z_\alpha^*)$ is the probability of a rational player choosing action 1. Similarly, denote the expected utility for an irrational player by $u_\alpha^I(z_\alpha^*)$, and

note that it is given by:

$$u_\alpha^I(z_\alpha^*) = 1_{\{z_\alpha^* \geq \frac{1}{2}\}} u(1, z_\alpha^*) + \left(1 - 1_{\{z_\alpha^* \geq \frac{1}{2}\}}\right) u(2, z_\alpha^*). \quad (11)$$

Recall that our motive is to compare the utilities of the players under α -RNE and classical NE. Keeping this in mind, we aim to investigate if, under α -RNE, players can achieve utility that is comparable to the social optimal utility (u^S) under the classical setting ($\alpha = 1$). Thus, define:

$$u^S := \sup_{z \in [0,1]} (zu(1, z) + (1 - z)u(2, z)). \quad (12)$$

We begin with comparing the utilities for a rational and irrational player under α -RNEs. Further, we compare the utilities with the social optimal utility.

Proposition 1. *For any $\alpha \in (0, 1)$, $u_\alpha^I(z_\alpha^*) \leq u_\alpha^R(z_\alpha^*)$ and $u_\alpha^I(z_\alpha^*) \leq u^S$, for all $z_\alpha^* \in \mathcal{N}_\alpha$. \square*

The above result asserts that rational players always obtain more utility than irrational players. Additionally, the irrational players can never achieve higher utility than the social optimal utility (u^S).

Recall from Theorem 2 that some new equilibria may get added or classical NEs may be deleted in the presence of irrational players. The next result states that if new equilibria (namely, α and $1 - \alpha$) are not added (i.e., $\mathcal{N}_\alpha \subseteq \mathcal{N}_1$), then both the rational players and more importantly, the irrational players attain exactly as much as a rational player gets in the classical setting.

Proposition 2. *For any $\alpha \in (0, 1]$, when $\mathcal{N}_\alpha \subseteq \mathcal{N}_1$, then $u^S \geq u_\alpha^R(z_\alpha^*) = u_\alpha^I(z_\alpha^*) = u_1^R(z_\alpha^*)$ for all $z_\alpha^* \in \mathcal{N}_\alpha$. \square*

Thus, for example, if more rational players are present in the system ($\alpha > 1/2$), then they are able to manipulate the irrational players in such a way that no player loses anything. Further, by the above result, it is clear that no player (not even a rational player) receives more than the social optimal utility (u^S), when $\mathcal{N}_\alpha \subseteq \mathcal{N}_1$.

Now, observe that the above result comments on all α -RNEs, except $z_\alpha^* \in \{\alpha, 1 - \alpha\}$ such that $z_\alpha^* \in \mathcal{N}_\alpha$ but $z_\alpha^* \notin \mathcal{Z}^h$. Thus, if $\mathcal{N}_1 \subseteq \mathcal{N}_\alpha$, then according to Proposition 1, it is evident that the irrational players definitely receive strictly less utility than the rational players at α -RNE. However, several interesting possibilities arise at $z_\alpha^* \in \{\alpha, 1 - \alpha\}$ under said conditions:

- (i) the rational players may outperform the utility attainable in social optimization (when all players are rational and their actions are governed by a central controller), i.e., $u_\alpha^R(z_\alpha^*) > u^S$. This holds for all the examples discussed in the subsequent section;
- (ii) the irrational players may experience no change or a loss ($u_\alpha^I(z_\alpha^*) < u_1^R(z_1^*)$) or a gain ($u_\alpha^I(z_\alpha^*) > u_1^R(z_1^*)$) when compared to some classical NE ($z_1^* \in \mathcal{N}_1$). The first two scenarios occur for the non-atomic routing game, while the latter applies to the other two games discussed in the coming section;

- (iii) interestingly, we shall see in the bandwidth sharing game that both rational and irrational players benefit at α -RNE such that for all $z_1^* \in \mathcal{N}_1$:

$$u_\alpha^R(z_\alpha^*) > u_1^R(z_1^*) \text{ and } u_\alpha^I(z_\alpha^*) > u_1^R(z_1^*).$$

Clearly, the utility of rational and irrational players surpasses and approaches closer to the social optimal utility (u^S) respectively. Thus, in such cases, we may declare that ‘it is rational to be irrational’.

