# Emergence and suppression of cooperation by action visibility in transparent games

Anton M. Unakafov*[1,2,3,4,5,6], Thomas Schultze[1,3], Igor Kagan[3,6], Sebastian Moeller[1,3,6], Stephan Eule†[2,3,4,6], and Fred Wolf†[2,3,4,5,7]

[1]Georg-Elias-Müller-Institute of Psychology, University of Goettingen, Gosslerstrasse 14, 37073, Goettingen, Germany
[2]Max Planck Institute for Dynamics and Self-Organization, Am Fassberg 17, 37077, Goettingen, Germany
[3]Leibniz ScienceCampus Primate Cognition, Kellnerweg 4, 37077, Goettingen, Germany
[4]Campus Institute for Dynamics of Biological Networks, Hermann Rein Strasse 3, 37075 Goettingen, Germany
[5]Max Planck Institute for Experimental Medicine, Hermann Rein Strasse 3, 37075 Goettingen, Germany
[6]German Primate Center - Leibniz Institute for Primate Research, Kellnerweg 4, 37077, Goettingen, Germany
[7]Bernstein Center for Computational Neuroscience, Am Fassberg 17, 37077, Goettingen, Germany

January 29, 2019

**Abstract**

Real-world agents, such as humans, animals and robots, observe each other during interactions and choose their own actions taking the partners' ongoing behaviour into account. Yet, classical game theory assumes that players act either strictly sequentially or strictly simultaneously (without knowing the choices of each other). To account for action visibility and provide a more realistic model of interactions under time constraints, we introduce a new game-theoretic setting called transparent game, where each player has a certain probability to observe the choice of the partner before deciding on its own action. Using evolutionary simulations, we demonstrate that even a small probability of seeing the partner's choice before one's own decision substantially changes evolutionary successful strategies. Action visibility enhances cooperation in a Bach-or-Stravinsky game, but disrupts cooperation in a more competitive iterated Prisoner's Dilemma. In both games, strategies based on the "Win–stay, lose–shift" and "Tit-for-tat" principles are predominant for moderate transparency, while for high transparency strategies of "Leader-Follower" type emerge. Our results have implications for studies of human and animal social behaviour, especially for the analysis of dyadic and group interactions.

One of the most interesting questions in evolutionary biology, social sciences, and economics is the emergence and maintenance of cooperation. A popular framework for studying cooperation (or the lack thereof) is Game Theory, which is frequently used to model interactions between "rational" decision-makers. In particular, a model for repeated interactions is provided by iterated games; two settings were previously used [1]:

- Simultaneous games: players act at the same time without having any information about the current choice of the partners. Consequently, all players must make a decision under uncertainty concerning the choices of others.
- Sequential games: players act in a certain order (either random or predefined [2]) and the player acting later in the sequence is guaranteed to see the choices of the preceding players. Here the burden of uncertainty only applies to the first player or – if there are more than two players – becomes lighter with every turn in the sequence.

Both settings place a simplifying restriction on the decisional context: either all players have no information about the choices of the partners (simultaneous game), or some players always have

*Corresponding author. E-mail: anton@nld.ds.mpg.de (AU)
†These authors contributed equally

more information than others (sequential game). This simplification might be disadvantageous for modelling certain behaviours, since humans and animals usually act neither strictly simultaneously nor sequentially, but observe the choices of each other and adjust their actions accordingly [3]. Indeed, the visibility of the partner's actions plays a crucial role in social interactions, both in laboratory experiments [4–7] and in natural environments [8–12].

For example, in soccer the penalty kicker must decide where to place the ball and the goalkeeper must decide whether to jump to one of the sides or to stay in the centre. Both players resort to statistics about the other's choices in the past, making this more than a simple one-shot game. Since the goalkeeper must make the choice while the opponent is preparing the shot, a simultaneous game provides a crude model for such interactions [13, 14]. However, in practice, both players observe each other's behaviour and try to anticipate the direction of the kick or of the goalkeeper's jump from subtle preparatory cues [7]. Thanks to these observations, professional goalkeepers manage to predict the direction of the shot better than chance [13–15]. While this example represents a zero-sum game, similar considerations apply to a wide range of different interactions in real life. Yet a framework for the treatment of such cases is missing in classical game theory.

To better predict and explain the outcomes of interactions between agents by taking the visibility factor into account, we introduce the concept of transparent games, where players can monitor actions of each other. The access to the information about choices of other players is therefore probabilistic; in particular, for a game between two players at each round three cases are possible:

1. Player 1 knows the choice of Player 2 before making its own choice.
2. Player 2 knows the choice of Player 1 before making its own choice.
3. Neither of players knows the choice of the partner.

Which of these cases applies depends on the reaction times of the players. If they act nearly at the same time, neither is able to use the information about partner's action; but a player who waits before making the choice has a higher probability to see the choice of the partner. Setting a time constraint (which is always present, either explicitly or implicitly, both in natural and in experimental situations) prevents players from waiting indefinitely for the partner's choice. Then, given the reaction time distributions for the players, one can infer the probability $p^i_{\text{see}}$ of Player $i$ to see the choice of the partner before making own choice.

