Economics Letters 255 (2025) 112510

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Conformism across games

Roberto Rozzi

University of Siena, Department of Economics and Statistics, Siena Piazza San Francesco 7, 53100, Italy

ARTICLE INFO

JEL classification:
Codes
C73
D74
D83

Keywords:
Conformism
Local stability
Behavioral rules

ABSTRACT

I study a population of conformist and rational players playing 2x2 games and calculate the locally stable fraction of conformists. I evaluate the fitness of each behavioral rule in all Nash Equilibria for each population share, discounting a cognitive cost to rational players. I find that conformists outperform rational players when in the minority because, in that case, the equilibrium is such that all strategies yield the same payoff. If the cognitive cost for rational players is sufficiently large, the only locally stable population composition is one in which each behavioral rule plays a different pure strategy in equilibrium.

1. Introduction

Empirical and experimental evidence shows that individuals are susceptible to conformity pressure even in situations that do not inherently require alignment (e.g., Mascagni, 2018; Farrow et al., 2017) and that people differ in their responsiveness to conformity (Efferson et al., 2015; Boucher et al., 2024; Rasooly and Rozzi, 2024). Despite extensive evidence of this heterogeneity, however, it is less clear why some individuals have a strong taste for conformity while others do not.

In this paper, I provide an evolutionary argument to explain the emergence of conformism across anti-coordination and conflict games (i.e., contexts not inherently requiring alignment). In the stable population composition, conformists can never be in the minority in these games because when the majority of agents are sophisticated, they push the behavior of all other agents towards an equilibrium where each strategy pays the same, which makes following the majority a lossless strategy. Since being more sophisticated involves a cognitive cost, the behavior of conformists always pays more.

Among the many contributions in the literature on conflict and anti-coordination games, Herold and Kuzmics (2020) is the closest to my paper. Their study demonstrates that individuals can use labels to establish roles in conflict situations to reduce costly disputes. My paper shows that individuals may take different roles depending on the behavioral rules they adopt. More broadly, my work is connected to papers studying the evolution of mental models. The pioneering work by Stahl (1993) initiated a series of studies proving that agents with different levels of sophistication might coexist in strategic contexts. Mohlin (2012) extends this result to theories of mind, while Heller (2015) proves a similar result when considering agents with different foresight abilities. Heifetz et al. (2007) shows that players can benefit from distorting the payoffs of the underlying game when making a decision. In my paper, I find that a less sophisticated rule (conformism) might perform better than a more sophisticated rule when the latter suffers a cognitive cost. I also show that behavioral rules might be inferred from the strategies in equilibrium.

My paper connects with works following the indirect evolutionary approach (Güth and Yaari, 1992; Güth, 1995) for studying the evolution of preferences that differ from material payoffs (see Alger and Weibull, 2019; Alger, 2023, for recent reviews). According to this literature, such preferences may or may not emerge depending on the observability of other players' preferences (Ok and Vega-Redondo, 2001; Ely and Yilankaya, 2001; Dekel et al., 2007) and the degree of assortativity of players' matching (Alger and Weibull, 2013; Alger et al., 2020). Similarly to papers like Ely and Yilankaya (2001) or Dekel et al. (2007), in my model, other players' preferences are not observable and players are randomly matched. Like those papers, the equilibrium action frequency includes that of the NE of the underlying game should the cognitive cost borne by rationals be sufficiently small. Unlike those papers, in my model, because rationals bear a cognitive cost, the number of conformists might become large enough to induce an equilibrium action frequency that differs from that of the NE of the underlying game.

My work also relates to papers studying the evolution of behavioral rules. Many studies from this literature show that humans (Duersch et al., 2012; Dong et al., 2015; LiCalzi and Mühlenbernd, 2019; Alós-Ferrer and Ritschel, 2021) and other animals (Laland, 2004; Rendell et al., 2010; Dridi and Lehmann, 2015) can benefit from imitating similar individuals in various contexts. I show that rigidly following the

E-mail address: roberto.rozzi@unisi.it.

https://doi.org/10.1016/j.econlet.2025.112510
Received 20 March 2025; Received in revised form 16 July 2025; Accepted 16 July 2025
Available online 11 August 2025
0165-1765/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

majority might lead to the same results as being rational even in anti-coordination or conflict games. Thus, if being more sophisticated comes with a cognitive cost to bear, conformity might provide an evolutionary advantage.