4 Examples

4.1 Non-atomic routing game

Consider a non-atomic routing game, which is a slight modification of Pigou’s network game [14]. One can travel from source (S) to destination (T) via hub 1 or 2. The users must opt for either the path via hub 1 or hub 2 to minimize their travel time.

It takes γz hour(s) to travel from S to T, for some $\gamma \in (1, \infty)$, while the travel time via hub 2 is just 1 hour. Consider that α -fraction of users are rational, and the rest adhere to the majority’s choice (as in (2)). Rational users base their decisions on optimizing the following utility function:

$$u(a, z) = (-\gamma z) 1_{\{a=1\}} + (-1) 1_{\{a=2\}}, \quad (13)$$

where z is the proportion of users travelling via hub 1.

Next, we provide the set of α -RNEs for the above game. The proof is omitted as it directly follows from Theorem 2.

Corollary 1. *Consider the non-atomic routing game and define $\Delta := 1/\gamma$. Then, the set of α -RNEs is given below in two regimes:*

(i) when $\Delta \leq \frac{1}{2}$:

$$\mathcal{N}_\alpha = \begin{cases} \{\alpha, 1 - \alpha\}, & \text{if } \alpha \leq \Delta, \\ \{\Delta, 1 - \alpha\}, & \text{if } \Delta < \alpha \leq \frac{1}{2}, \\ \{\Delta\}, & \text{if } \alpha > \frac{1}{2}; \end{cases}$$

(ii) when $\Delta > \frac{1}{2}$:

$$\mathcal{N}_\alpha = \begin{cases} \{\alpha, 1 - \alpha\}, & \text{if } \alpha \leq 1 - \Delta, \\ \{\Delta, \alpha\}, & \text{if } 1 - \Delta < \alpha < \frac{1}{2}, \\ \{\Delta\}, & \text{if } \alpha \geq \frac{1}{2}. \end{cases} \quad \square$$

At first, note that the classical NE is unique and equals Δ . A new equilibrium emerges when $\alpha > \min\{\Delta, 1 - \Delta\}$. Otherwise, the classical NE is removed, and two new equilibria emerge.

Thus, under classical setting ($\alpha = 1$), Δ -fraction of users choose to travel through hub 1. However, if we consider the game with the rational and irrational

users, then the congestion on the path via hub 1 can either remain the same (as before), or it can decrease to α -level, or increase to $(1 - \alpha)$ -level. Even at times when congestion is lesser, the irrational users are at a loss in this game as⁴ $u_\alpha^I(z_\alpha^*) \leq u_1^R(\Delta)$, for all $z_\alpha^* \in \mathcal{N}_\alpha$.

Rational users exploit the presence of irrational users and benefit in multiple ways:

(i) rational users take an equal or less amount of time to travel, compared to irrational users (as $u_\alpha^R(z_\alpha^*) \geq u_\alpha^I(z_\alpha^*)$ for any $z_\alpha^* \in \mathcal{N}_\alpha$, by Proposition 1).

(ii) *rational users take lesser or equal amount of time to travel than under all-rational case* (as⁵ $u_\alpha^R(z_\alpha^*) \geq u_1^R(\Delta)$, for all $z_\alpha^* \in \mathcal{N}_\alpha$).

(iii) the utility $u_\alpha^R(\alpha)$ is strictly decreasing in α . Thus, if α emerges as the α -RNE, then lower α (more herding players) implies that rational players take less time to reach the destination. In fact, if $\alpha < (1 - \Delta/4)$, then rational users receive more utility at α than the social optimal utility ($u^S = \Delta/4 - 1$).

4.2 Participation game and mechanism design

Motivated by [4], we consider the game where each player has to decide whether to participate or not in an activity. Let us designate $a = 1$ as the action indicating participation and $a = 2$ as the action of non-participation in the activity. Thus, z denotes the proportion of participants.