Transparent games provide a general framework that also includes classical game-theoretical settings: simultaneous games correspond to $p^1_{\text{see}} = p^2_{\text{see}} = 0.0$, while sequential games result in $p^1_{\text{see}} = 0$, $p^2_{\text{see}} = 1$ for a fixed order of decisions in each round (Player 1 always moves first, Player 2 – second) and in $p^1_{\text{see}} = p^2_{\text{see}} = 0.5$ for a random sequence of decisions. Note that the latter case (sequential game with random order of decisions) differs qualitatively from the general case of transparent games. The critical difference is that in transparent games there is an inherent uncertainty for the first movers because they cannot know in advance whether the other players will see their decision or not in a specific round. The uncertainty of the first movers about whether the second mover sees their choice or not, may favour behaviours qualitatively different from those that yield the best performance in games with either full unidirectional transparency (sequential games) or with no transparency (simultaneous games).

The main question then is whether the probabilistic access to the information in transparent games leads to the success of different behavioural strategies as compared to classic games. To answer this question, we study transparent versions of two classic two-player games: the iterated Prisoner's dilemma (iPD) [16] and the iterated Bach-or-Stravinsky game (iBoS, also known as Battle of the Sexes and as Hero) [17]. We selected iPD and iBoS because they are counted among the most interesting games where cooperation is possible (non-zero-sum games) [17, 18], and because they require two distinct types of cooperative behaviour [19, 20]. While iPD is traditionally used for studying cooperation [16], iBoS is sometimes considered as a more suitable model [21, 22]. We employ evolutionary simulations, which allow evaluating optimal strategies using principles of natural selection,

and consider memory-one strategies [23, 24] that take into account own and partner's choices at the previous round of the game.

We find that even a small probability of seeing the choice of the partner before one's own decision changes the optimal behaviour in the iPD and iBoS games. Moreover, if the probability to see the partner's choice is high, it enhances cooperation in the generally cooperative iBoS, but disrupts cooperation in the more competitive iPD. Different transparency levels also bring qualitatively different strategies to success. In particular, we show that strategies based on the "Win–stay, lose–shift" and "Tit-for-tat" principles are the most successful in both games for low and moderate transparency, while for high transparency a new class of strategies, which we term "Leader-Follower" strategies, evolves. Although frequently observed in humans and animals (see, for instance, [25]), these strategies have up to now remained beyond the scope of game-theoretical studies, but naturally emerge in our transparent games framework.

## Results

### Transparent games without memory: analytical results

To substantiate further presentation of results we first summarize here analytical results for one-shot transparent versions of Prisoner's Dilemma (PD) and Bach-or-Stravinsky (BoS) game, with payoff matrices shown in Fig. 1. For details and proofs see "Methods" section.

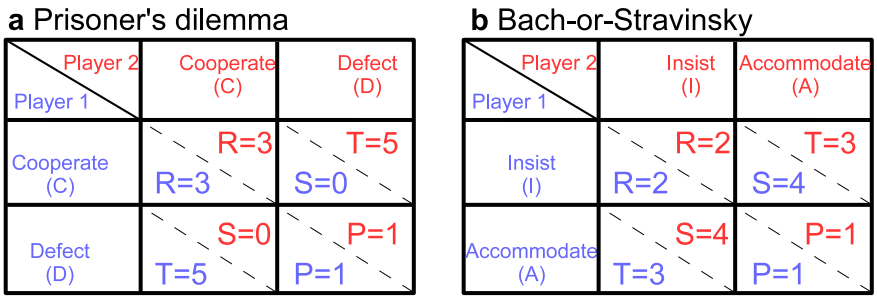



Figure 1: **Payoff matrices for Prisoner's Dilemma and Bach-or-Stravinsky game.** **a** In Prisoner's Dilemma, players adopt roles of prisoners suspected of committing a crime, arrested and kept in isolated rooms. Due to lack of evidence, prosecutors offer each prisoner an option to minimize the punishment by making a confession. A prisoner can either betray the other by defecting (D), or cooperate (C) with the partner by remaining silent. The maximal charge is five years in prison, and the payoff matrix represents the number of years deducted from it (for instance, if both players cooperate (CC, upper left), each gets a two-year sentence, because 3 years of prison time have been deducted). The letters $R$,$T$,$S$ and $P$ denoting payoff values stand for Reward, Temptation, Saint and Punishment, respectively. **b** In Bach-or-Stravinsky game two people are choosing between Bach and Stravinsky music concerts. Player 1 prefers Bach, Player 2 – Stravinsky; yet, both prefer going to the concert together. To make the game symmetric we convert musical tastes to the behavioural descriptions: insisting (I) on own preference or accommodating (A) the preference of the partner. Here cooperation is achieved when players choose different actions: either (I, A) or (A, I).