2. Model

I study a model where players are randomly matched in pairs to play the given 2×2 symmetric game parametrized by the payoffs $p \in (1/2, 1)$ and $q \in (0, 1]$. Depending on p and q, the game might be one of anti-coordination or conflict (this distinction does not affect the results).

	H	D
H	0,0	p,q
D	q,p	1-p,1-p

I consider a continuum of players. Each player is either one of two types: C or R. Let $\alpha \in [0, 1]$ denote the fraction of type C. Throughout the paper, I will refer to C types as conformists and R types as rationals. I call $\sigma_{H|C} \in [0, 1]$ the fraction of C types playing H and by $\sigma_{H|R} \in [0, 1]$ the fraction of R types playing H. The total fraction of players playing H is $\sigma_H = \alpha\sigma_{H|C} + (1 - \alpha)\sigma_{H|R}$.

Let $\hat{\sigma}_H$ be the expected fraction of players playing H. Types choose strategies according to the following utility functions:

$$U_C(H, \hat{\sigma}_H) = \hat{\sigma}_H, \quad U_C(D, \hat{\sigma}_H) = 1 - \hat{\sigma}_H;$$

$$U_R(H, \hat{\sigma}_H) = (1 - \hat{\sigma}_H)p, \quad U_R(D, \hat{\sigma}_H) = \hat{\sigma}_H q + (1 - \hat{\sigma}_H)(1 - p).$$

Let $\boldsymbol{\sigma}^* = (\sigma^*_{H|C}, \sigma^*_{H|R})$ be a Nash Equilibrium (NE) implying a value σ^*_H. Throughout the paper, I refer to a type-monomorphic NE, as a NE where each type plays a different pure strategy. Importantly, a player's type can be inferred from their strategy in a type-monomorphic NE. I calculate the NE for every value of α, and then evaluate the fitnesses of each type in every NE as follows:

$$\overline{\Pi}_C(\sigma^*_{H|C}, \sigma^*_H) = \sigma^*_{H|C}\left((1 - \sigma^*_H)p\right) + (1 - \sigma^*_{H|C})\left(\sigma^*_H q + (1 - \sigma^*_H)(1 - p)\right),$$

$$\begin{aligned}\overline{\Pi}_R(\sigma^*_{H|R}, \sigma^*_H) \;=\; & \sigma^*_{H|R}\left((1 - \sigma^*_H)p\right) + (1 - \sigma^*_{H|R}) \\ & \times \left(\sigma^*_H q + (1 - \sigma^*_H)(1 - p)\right) - \kappa.\end{aligned}$$

The fitness of R types is discounted by $\kappa > 0$. Such a cost represents the cognitive effort associated with engaging in a more demanding task, since rationals compute both $\hat{\sigma}_H$ and the payoffs of the game, while conformists only compute $\hat{\sigma}_H$.

The main analysis involves finding locally stable values of α. Following an evolutionary logic, $\alpha^* \in (0, 1)$ is locally stable if the following three conditions hold:

1. $\overline{\Pi}_C(\sigma^*_{H|C}, \sigma^*_H) = \overline{\Pi}_R(\sigma^*_{H|R}, \sigma^*_H)$ for $\alpha = \alpha^*$;
2. $\overline{\Pi}_C(\sigma^*_{H|C}, \sigma^*_H) > \overline{\Pi}_R(\sigma^*_{H|R}, \sigma^*_H)$ for $\alpha < \alpha^*$;
3. $\overline{\Pi}_C(\sigma^*_{H|C}, \sigma^*_H) < \overline{\Pi}_R(\sigma^*_{H|R}, \sigma^*_H)$ for $\alpha > \alpha^*$.

For $\alpha^* = 1$, only the second condition must hold, while for $\alpha^* = 0$, only the third.

3. Results

The following proposition collects the main results of the paper about α^* and σ^*_H.

Proposition 1. *Let $\overline{\sigma}$ be such that $U_R(H, \overline{\sigma}) = U_R(D, \overline{\sigma})$.*

1. *For all $\kappa > 0$, $\alpha^* \in [1/2, 1]$.*
2. *If $\overline{\sigma} > 1/2$,*
   - *(a) if $0 < \kappa < 2p - 1 - q$, then $\alpha^* \in [1/2, \overline{\sigma})$ and $\sigma^*_H \in \{\overline{\sigma}, 1 - \alpha^*\}$;*
   - *(b) if $\kappa \geq 2p - 1 - q$, then $\alpha^* \in [\overline{\sigma}, 1]$ and $\sigma^*_H \in \{\alpha^*, 1 - \alpha^*\}$.*
3. *If $\overline{\sigma} < 1/2$,*
   - *(a) if $0 < \kappa < q + 1 - 2p$, then $\alpha^* \in [1/2, 1 - \overline{\sigma})$ and $\sigma^*_H \in \{\overline{\sigma}, \alpha^*\}$;*
   - *(b) if $\kappa \geq q + 1 - 2p$, then $\alpha^* \in [1 - \overline{\sigma}, 1]$ and $\sigma^*_H \in \{\alpha^*, 1 - \alpha^*\}$.*