Each non-participant gets a (perceived) utility equal to 1. To increase participation, the game designer provides a fixed utility $C < 1$ to each participant, and additionally, it offers a reward of $P > 0$, which is equally distributed among all the participants. Hence, the utility function can be expressed as follows:

$$u(a, z) = \left(C + \frac{P}{z} 1_{\{z > 0\}} \right) 1_{\{a=1\}} + 1_{\{a=2\}}. \quad (14)$$

For the above game, the set of α -RNEs depends on the value of P . We consider two disjoint regimes: (i) $P \geq 1 - C$ and (ii) $P < 1 - C$, and present the results for the respective regimes below. The proofs again follow from Theorem 2.

Corollary 2. *Consider the participation game with $P \geq 1 - C$. Then, we have:*

(i) *when $\alpha \in [1/2, 1]$, $\mathcal{N}_\alpha = \mathcal{N}_1 = \{0, 1\}$, and*

(ii) *when $\alpha \in (0, 1/2)$, $\mathcal{N}_\alpha = \mathcal{N}_1 \cup \{\alpha\}$. □*

Under classical NE, dichotomy occurs: either everyone participates or no one participates. This situation may be undesirable for the designer as the chances of zero participation are 50%. Interestingly, when irrational players also play the game and constitute the majority, the designer can exploit the inherent herding

⁴ Here, $u_\alpha^I(z_\alpha^*) = -1$ for $z_\alpha^* \in \{\Delta, \alpha\}$. Further, $u_\alpha^I(1 - \alpha) = -(1 - \alpha)/\Delta < u_1^R(\Delta)$; to verify this, note the conditions when $1 - \alpha \in \mathcal{N}_\alpha$ from Corollary 1.

⁵ Here, $u_1^R(\Delta) = -1 = u_\alpha^R(z_\alpha^*)$ for $z_\alpha^* \in \{\Delta, 1 - \alpha\}$. Further, $u_\alpha^R(\alpha) = -\alpha/\Delta > u_1^R(\Delta)$; to verify this, note the conditions when $\alpha \in \mathcal{N}_\alpha$ from Corollary 1.

behavior of irrationals and possibly induce α -level of participation. Thus, the likelihood of non-zero participation increases due to herding.

Further, the rational players benefit when $\mu = (\alpha, 1 - \alpha)$ is the α -RNE, as they receive higher⁶ utility ($u_\alpha^R(\alpha) = C + P/\alpha$) than the social optimal utility ($u^S = 1 + P$), see (12).

Corollary 3. *Consider the participation game with $P < 1 - C$. Then, we have:*

(i) when $\alpha \in (1/2, 1]$, $\mathcal{N}_\alpha = \mathcal{N}_1 = \left\{0, \frac{P}{1-C}\right\}$,

(ii) when $\alpha = 1/2$,

$$\mathcal{N}_\alpha = \begin{cases} \mathcal{N}_1 \cup \left\{\frac{1}{2}\right\}, & \text{if } P < \frac{1-C}{2}, \\ \mathcal{N}_1, & \text{otherwise, and,} \end{cases}$$

(iii) when $\alpha \in (0, 1/2)$, define $P_1 := \alpha(1 - C)$ and $P_2 := (1 - \alpha)(1 - C)$. Then:

$$\mathcal{N}_\alpha = \begin{cases} \mathcal{N}_1 \cup \{1 - \alpha\}, & \text{if } 0 < P \leq P_1, \\ \{0, 1, \alpha, 1 - \alpha\}, & \text{if } P_1 < P < P_2, \\ \mathcal{N}_1 \cup \{\alpha\}, & \text{if } P_2 \leq P < 1 - C. \end{cases} \quad \square$$

As observed before, here also, either new equilibria are added, or some classical NEs are deleted when $\alpha \leq 1/2$. As one can anticipate, lesser rewards imply lesser utility for the players. When $P < 1 - C$, rational players get lesser utility (at $\alpha, 1 - \alpha$) than the social optimal utility; recall previously, $u_\alpha^R(\alpha) > u^S$.