In each game players choose between two actions, $A_1$ or $A_2$ (corresponding to cooperation and defection in PD or to insisting and accommodating in BoS, respectively) according to their strategies. In a one-shot transparent game strategy is represented by a vector $(s_1; s_2; s_3)$, where $s_1$, $s_2$, $s_3$ are probabilities to select $A_1$ without seeing partner's choice, seeing partner selecting $A_1$, and seeing partner selecting $A_2$, respectively. For example, strategy $(1; 1; 0)$ in transparent PD means that player cooperates unless seeing that partner defects. The optimal behaviour for players is described by a Nash Equilibrium (NE), a pair of strategies, such that neither player can get a higher payoff by unilaterally changing its strategy. In one-shot transparent PD all Nash equilibria are comprised by

defecting strategies $(0; x; 0)$ with $0 \leq x \leq \frac{1-p_{\text{see}}}{p_{\text{see}}} \frac{P-S}{R-S}$ (Proposition 2). This means that cooperation does not survive, similar to the classic PD. However, in a finite population playing one-shot PD [26], cooperators have better chances in the transparent PD with high $p_{\text{see}}$ than in the classic settings (Proposition 3).

Nash equilibria for the one-shot transparent BoS depend on $p_{\text{see}}$ (Proposition 4). For $p_{\text{see}} < \frac{T-R}{T+S-2R}$ there are three NE: (a) Player 1 uses $(0; 0; 1)$, Player 2 uses $(1; 0; 1)$; (b) vice versa; (c) both players use strategy $(x; 0; 1)$ with $x = \frac{(S-P)-p_{\text{see}}(T+S-2P)}{(1-2p_{\text{see}})(T+S-P-R)}$. This generalises the classic case of one-shot simultaneous BoS. However for $p_{\text{see}} \geq \frac{T-R}{T+S-2R}$ the only NE is provided by $(1; 0; 1)$. In particular for the payoff matrix in Fig. 1b, there are three NE for $p_{\text{see}} < 1/3$ and one NE otherwise. As we will show below, the values of NE in one-shot BoS mainly determine the dynamics in the iterated game as well.

To summarize this part, introducing action transparency influences optimal behaviour already in simple one-shot games. As a next step we considered iterated transparent games, where players take into account results of the last interaction.

## Transparent games with memory: evolutionary simulations

Since the dynamics of strategies with memory is too complicated to be solved analytically, we used evolutionary simulations [24] to investigate strategies evolving in the transparent versions of iterated PD and BoS (iPD and iBoS). In both games we assume players to have equal mean reaction times (see "Methods" section). Then the probability $p_{\text{see}}$ to see the choice of the partner is equal for all players, which in a dyadic game results in $p_{\text{see}} \leq 0.5$.

We studied an infinite population of players using the methods described in [23,24]. For any generation $t$ the population consists of $n(t)$ "species" of players, each defined by a strategy vector $\mathbf{s}_i$ and frequency $x_i(t)$ in the population with $\sum_{i=1}^{n(t)} x_i(t) = 1$. For each species $i$ the strategy is represented by a vector $\mathbf{s}_i = (s_k^i)_{k=1}^{12}$, where $k$ enumerates the 12 different situations in which the player can be when making the choice. These depend on the outcome of the previous round, on whether or not the player can see the current choice of the partner, and on what this choice is if it is visible. The $s_k^i$ thus represent the conditional probabilities to select action $\mathrm{A}_1$, specifically

1. $s_1^i, \ldots, s_4^i$ are probabilities to select $\mathrm{A}_1$ without seeing partner's choice, given that in the previous round the joint choice of the player and the partner was $\mathrm{A}_1\mathrm{A}_1$, $\mathrm{A}_1\mathrm{A}_2$, $\mathrm{A}_2\mathrm{A}_1$, and $\mathrm{A}_2\mathrm{A}_2$ respectively (the first action specifies the choice of the player, and the second – the choice of the partner);
2. $s_5^i, \ldots, s_8^i$ are probabilities to select $\mathrm{A}_1$, seeing partner selecting $\mathrm{A}_1$ and given the outcome of the previous round (as before).
3. $s_9^i, \ldots, s_{12}^i$ are probabilities to select $\mathrm{A}_1$, seeing partner selecting $\mathrm{A}_2$ and given the outcome of the previous round.

Probabilities to select $A_2$ are represented by $(1 - s_k^i)$, respectively. To ensure numerical stability of the simulations, it is common to introduce a minimal possible error $\varepsilon$ in the strategies such that $\varepsilon \leq s_k^i \leq 1 - \varepsilon$, with $\varepsilon = 0.001$, see [23, 24]. The fact that players cannot have pure strategies and are prone to errors is also closely related to the "trembling hand" effect preventing players from using pure strategies [23, 27].

For every value of $p_{\text{see}} = 0.0, 0.1, \ldots, 0.5$ we performed 80 runs of evolutionary simulations tracing $10^9$ generations in each run. We began each run of simulations with five species having equal initial frequencies $x_1(1) = \ldots = x_5(1) = 0.2$ and random strategies $\mathbf{s}_i$. The frequency of the strategies $x_i(t)$ evolved in time according to the replicator dynamics equation (see "Methods" section). If $x_i(t)$ dropped below 0.001, the species was assumed to die out, and its share in the population was

distributed proportionally among the remaining strategies. On average every 100 generations new species with random strategies were entered in the population. Details of our simulations can be found in the "Methods" section.