Population composition. Proposition 1 shows that conformists are never the minority in α^* and can become more than the majority should κ be large enough. I prove the result in the Appendix and provide an intuition here. If $\alpha < 1/2$, conformists perform better than rationals since $\kappa > 0$. This result arises because, whenever $\alpha < 1/2$, rationals influence the NE towards a mixed-strategy one, where all strategies give the same payoff. Thus, since playing rationally involves a cognitive cost, conformity pays more. When conformists become the majority, they may drive the selection of the NE towards type-monomorphic equilibria in which conformists all play either H or D, and rationals best respond by all playing D or H, respectively. Specifically, for $1/2 \leq \alpha < \max\{\overline{\sigma}, 1 - \overline{\sigma}\}$, there is a multiplicity of equilibria between a mixed NE and the type-monomorphic ones (see Proof of Proposition 1), while for $\alpha > \max\{\overline{\sigma}, 1 - \overline{\sigma}\}$ only type-monomorphic NE are possible. Since rationals earn a better material payoff in the type-monomorphic equilibria, how large α^* is compared to $1/2$, depends on κ.

Equilibrium action frequencies. σ^*_H depends on α^*, and thus, on κ. The multiplicity of NE for $\alpha > 1/2$ prevents clear predictions for the equilibrium action frequency. However, from Proposition 1, it emerges that for κ reasonably low, $\overline{\sigma}$ (i.e., the mixed NE of the underlying game) is always included in the set of equilibrium action frequencies (in line with Ely and Yilankaya, 2001; Dekel et al., 2007), while as κ increases, $\overline{\sigma}$ ceases to be in such a set. This shift occurs because conformists become more prevalent in the population. Since such types play blindly following the majority, they play H or D depending on the realized NE. These results could have important welfare implications. Indeed, conformists do not play the "rational" strategy: their behavior induces a σ^*_H, that compared to $\overline{\sigma}$, increases cooperation levels in case $\sigma^*_{H|C} = 0$ and $\sigma^*_H = 1 - \alpha^*$, and decreases such levels if $\sigma^*_{H|C} = 1$ and $\sigma^*_H = \alpha^*$.

Type-monomorphic strategy distribution. One important consideration about the cognitive cost borne by rationals is that, should κ be sufficiently large, we could infer the behavioral rule of a player based on their strategies. Indeed, conformists are the predominant behavioral rule for $\kappa > \max\{2p-1-q;\ q+1-2p\}$, and thus, $\alpha^* \geq \max\{\overline{\sigma}, 1-\overline{\sigma}\}$. Since the majority of players rigidly follow the crowd, they all blindly play H, or D (depending on the NE). More sophisticated agents best respond to conformists' strategy by playing the opposite strategy (i.e., D, or H). Thus, for $\kappa > \max\{2p - 1 - q;\ q + 1 - 2p\}$, the NE is either $\boldsymbol{\sigma}^* = (1, 0)$, or $\boldsymbol{\sigma}^* = (0, 1)$. In these cases, we can distinguish conformists from rationals simply by observing which strategy the majority plays. This result is summarized in the next corollary.

Corollary 1. *If $\kappa > \max\{2p-1-q;\ q+1-2p\}$, α^* is such that $\boldsymbol{\sigma}^* = (1, 0)$ or $\boldsymbol{\sigma}^* = (0, 1)$.*

Connection to Harsanyi-style purification. It is worth mentioning that the result in Proposition 1 connects to the theoretical insights on mixed-strategy equilibria originally provided by Harsanyi (1973). Harsanyi proved that mixed strategies in games with complete information can be interpreted as the limit of pure strategies in games with incomplete information. When players face small private uncertainty about their own payoffs, they respond deterministically, leading to a pure-strategy Bayesian equilibrium. As the uncertainty vanishes, the distribution of play converges to the original mixed-strategy equilibrium. Sandholm (2007) extended this logic by characterizing the evolutionary stability of such equilibria. My model offers a complementary perspective: mixed-strategy profiles emerge from the coexistence of distinct behavioral rules, each selecting a different strategy in equilibrium. In the

stable population composition (α^*), each behavioral rule selects a different strategy, yielding a mixed-strategy profile at the aggregate level even if each behavioral rule prescribes a pure strategy in equilibrium (e.g., $\sigma_H^* \in (0,1)$).