Importantly, note that *lesser reward means higher participation* under the realistic setting with $\alpha < 1/2$. To be precise, $(1 - \alpha)$ -level (more than 50%) of participation can occur when $P < P_2$ (also compare it with Corollary 2). This happens because rational players tend to lose interest in participation with low reward, so they choose $a = 2$; irrational players (which form the majority) then choose $a = 1$ due to herding.

In all, less reward proves detrimental for the players but advantageous for the game designer.

4.3 Bandwidth sharing game — is it rational to be irrational?

Consider a communication network where players share the bandwidth to transmit their signals/information. The players can either transmit at the maximum capacity (which equals 1) of the shared channel, or they can transmit at a lower level, which equals $1/2$. We refer to the two actions as $a = 1$ and $a = 2$, respectively. In the first case, the communication of others can get interfered with, while in the latter case, no disruption occurs. The overall utility derived by any

⁶ Given $P \geq 1 - C$. Therefore, $\Delta := \frac{P}{1-C} > 1$. Now, observe $u_1^R(\alpha) - u^S = C - 1 - P + \frac{P}{\alpha} > 0$, only if $\alpha < \frac{1}{1 + \frac{1-C}{P}} = \frac{\Delta}{1 + \Delta}$. Since $\alpha \in \mathcal{N}_\alpha$ when $\alpha < \frac{1}{2}$, and $\frac{\Delta}{1 + \Delta} > \frac{1}{2}$, therefore, $u_1^R(\alpha) > u^S$.

player depends upon its maximum capacity discounted by the overall interference caused by the opponents, and hence, the utility function is as follows:

$$u(a, z) = \left(1_{\{a=1\}} + \frac{1}{2} 1_{\{a=2\}} \right) (1 - z). \quad (15)$$

The above game is a simplified version of the bandwidth sharing game discussed in [14] for classical strategic form setup; here we have also modified it for the mean-field setting. Now, we will provide the set of α -RNEs for this game, which can be derived by Theorem 2.

Corollary 4. *Consider the bandwidth-sharing game. The set of α -RNEs is as below:*

- (i) when $\alpha \in [1/2, 1]$, $\mathcal{N}_1 = \mathcal{N}_\alpha = \{1\}$, and
- (ii) when $\alpha \in (0, 1/2)$, $\mathcal{N}_\alpha = \mathcal{N}_1 \cup \{\alpha\}$. □

It is easy to solve the social optimization problem (12) for this game: $u^S = 1/2$, which is realized when no one transmits at capacity 1.

When there are only rational players (i.e., $\alpha = 1$), clearly from (15), the unique classical NE equals $z_1^* = 1$ and the corresponding utility $u_1^R(1) = 0$. Now, consider $\alpha < 1/2$. Then, by (10), (11), one can calculate that $u_\alpha^R(\alpha) = 1 - \alpha$ and $u_\alpha^I(\alpha) = 1/2(1 - \alpha)$.

Interestingly, the rational players have strictly improved their utility compared to the scenario with all rational players, as $u_1^R(1) < u_\alpha^R(\alpha)$. More interestingly, the utility of irrational players is also higher than that of the rational players under classical NE, as $u_1^R(1) < u_\alpha^I(\alpha)$. *Thus, in this case, it is ‘rational’ to be irrational!*

Moreover, as α approaches 0, $u_\alpha^I(\alpha) \rightarrow u^S$. This is a surprising outcome— at a selfish equilibrium, the players are achieving near social optimal utilities. Notably, the existence of a small fraction of rational players and a large fraction of herding players achieves this feat.

As said before, $u_1^R(z_1^*) \leq u^S$ for all $z_1^* \in \mathcal{N}_1$. However, here, $u^S < u_\alpha^R(\alpha)$ due to the presence of irrational players. Conclusively, the introduction of irrationality can be beneficial in some cases.