Since the strategies in the evolutionary simulations were generated randomly, convergence to the theoretical optimum may take many generations and the observed successful strategies may deviate from the optimum. Therefore, we provide a coarse-grained description of strategies using the following notation: symbol 0 for $s^i_k \leq 0.1$, symbol 1 for $s^i_k \geq 0.9$, symbol * is used as a wildcard character to denote an arbitrary probability.

Let us exemplify this notation for the well-known strategies in the iPD.

1. The Generous tit-for-tat (**GTFT**) strategy is encoded by ($1a1c$;1***;****), where $0.1 < a, c < 0.9$. Indeed, GTFT cooperates with cooperators and forgives defectors. To satisfy the first property, the probability to cooperate after the partner cooperated in the previous round should be rather high, say above 0.9, thus the corresponding entries of the strategy $s^i_1$, $s^i_3$, $s^i_5$ are encoded by 1. To satisfy the second property, the probability to cooperate after the partner defected should be somewhere between zero and one with the optimal value 1/3 [23]. Since evolving towards this optimum may take long, we allow a broad range of values for $s^i_2$ and $s^i_4$, for instance $[0.1, 0.9]$. We leave $s^i_6, \ldots, s^i_{12}$ arbitrary since for low values of $p_{\text{see}}$ these entries have little influence on the strategy performance, meaning that their evolution towards optimal values may take especially long.

2. Similarly, Firm-but-fair (**FbF**) by ($101c$;1***;****), where $0.1 < c < 0.9$.

3. Tit-for-tat (**TFT**) is a "non-forgiving" version of GTFT, encoded by (1010;1***;****).

4. Win – stay, lose – shift (**WSLS**) is encoded by ($100c$;1***;****) with $c \geq 2/3$. Indeed, in the canonical simultaneous iPD (which corresponds to the case when current partner's choice is unknown) WSLS repeats its own previous action if it resulted in relatively high rewards of $R = 3$ (that is previous action was cooperation and $s^i_1 \geq 0.9$) or $T = 5$ (previous action was defection, $s^i_3 \leq 0.1$), and switches to another action otherwise ($s^i_2 \leq 0.1$, $s^i_4 = c \geq 2/3$). Note that the condition for $s^i_4$ is relaxed compared to $s^i_2$ since payoff $P = 1$ corresponding to mutual defection is not so bad compared to $S = 0$ and may not require immediate switching. Additionally, we set $s^i_5 \geq 0.9$ to ensure that WSLS cooperates with itself in the transparent iPD as it does in the simultaneous iPD.

   We also consider a relaxed version of WSLS, which we term "generous WSLS" (**GWSLS**). It follows WSLS principle only in a general sense and is encoded by ($1abc$;1***;****) with $c \geq 2/3$, $a, b < 2/3$ and either $a > 0.1$ or $b > 0.1$.

5. The Always Defect strategy (**AllD**) is encoded by (0000;**00;**00), meaning that the probability to cooperate when not seeing partner's choice or after defecting is below 0.1, and other behaviour is not specified.

We aimed to avoid over-generalisation in strategy descriptions. Therefore only the strategies that actually persisted in the population for some of our simulations were included in the coarse-grained description.

## Transparency suppresses cooperation in Prisoner's Dilemma

Simulation results for the transparent iPD are presented in Table 1. Most of the effective strategies were known from earlier studies on non-transparent games, but for high transparency ($p_{\text{see}} \to 0.5$) a new previously unknown strategy emerged. We dub this strategy "Leader-Follower" (**L-F**) since when two L-F players meet for $p_{\text{see}} = 0.5$, the player acting first (the Leader) defects, while the second player (the Follower) sees this and makes a "self-sacrificing" decision to cooperate. Note that for the next round the roles of the individuals may switch, ensuring balanced benefits of exploiting

sacrificial second move. We characterized as L-F all strategies with profile (*00$c$;****;*11$d$) with $c < 1/3$ and $d < 2/3$. Indeed, for $p_{\text{see}} = 0.5$ these entries are most important to describe the L-F strategy: after unilateral defection the Leader always defects ($s^i_2, s^i_3 \leq 0.1$) and the Follower always cooperates ($s^i_{10}, s^i_{11} \geq 0.9$). Meanwhile mutual defection most likely takes place when playing against a defector, thus both Leaders and Followers have low probability to cooperate after mutual defection ($s^i_4 = c < 1/3$, $s^i_{12} = d < 2/3$). Behaviour after mutual cooperation is only relevant when L-F is playing against another strategy, and success for different types of behaviour depends on the composition of the population. For instance, (100$c$;111*;100$d$) is optimal in a cooperative population. Note that L-F did not emerge for the sequential iPD in [2, 28, 29], since in these studies, players were bound to the same strategy regardless of whether they made their choice before or after the partner. In contrast, transparent games allow different sub-strategies ($s_1, \ldots, s_4$), ($s_5, \ldots, s_8$) and ($s_9, \ldots, s_{12}$) for these situations.