4. Discussion

In this paper, I provided a model to explain the emergence of conformist behavior outside coordination games. I showed that, even in anti-coordination and conflict games, rigidly following the majority pays more than being rational. Indeed, as long as conformists are the minority, rationals push towards an equilibrium where all strategies give the same payoff, and thus, less cognitively demanding behavioral rules (such as conformism) pay more. Due to this mechanism, conformism emerges in anti-coordination and conflict games, leading to people conforming to norms across different contexts.

My results raise further questions: how intensely does conformism grow in individuals? Which types of conformism grow in individuals? Research on social pressure shows that individuals perceive social pressure differently depending on the context (Boucher et al., 2024). Thus, future work could provide evolutionary insights on why humans are influenced by their peers' decisions differently depending on the context and how much they become influenced by those decisions. This insight is crucial for policymakers, as incentive-based interventions — such as those addressing climate change — may not only alter behavior but also reshape underlying behavioral rules, with long-term consequences.

Acknowledgments

I deeply thank Pietro Dindo and Ennio Bilancini for enlightening comments and suggestions during and after the supervision of my thesis. I wish to express my gratitude to Marco LiCalzi, Heinrich Nax, and Salvatore Modica for their useful reviews of my thesis. I thank Joseph E. Harrington, the editor in charge of the submission, as well as an anonymous referee for enlightening comments during the submission process. I thank Federico Innocenti, Jonathan Newton, and Itzhak Rasooly for their comments and feedback on the paper, Michael Kopel for his discussion at Oligo 2023, and all the participants at that conference. Furthermore, I also thank all audiences and participants at The Lisbon Meetings in Game Theory and Applications, The Conference on Economic Design 2023, CEPET Workshop 2023, XVII Grass Workshop, and ASSET Meeting 2023.

Appendix. Proofs

Proof of Proposition 1.

Nash Equilibria

I start by considering $\overline{\sigma} > 1/2$. In such a case, for $\alpha \in [0, 1/2]$, there exists only one NE such that $\sigma^*_{H|C} = 1$ and $\sigma^*_{H|R} = \frac{\overline{\sigma}-\alpha}{1-\alpha}$. To verify that this is an equilibrium, note that $U_C(H,\overline{\sigma}) > U_C(D,\overline{\sigma})$ for $\overline{\sigma} > 1/2$ and that $U_R(H,\overline{\sigma}) = U_R(D,\overline{\sigma})$. To show that there are no other NE, note that, it must be that $\sigma^*_{H|C} = 1$ or $\sigma^*_{H|C} = 0$. Moreover, $\forall \sigma_H < \overline{\sigma}$ $U_R(H,\sigma_H) > U_R(D,\sigma_H)$ and $\forall \sigma_H > \overline{\sigma}$ $U_R(H,\sigma_H) < U_R(D,\sigma_H)$. Thus, $\nexists \boldsymbol{\sigma}^*$ such that all R best reply to $\boldsymbol{\sigma}^*$ if not for the one such that $\sigma^*_H = \overline{\sigma}$.

For $\alpha \in (1/2, \overline{\sigma})$, $\sigma^* = \left(1, \frac{\overline{\sigma}-\alpha}{1-\alpha}\right)$ is still a NE, but also $\sigma^* = (0,1)$ is a NE. Indeed, $U_C(H, 1-\alpha) < U_C(D, 1-\alpha)$ and $U_R(H, 1-\alpha) > U_R(D, 1-\alpha)$ if $\alpha > \frac{1}{2}$. For $\alpha \in [\overline{\sigma}, 1]$, $\sigma^* = (0,1)$ is a NE but $\sigma^* = \left(1, \frac{\overline{\sigma}-\alpha}{1-\alpha}\right)$ is not. Moreover, given that $\alpha > \overline{\sigma} > 1/2$, $U_C(H,\alpha) > U_C(D,\alpha)$ and $U_R(H,\alpha) < U_R(D,\alpha)$. Thus, also $\sigma^* = (1,0)$ is a NE.