5 Conclusions

This paper studies the mean-field game involving α -fraction of rational and $(1 - \alpha)$ -fraction of irrational players. While rational players adhere to classical game theory principles, irrational players exhibit herding behavior by blindly choosing the action chosen by the majority. We introduce a novel equilibrium concept, termed α -Rational Nash equilibrium (α -RNE), which extends the NE for the classical mean-field games by capturing the responses of the irrational players.

The analysis in this paper focuses on the games with two actions, which captures many yes/no, this/that like scenarios encountered in daily lives.

Our findings reveal that the presence of irrational players can alter the set of equilibria compared to the set of classical NEs. New equilibria may emerge, while some classical equilibria may disappear when more irrational players are in the system. Otherwise, the set of equilibria does not change. We also provide easy conditions to identify α -RNEs for such games.

The price of anarchy (PoA) is known to be significant in many classical games, and arises due to the rational behaviour of the players involved. With the inclusion of ‘herding-irrational’ players, an immediate interesting question is about the way the PoA manifests. And as one may anticipate, we found some games for which the PoA reduces (in fact to zero) as the fraction of rational players decreases.

Future directions: A key next step is to extend the notion of this paper to more general games, in particular, to games with more than two actions. One can also attempt to formulate an appropriate Stackelberg game, whose outcome matches with the equilibria introduced here. In [3], one special kind of dynamics is studied which converges to a subset of α -RNEs. It will be interesting to consider other behavioral learning dynamics and investigate the properties of their stationary points. Another promising direction would be to examine elaborate mechanism design scenarios involving herding players.

A Appendix

Note: Only in this section, we refer $\text{Arg max}_{a \in \mathcal{A}} u(a, z)$ as \mathcal{A}_u , in short.

Proof of Theorem 1: Before we start the proof, note that when $\alpha = 1$, (4) implies $z = y$. By (5), for any z to be in \mathcal{N}_1 , following should hold:

$$\text{support}(\mu) \subseteq \mathcal{A}_u, \text{ for } \mu = (z, 1 - z). \quad (16)$$

Now, consider any $z \in \mathcal{N}_1$. Then, three different cases arise based on the value of $h(z)$. Suppose $h(z) > 0$. By (7), $\mathcal{A}_u = \{1\}$. Under (16), $\text{support}(\mu) = \{1\}$ (since $z \in \mathcal{N}_1$); thus, $z = 1$. Similarly, one can show that $z = 0$ if $h(z) < 0$. Lastly, suppose $h(z) = 0$. Then, $\mathcal{A}_u = \{1, 2\}$, again by (7). Since $z \in \mathcal{N}_1$, therefore, under (16), either $\text{support}(\mu) = \{1\}$ or $\{2\}$ or $\{1, 2\}$. In the first two cases, $z = 1, 0$ respectively as above. In the last case, $z \in (0, 1)$. Combining all the implications from above, we get (9).

We now prove the claims for the converse of (9). Suppose $z \in \mathcal{Z}^h \cap (0, 1)$. Then, $\text{support}(\mu) = \{1, 2\} = \mathcal{A}_u$. Thus, $z \in \mathcal{N}_1$. Next, suppose $z = 0$. Then, $\text{support}(\mu) = \{2\}$. If $h(0) \leq 0$, then $\mathcal{A}_u = \{1, 2\}$ or $\{2\}$; thus, $0 \in \mathcal{N}_1$. However, if $h(0) > 0$, then $\mathcal{A}_u = \{1\}$. Clearly, it contradicts (16) and thus, $0 \notin \mathcal{N}_1$ in this case. One can similarly prove the claim for $z = 1$. Conclusively, (i)-(iii) hold. \square

Proof of Theorem 2: We divide the proof into two cases.