| $p_{\text{see}}$ | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|---|
| WSLS, (100$c$;1***;****) | **62.9** | **79.8** | **80.3** | **56.3** | 12.6 | 3.8 |
| GWSLS, (1$abc$;1***;****) | 0.0 | 1.0 | 6.6 | 22.4 | **16.3** | 6.0 |
| GTFT, (1$a$1$c$;1***;****) | 36.5 | 0.1 | 0.1 | 0.3 | 0.7 | 0.5 |
| TFT, (1010;1***;****) | 0.0 | 0.0 | 0.0 | 1.9 | 1.6 | 1.5 |
| FbF, (101$c$;1***;****) | 0.0 | 0.0 | 0.1 | 1.0 | 6.0 | 1.8 |
| AllD, (0000;**00;**00) | 0.1 | 0.0 | 0.0 | 0.0 | 2.4 | 1.9 |
| L-F, (*00$c$;****;*11$d$) | 0.0 | 0.2 | 0.0 | 0.0 | 0.6 | **17.8** |
| Rare and transient strategies | 0.5 | 18.9 | 12.9 | 18.1 | 59.8 | 66.7 |

Table 1: **Frequencies of stable strategies in the iterated Prisoner's Dilemma for different transparency levels.** The frequencies were computed over $10^9$ generations in 80 runs; strategies were counted as stable if they survived for more than 1000 generations after they emerged in the population. The frequency of the most successful strategy for each $p_{\text{see}}$ value is shown in **bold**.

As in the simultaneous iPD, WSLS was predominant in the transparent iPD for low and moderate $p_{\text{see}}$, which is reflected by the distinctive WSLS profiles in the final strategies of the population (Fig. 2). Note that GTFT, another successful strategy in the simultaneous iPD, disappeared completely for $p_{\text{see}} > 0$. For $p_{\text{see}} \geq 0.4$, the game resembled the sequential iPD and the results changed accordingly. Similar to the sequential iPD [2, 28, 29], the frequency of WSLS waned, the FbF strategy emerged, cooperation became less frequent and took longer to establish itself (Fig. 3a). For $p_{\text{see}} = 0.5$ the population was taken over either by L-F, WSLS-based strategies or (rarely) by FbF, which is reflected by the mixed profile in Fig. 2. Note that the share of distinctly described strategies decreased with increasing $p_{\text{see}}$, which indicates that iPD becomes unstable for high transparency, see Supplementary Fig. 1. This instability means that most strategies appear in the population only transiently and rapidly replace each other; in these circumstances performance of L-F (17.8% of the population across all generations) is quite good.

To provide a theoretical justification for our results, we analytically compared strategies most frequently emerging in simulations. Pairwise comparison of strategies (Fig. 4) helps to explain the superiority of WSLS for $p_{\text{see}} < 0.5$, the disappearance of GTFT for $p_{\text{see}} > 0.0$, and the abrupt increase of L-F frequency for $p_{\text{see}} = 0.5$.

For $p_{\text{see}} \leq 0.3$ cooperation evolved relatively quickly thanks to the predominance of WSLS. Fig. 3a shows that further increase of $p_{\text{see}}$ apparently undermined cooperation in iPD, this is why in the realistic iPD-prototype a face-to-face interrogation would be used. However, Leader-Follower is in a sense a cooperative strategy for iPD: it just alternates between cooperation and defection instead of using a synchronized cooperation. This brings to L-F an average payoff of $(S + T)/2$ when playing against itself. Alternation is generally sub-optimal in iPD since $R > (S + T)/2$; for instance, the results above are for $R = 3 > (S + T)/2 = 2.5$. To check the influence of the payoff on the strategies predominance, we have varied values of $R$ by keeping $T$, $S$ and $P$ the same as in Fig. 1 as it was done in [23] for simultaneous iPD. Fig. 5 shows that for $R > 3.2$, evolution in the transparent iPD favours