The intuition for $\overline{\sigma} < 1/2$ is similar but the NE are different. For $\alpha \in [0, 1/2]$, the only NE is $\left(0, \frac{\overline{\sigma}}{1-\alpha}\right)$. For $\alpha \in (1/2, 1-\overline{\sigma})$, the are two NE: $\left(0, \frac{\overline{\sigma}}{1-\alpha}\right)$ and $(1,0)$. For $\alpha \in [1-\overline{\sigma}, 1]$, the are two NE: $(0,1)$ and $(1,0)$.

Local stability of α.

I start the proof by considering $\alpha < 1/2$. For such values of α, $\sigma^*_H = \overline{\sigma}$ in all NE. Since for $\sigma_H = \overline{\sigma}$, the average payoff for H is equal to the one for D, for all $\alpha\sigma^*_{H|C} + (1-\alpha)\sigma^*_{H|R} = \overline{\sigma}$,

$$\overline{\Pi}_R(\sigma^*_{H|R}, \overline{\sigma}) = \overline{\Pi}_C(\sigma^*_{H|C}, \overline{\sigma}) - \kappa.$$

Thus, $\alpha^* \in [\frac{1}{2}, 1]$ for all $\kappa > 0$. Next, I consider the fitnesses in the two NE $(1,0)$ and $(0,1)$

$$\overline{\Pi}_R(0,\alpha) = \alpha q + (1-\alpha)(1-p) - \kappa, \quad \overline{\Pi}_C(1,\alpha) = (1-\alpha)p;$$

$$\overline{\Pi}_R(1,1-\alpha) = \alpha p - \kappa, \quad \overline{\Pi}_C(0,1-\alpha) = (1-\alpha)q + \alpha(1-p).$$

Note that for $(1,0)$ and $(0,1)$,

1. $\overline{\Pi}_R(0,\alpha) + \kappa > \overline{\Pi}_C(1,\alpha)$;
2. $\overline{\Pi}_R(1,1-\alpha) + \kappa > \overline{\Pi}_C(0,1-\alpha)$.

Moreover, there exists, and it is unique, an α^* such that

1. $\overline{\Pi}_R(0,\alpha^*) = \overline{\Pi}_C(1,\alpha^*)$;
2. $\overline{\Pi}_R(0,\alpha) > \overline{\Pi}_C(1,\alpha)$ if $\alpha > \alpha^*$;
3. $\overline{\Pi}_R(0,\alpha) < \overline{\Pi}_C(1,\alpha)$ if $\alpha \in (1/2, \alpha^*]$.

The same can be said for $\overline{\Pi}_R(1,1-\alpha)$ and $\overline{\Pi}_C(0,1-\alpha)$. Thus, once we find α^*, we know that such a solution is the unique satisfying the conditions for local stability. More precisely,

$$\overline{\Pi}_R(0,\alpha) = \overline{\Pi}_C(1,\alpha) \text{ for } \alpha^* = \frac{\kappa + 2p - 1}{2p + q - 1}; \tag{1}$$

$$\overline{\Pi}_R(1,1-\alpha) = \overline{\Pi}_C(0,1-\alpha) \text{ for } \alpha^* = \frac{\kappa + q}{2p + q - 1}. \tag{2}$$

From (1) and (2), we can find a κ_1 and a κ_2 such that for $\kappa > \max\{\kappa_1, \kappa_2\}$ conformists are always better off than rationals for $\alpha \in (1/2, \max\{\overline{\sigma}, 1-\overline{\sigma}\}]$ and might be better than rationals for $\alpha \in (\max\{\overline{\sigma}, 1-\overline{\sigma}\}, 1]$. Such values should be such that $\frac{\kappa+q}{2p+q-1} > \frac{2p-1}{2p+q-1} = \overline{\sigma}$ and $\frac{\kappa+2p-1}{2p+q-1} > \frac{q}{2p+q-1} = 1-\overline{\sigma}$. Hence, $\kappa_1 = 2p-1-q$ and $\kappa_2 = q+1-2p$.

Moreover, $\kappa_1 < 0$ and $\kappa_2 > 0$ for $\overline{\sigma} < 1/2$ and vice-versa for $\overline{\sigma} > 1/2$. Thus, it can never be that $\kappa < \min\{\kappa_1, \kappa_2\}$. We can only conclude that $\alpha^* \in [1/2, \max\{\overline{\sigma}, 1-\overline{\sigma}\})$ and $\sigma^*_H \in \{\overline{\sigma}, 1-\alpha^*\}$ for $\kappa \in (0, \kappa_1)$ for $\overline{\sigma} > 1/2$, while $\alpha^* \in [1/2, \max\{\overline{\sigma}, 1-\overline{\sigma}\})$ and $\sigma^*_H \in \{\overline{\sigma}, \alpha^*\}$ for $\kappa \in (0, \kappa_2)$ for $\overline{\sigma} < 1/2$. □

Data availability

No data was used for the research described in the article.