Case 1. when $\alpha > \frac{1}{2}$

Claim 1: $\mathcal{N}_\alpha \subseteq \mathcal{N}_1$

Suppose $z \in \mathcal{N}_\alpha$. Firstly, let $h(z) > 0$. By (7), $\mathcal{A}_u = \{1\}$. Under (5), $\text{support}(\mu^R) = \{1\}$; thus, $y^*(z) = 1$. By (6), $z = 1$ (as $\alpha > \frac{1}{2}$). Thus, $z = 1 \in \mathcal{N}_1$,

as $\{1\} \subseteq \mathcal{N}_1$ (see Theorem 1). Similarly, one can show that $z = 0 \in \mathcal{N}_1$ when $h(z) < 0$. Further, if $h(z) = 0$, then, $z \in \mathcal{Z}^h$. By Theorem 1(i), it is clear that $z \in \mathcal{N}_1$. In all, $\mathcal{N}_\alpha \subseteq \mathcal{N}_1$.

Claim 2: $\mathcal{N}_1 \subseteq \mathcal{N}_\alpha$

Let $z \in \mathcal{N}_1$. Then by (9), either $z \in \mathcal{Z}^h$ or $z \in \{0, 1\}$. Say $z \in \mathcal{Z}^h \cap [0, \frac{1}{2})$. Here, $y^*(z) = \frac{z}{\alpha}$; observe $y^*(z) = 0$ if $z = 0$ and $y^*(z) \in (0, 1)$ otherwise. Thus, $\text{support}(\mu^R) = \{2\}$ or $\{1, 2\} \subseteq \mathcal{A}_u$, as $z \in \mathcal{Z}^h$. By Definition 1, $z \in \mathcal{N}_\alpha$. One can prove in a similar manner that any $z \in \mathcal{Z}_\alpha^h \cap [\frac{1}{2}, 1)$ is also in \mathcal{N}_α .

Now, say $z = 0$. Observe $y^*(0) = 0$. Thus, $\text{support}(\mu^R) = \{2\}$. Recall $z = 0 \in \mathcal{N}_1$ only if $h(0) \leq 0$, by Theorem 1(ii). Thus, $\mathcal{A}_u = \{1, 2\}$ or $\{2\}$. Then, as above, $z \in \mathcal{N}_\alpha$. The proof similarly follows when $z = 1$. Hence, $\mathcal{N}_1 \subseteq \mathcal{N}_\alpha$.

Conclusively, $\mathcal{N}_1 = \mathcal{N}_\alpha$.

Case 2. when $\alpha \leq \frac{1}{2}$

(i) Say $z \in \mathcal{N}_\alpha$. Then, z satisfies (5). Now, we divide the proof based on the values of $h(z)$.

Firstly, let $h(z) > 0$. Then, as in the case with $\alpha > \frac{1}{2}$, $y^*(z) = 1$. By (6), either $z = \alpha$ if $\alpha < \frac{1}{2}$ or $z = 1$. Observe that $\alpha \notin \mathcal{N}_1$ and $\{1\} \subseteq \mathcal{N}_1$. This implies that $z \in \mathcal{N}_1 \cup \{\alpha, 1 - \alpha\} \setminus (\alpha, 1 - \alpha)$. Similarly, one can prove (??) when $h(z) < 0$. At last, let $h(z) = 0$. Then, $\mathcal{A}_u = \{1, 2\}$, again by (7). Under (5), three possibilities arise:

- $\text{support}(\mu^R) = \{1\}$: here, $z = \alpha$ if $\alpha < \frac{1}{2}$ or $z = 1$, as when $h(z) > 0$.
- $\text{support}(\mu^R) = \{2\}$: here, $z = 1 - \alpha$ if $\alpha \leq \frac{1}{2}$ or $z = 0$, as when $h(z) < 0$.
- $\text{support}(\mu^R) = \{1, 2\}$: here, $y^* \in (0, 1)$. This implies that either $z < \alpha$ and $z < \frac{1}{2}$, or $z > 1 - \alpha$ and $z \geq \frac{1}{2}$.

In all three cases, one can easily see that $z \in \mathcal{N}_1 \cup \{\alpha, 1 - \alpha\} \setminus (\alpha, 1 - \alpha)$. Conclusively, (i) holds.

(ii.a) Suppose $z \in \mathcal{N}_1 \setminus (\alpha, 1 - \alpha)$. Then by (9), $z \in \mathcal{Z}^h \setminus (\alpha, 1 - \alpha)$ or $z \in \{0, 1\}$.