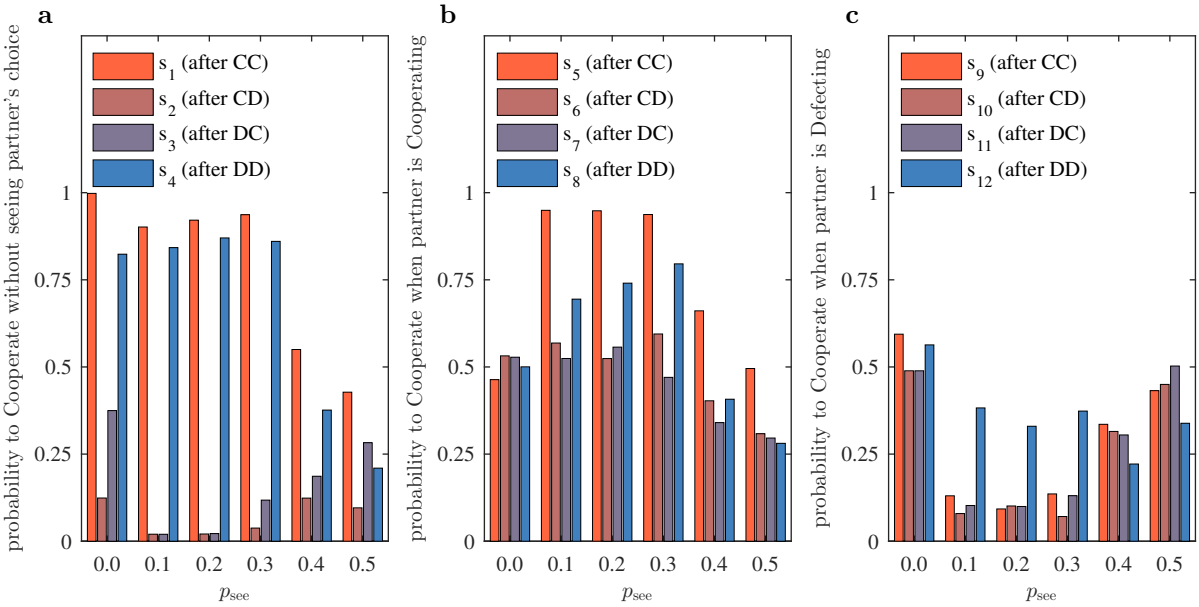



Figure 2: **iPD strategies present in the final population.** Strategies are taken for the $10^9$-th generation and averaged over all runs. CC, CD, DC and DD stand for the outcomes "both players Cooperated", "self Cooperated, partner Defected", "self Defected, partner Cooperated" and "both Defected", respectively. **a** Strategy entries $s_1, \ldots, s_4$ are close to (1001) for $p_{\text{see}} = 0.1, \ldots, 0.3$ demonstrating the dominance of WSLS. Deviations from this pattern for $p_{\text{see}} = 0.0$ and $p_{\text{see}} = 0.4$ indicate the presence of the GTFT ($1a1b$) and FbF ($101b$) strategies, respectively. For $p_{\text{see}} \geq 0.5$ strategy entries $s_1, \ldots, s_4$ are quite low due to the extinction of cooperative strategies. **b** Entries $s_5, \ldots, s_8$ are irrelevant for $p_{\text{see}} = 0.0$ (resulting in random values around 0.5) and indicate the same WSLS-like pattern for $p_{\text{see}} = 0.1, \ldots, 0.3$. Note that $s_6, s_7 > 0$ meaning that in transparent settings WSLS-players tend to cooperate seeing that the partner is cooperating even when this is against the WSLS principle. The decrease of reciprocal cooperation for $p_{\text{see}} \geq 0.4$ indicates the decline of WSLS and cooperative strategies in general. **c** Entries $s_9, \ldots, s_{12}$ are irrelevant for $p_{\text{see}} = 0.0$ (resulting in random values around 0.5) and are quite low for $p_{\text{see}} = 0.1, \ldots, 0.3$ ($s_{12}$ is irrelevant in a cooperative population). Increase of $s_9, \ldots, s_{11}$ for $p_{\text{see}} > 0.3$ indicates that mutual cooperation in the population is replaced by unilateral defection.

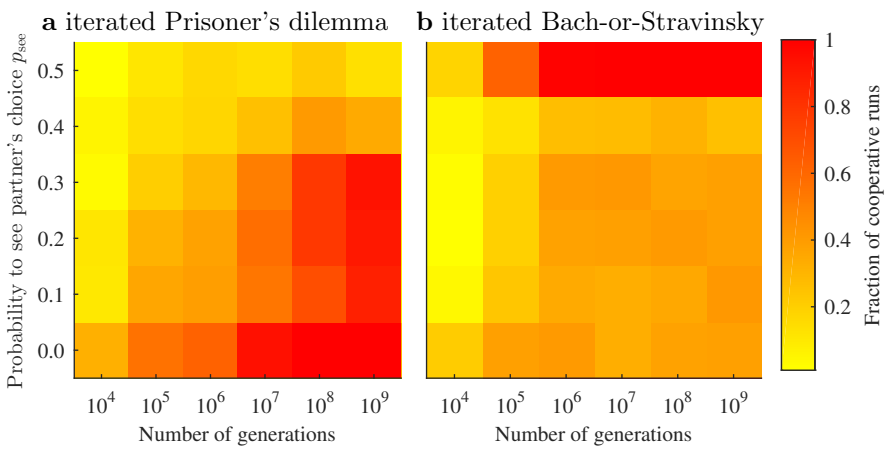



Figure 3: **Fraction of runs for which cooperation was established in iPD and iBoS.** We assumed that cooperation was established in the population if the average payoff was above $0.9 \cdot 3$ for iPD and above $0.95 \cdot 3.5$ for iBoS (90% and 95% of maximal possible value). The threshold for iPD was proposed in [23], while for iBoS we set a higher threshold due to the less competitive nature of this game.