References

Alger, I., 2023. Evolutionarily stable preferences. Philos. Trans. R. Soc. B 378 (1876), 20210505.
Alger, I., Weibull, J.W., 2013. Homo moralis—preference evolution under incomplete information and assortative matching. Econometrica 81 (6), 2269–2302.
Alger, I., Weibull, J.W., 2019. Evolutionary models of preference formation. Annu. Rev. Econ. 11 (1), 329–354.
Alger, I., Weibull, J.W., Lehmann, L., 2020. Evolution of preferences in structured populations: Genes, guns, and culture. J. Econom. Theory 185, 104951.
Alós-Ferrer, C., Ritschel, A., 2021. Multiple behavioral rules in Cournot oligopolies. J. Econ. Behav. Organ. 183, 250–267.
Boucher, V., Rendall, M., Ushchev, P., Zenou, Y., 2024. Toward a general theory of peer effects. Econometrica 92 (2), 543–565.
Dekel, E., Ely, J.C., Yilankaya, O., 2007. Evolution of preferences. Rev. Econ. Stud. 74 (3), 685–704.
Dong, Y., Li, C., Tao, Y., Zhang, B., 2015. Evolution of conformity in social dilemmas. PLoS One 10 (9), e0137435.
Dridi, S., Lehmann, L., 2015. A model for the evolution of reinforcement learning in fluctuating games. Anim. Behav. 104, 87–114.
Duersch, P., Oechssler, J., Schipper, B.C., 2012. Unbeatable imitation. Games Econom. Behav. 76 (1), 88–96.
Efferson, C., Vogt, S., Elhadi, A., Ahmed, H.E.F., Fehr, E., 2015. Female genital cutting is not a social coordination norm. Sci. 349 (6255), 1446–1447.
Ely, J.C., Yilankaya, O., 2001. Nash equilibrium and the evolution of preferences. J. Econom. Theory 97 (2), 255–272.

Farrow, K., Grolleau, G., Ibanez, L., 2017. Social norms and pro-environmental behavior: A review of the evidence. Ecol. Econom. 140, 1–13.
Güth, W., 1995. An evolutionary approach to explaining cooperative behavior by reciprocal incentives. Int. J. Game Theory 24 (4), 323–344.
Güth, W., Yaari, M., 1992. Explaining reciprocal behavior in simple strategic games: An evolutionary approach. Explain. Process. Chang.: Appr. Evol. Econ. 23–34.
Harsanyi, J.C., 1973. Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. Int. J. Game Theory 2 (1), 1–23.
Heifetz, A., Shannon, C., Spiegel, Y., 2007. What to maximize if you must. J. Econom. Theory 133 (1), 31–57.
Heller, Y., 2015. Three steps ahead. Theor. Econ. 10 (1), 203–241.
Herold, F., Kuzmics, C., 2020. The evolution of taking roles. J. Econ. Behav. Organ. 174, 38–63.
Laland, K.N., 2004. Social learning strategies. Anim. Learn. Behav. 32 (1), 4–14.
LiCalzi, M., Mühlenbernd, R., 2019. Categorization and cooperation across games. Games 10 (1), 5.
Mascagni, G., 2018. From the lab to the field: A review of tax experiments. J. Econ. Surv. 32 (2), 273–301.
Mohlin, E., 2012. Evolution of theories of mind. Games Econom. Behav. 75 (1), 299–318.
Ok, E.A., Vega-Redondo, F., 2001. On the evolution of individualistic preferences: An incomplete information scenario. J. Econom. Theory 97 (2), 231–254.
Rasooly, I., Rozzi, R., 2024. Masks, cameras and social pressure. J. Econ. Behav. Organ. 226, 106699.
Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M.W., Fogarty, L., Ghirlanda, S., Lillicrap, T., Laland, K.N., 2010. Why copy others? Insights from the social learning strategies tournament. Sci. 328 (5975), 208–213.
Sandholm, W.H., 2007. Evolution in Bayesian games II: Stability of purified equilibria. J. Econom. Theory 136 (1), 641–667.
Stahl, D.O., 1993. Evolution of smartn players. Games Econom. Behav. 5 (4), 604–617.