Firstly, say $z \in \mathcal{Z}^h \setminus (\alpha, 1 - \alpha)$ such that $z < \frac{1}{2}$. Here, $y^*(z) = \frac{z}{\alpha}$; observe $y^*(z) = 1$ if $z = \alpha$ and $y^*(z) < 1$ otherwise. Thus, $\text{support}(\mu^R) = \{1\}$ or $\{1, 2\}$; in either case, $\text{support}(\mu^R) \subseteq \mathcal{A}_u$, as $z \in \mathcal{Z}^h$. This implies that z satisfies (5). By Definition 1, $z \in \mathcal{N}_\alpha$. One can prove in a similar manner that any $z \in \mathcal{Z}^h \cap [\frac{1}{2}, 1) \setminus (\alpha, 1 - \alpha)$ is also in \mathcal{N}_α .

When $z = 0$ or 1 , the claim holds exactly as in the case with $\alpha > 1/2$.

(ii.b) Let $\alpha < 1/2$. Say $\alpha \in \mathcal{N}_\alpha$. Then, $y^*(\alpha) = 1$, and hence $\text{support}(\mu^R) = \{1\} \subseteq \mathcal{A}_u$ only if $h(\alpha) \geq 0$. Conversely, say $h(\alpha) \geq 0$. Then, $\mathcal{A}_u = \{1, 2\}$ or $\{1\}$. It is now easy to observe that $\alpha \in \mathcal{N}_\alpha$ as $y^*(\alpha) = 1$ ensures (5) is satisfied.

(ii.c) The proof follows as in part (ii.c). □

Proof of Proposition 1: Firstly, consider $z^* \in \mathcal{N}_\alpha \cap \mathcal{Z}^h$. Then, by definition of \mathcal{Z}^h , $u(1, z^*) = u(2, z^*)$. Thus, note from (10), (11) that $u_\alpha^R(z^*) = u(1, z^*) = u_\alpha^L(z^*)$. Secondly, consider $z^* = 0 \in \mathcal{N}_\alpha \setminus \mathcal{Z}^h$. Then, $y^*(0) = 0$, and thus, $u_\alpha^R(0) =$

$u(2, 0) = u_\alpha^I(0)$. One can similarly prove that $u_\alpha^R(1) = u(1, 1) = u_\alpha^I(1)$, when $1 \in \mathcal{N}_\alpha \setminus \mathcal{Z}^h$.

Next, suppose $z^* = \alpha \in \mathcal{N}_\alpha \setminus \mathcal{Z}^h$. Then, by Theorem 2, $h(\alpha) > 0$, i.e., $u(1, \alpha) > u(2, \alpha)$, and $\alpha < \frac{1}{2}$. The latter implies that $y^*(\alpha) = 1$; thus:

$$u_\alpha^R(\alpha) = u(1, \alpha) > u(2, \alpha) = u_\alpha^I(\alpha).$$

Lastly, one can similarly prove the claim for $1 - \alpha \in \mathcal{N}_\alpha \setminus \mathcal{Z}^h$.

Define $f(z) := zu(1, z) + (1 - z)u(2, z)$ for all $z \in [0, 1]$. Consider any $z^* \in \mathcal{N}_\alpha$, then, $u^S = \sup_{z \in [0, 1]} f(z) = \max \left\{ \sup_{z \in [0, 1] - \{z^*\}} f(z), f(z^*) \right\} \geq f(z^*) \geq u_\alpha^I(z^*)$, see (12). \square

Proof of Proposition 2: Suppose $\alpha > \frac{1}{2}$. By Theorem 2, $\mathcal{N}_\alpha = \mathcal{N}_1$. Further, $u^S \geq u_\alpha^R(z^*) = u_\alpha^I(z^*)$, for each $z^* \in \mathcal{N}_\alpha$, see proof of Proposition 1. From (10):

$$u_1^R(z^*) = \begin{cases} u(2, z^*), & \text{if } z^* = 0 \in \mathcal{N}_\alpha \setminus \mathcal{Z}^h, \\ u(1, z^*), & \text{otherwise.} \end{cases}$$

Then, observe $u_1^R(z^*) = u_\alpha^I(z^*)$, again from the proof of Proposition 1. This completes the proof for $\alpha > \frac{1}{2}$.