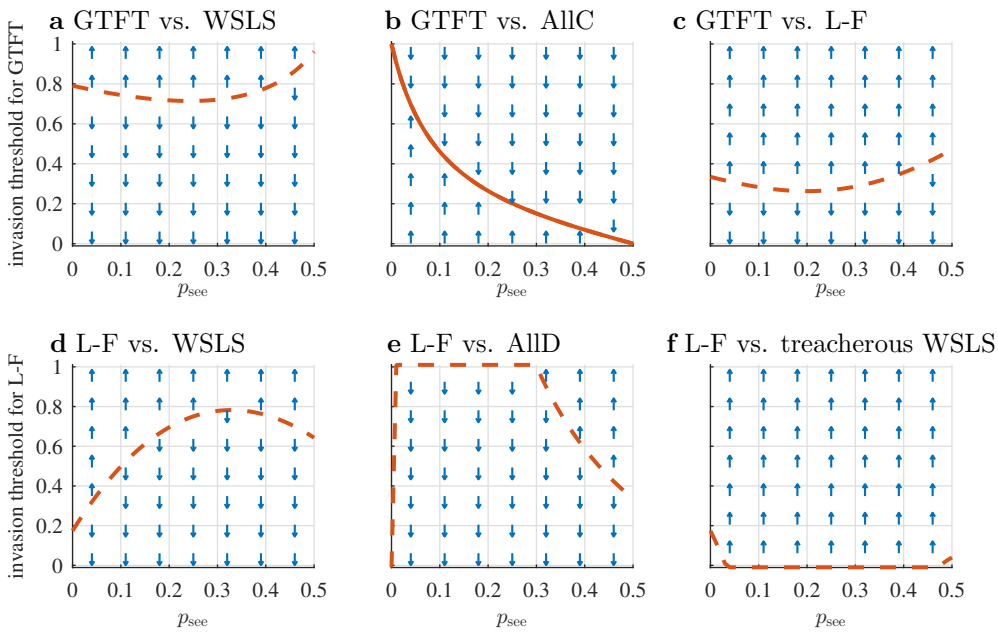



Figure 4: **Invasion thresholds computed for iPD strategies playing against each other.** The thresholds indicated by dashed lines show minimal frequency of the given strategy necessary for taking over the population against specific competitor strategies (the two strategies are bistable, see "Methods" section). The solid-line threshold shows the stable equilibrium frequency of the given strategy against specific competitor strategies (the strategies coexist, see "Methods" section). Arrows indicate whether the frequency of a strategy increases or decreases. In all strategies, 1 stands for 0.999 and 0 – for 0.001. **a** WSLS $(100\frac{9}{10}; 1001; 0000)$ has an advantage over GTFT $(1\frac{1}{3}1\frac{1}{3}; 1\frac{1}{3}1\frac{1}{3}; 0000)$: the former takes over the whole population even if its fraction is 0.25. WSLS even dominates GTFT for $p_{\text{see}} = 0.5$. **b** GTFT coexists with (prudent) AllC $(1111; 1111; 0000)$, which is more successful for $p_{\text{see}} \geq 0.1$. **c,d** L-F $(\frac{1}{3}000; \frac{2}{3}000; 111\frac{1}{3})$ performs almost as good as GTFT and WSLS, **e** but can resist the AllD strategy $(0000; 0000; 0000)$ only for high transparency. **f** Note that WSLS may lapse into its treacherous version, $(100\frac{9}{10}; 0000; 0000)$. This strategy dominates WSLS for $p_{\text{see}} > 0$ but is generally weak and cannot invade when other strategies are present in the population. Notably, when treacherous WSLS takes a part of the population, it is quickly replaced by L-F, which partially explains L-F success for high $p_{\text{see}}$.

cooperation, but $R \leq 3.2$ is sufficiently close to $(S + T)/2$ to make L-F a safe and efficient strategy.

## Cooperation emergence in the transparent Bach-or-Stravinsky game

Our simulations revealed that four memory-one strategies are most effective in iBoS for various levels of transparency. In contrast to iPD there exist only few studies of iBoS strategies, therefore we describe the observed strategies in detail.

1. **Turn-taker** aims to enter a fair coordination regime, where players alternate between IA (Player 1 insists and Player 2 accommodates) and AI (Player 1 accommodates and Player 2 insists) states. In the simultaneous iBoS, this strategy takes the form $(q01q)$, where $q = 5/8$ guarantees maximal reward in a non-coordinated play against a partner with the same strategy for the payoff matrix in Fig. 1b. We classify as Turn-takers all strategies encoded by (*01*;*0**;**1*). Turn-taking was shown to be successful in the simultaneous iBoS for a finite population of agents with pure strategies (i.e., having 0 or 1 entries only, with no account for mistakes) and a memory spanning three previous rounds [20].