Next, consider $\alpha \leq \frac{1}{2}$. Then, $\mathcal{N}_\alpha \subseteq \mathcal{N}_1$ only if $\alpha, 1 - \alpha \notin \mathcal{N}_\alpha$, see Theorem 2. Thus, the proof follows as in case with $\alpha > \frac{1}{2}$. \square

References

1. Abdallah, M., Cason, T., Bagchi, S., Sundaram, S.: The effect of behavioral probability weighting in a sequential defender-attacker game. In: 2020 59th IEEE Conference on Decision and Control (CDC). pp. 3255–3260 (2020)
2. Abdallah, M., Naghizadeh, P., Hota, A.R., Cason, T., Bagchi, S., Sundaram, S.: The impacts of behavioral probability weighting on security investments in interdependent systems. In: 2019 American Control Conference (ACC). pp. 5260–5265. IEEE (2019)
3. Agarwal, K., Avrachenkov, K., Vyas, R., Kavitha, V.: Two choice behavioral game dynamics with myopic-rational and herding players. Proceedings of the ACM on Measurement and Analysis of Computing Systems (2025), <https://inria.hal.science/hal-04883482>
4. Agarwal, K., Kavitha, V.: Single-out fake posts: participation game and its design. In: 2023 American Control Conference (ACC). pp. 2344–2350. IEEE (2023)
5. Banerjee, A.V.: A simple model of herd behavior. The Quarterly Journal of Economics **107**(3), 797–817 (1992)
6. Bénabou, R., Tirole, J.: Intrinsic and extrinsic motivation. The Review of Economic Studies **70**(3), 489–520 (2003)
7. Bonau, S.: A case for behavioural game theory. Journal of Game Theory **6**(1), 7–14 (2017)
8. Camerer, C.F.: Behavioral game theory: Experiments in strategic interaction. Princeton university press (2011)
9. Carmona, R., Delarue, F.: Probabilistic Theory of Mean Field Games with Applications I-II. Springer (2018)

10. Eliaz, K.: Fault tolerant implementation. *The Review of Economic Studies* **69**(3), 589–610 (2002)
11. Eyster, E., Rabin, M.: Rational and naive herding (2009)
12. Feldman, M., Chuang, J.: Overcoming free-riding behavior in peer-to-peer systems. *ACM sigecom exchanges* **5**(4), 41–50 (2005)
13. Morone, A., Samanidou, E.: A simple note on herd behaviour. *Journal of Evolutionary Economics* **18**, 639–646 (2008)
14. Narahari, Y.: *Game Theory and Mechanism Design*, vol. 4. World Scientific (2014)
15. Noti, G.: From behavioral theories to econometrics: Inferring preferences of human agents from data on repeated interactions. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 35, pp. 5637–5646 (2021)
16. Sandholm, W.H.: *Population Games and Evolutionary Dynamics*. MIT Press (2010)
17. Sanjab, A., Saad, W., Başar, T.: Prospect theory for enhanced cyber-physical security of drone delivery systems: A network interdiction game. In: *2017 IEEE international conference on communications (ICC)*. pp. 1–6 (2017)
18. Schultz, W.: Introduction. *neuroeconomics: The promise and the profit*. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**(1511), 3767–3769 (2008)
19. Thaler, R.H.: From cashews to nudges: The evolution of behavioral economics. *American Economic Review* **108**(6), 1265–1287 (2018)
20. Vasal, D., Berry, R.: Alpha-robust equilibrium in anonymous games. Available at SSRN 3643821 (2020)
21. Vasal, D., Berry, R.: Fault tolerant equilibria in anonymous games: best response correspondences and fixed points. *arXiv preprint arXiv:2005.06812* (2020)
22. Verginer, L., Vaccario, G., Ronzani, P.: The robotic herd: Using human-bot interactions to explore irrational herding. *SocArXiv* (2023)