2. Challenger takes the form (1101) in the simultaneous iBoS. When two players with this strategy meet, they initiate a "challenge": both insist until one of the players makes a mistake (that

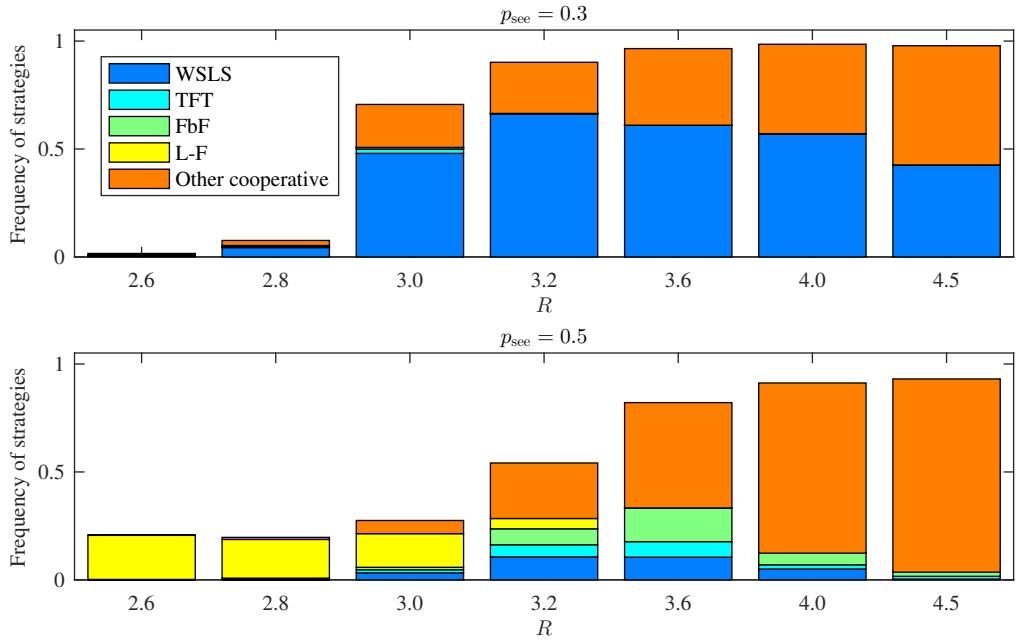



Figure 5: **Frequencies of stable strategies in the iterated Prisoner's Dilemma for various values of reward $R$ for mutual cooperation** for $p_{see} = 0.3$ and for $p_{see} = 0.5$. Values of $T$, $S$ and $P$ are the same as in Fig. 1, values of $R$ are in range $(S + T)/2 < R < T$ that defines the Prisoner's Dilemma payoff. The frequencies were computed over $10^9$ generations in 40 runs; strategies were counted as stable if they survived for more than 1000 generations after they emerged in the population. Here we present results for the TFT strategy (1010;1***;****) since GTFT did not emerge in these simulations. We describe all strategies having a pattern (1*1*;1***;****) or (1**1;1***;****) but different from WSLS, TFT and FbF as "other cooperative". While for $p_{see} = 0.3$ population for low $R$ mainly consists of defectors, for $p_{see} = 0.5$ L-F provides an alternative to defection. For $R \geq 3.2$ mutual cooperation becomes much more beneficial, which allows cooperative strategies to prevail for all transparency levels.

is, accommodates). Then, the player making the mistake (loser) submits and continues accommodating, while the winner continues insisting. This period of unfair coordination beneficial for the winner ends when the next mistake of either player (the winner accommodating or the loser insisting) triggers a new "challenge". This strategy is encoded by (11$b$*;****;*1**) and has two variants: **Challenger** "obeys the rules" and does not initiate the challenge after losing ($b \leq 0.1$), while **Aggressive Challenger** may switch to insisting ($0.1 < b \leq 1/3$). Challenging strategies were theoretically predicted to be successful in simultaneous iBoS [30, 31].

3. The Leader-Follower (**L-F**) strategy $\mathbf{s} = (1111; 0000; 1111)$ was not considered previously. In a game between two players with this strategy, the faster player insists and the slower player accommodates. In simultaneous game, this strategy lapses into inefficient stubborn insisting since all players consider themselves leaders, but in transparent settings with high $p_{see}$ this strategy provides an effective and fair cooperation (because of the equal mean reaction times). In particular, for $p_{see} > 1/3$ L-F strategy is a Nash Equilibrium in one-shot game (see Proposition 4 in Methods section), and is an evolutionary stable strategy for $p_{see} = 0.5$.

   When the entire population adopts an L-F strategy, most entries of the strategy vector become irrelevant since (i) only IA and AI states are visited and (ii) the faster player never accommodates. Therefore, we classify all strategies encoded by (*11*;*00*;****) as L-F.

4. **Challenging Leader-Follower** is a hybrid of the Challenger and L-F strategies encoded by (11$b$*;0$c$0*;*1**), where $1/3 < b \leq 0.9$, $c \leq 1/3$. This strategy tends to insist without seeing partner's choice, and tends to accommodate when seeing that partner insists; both these strategies are stronger than for Aggressive Challengers, but not as strong as for Leader-Followers.

The results of the simulations are presented in Table 2. The entries of the averaged-over-all-runs strategy established in the population (Fig. 6) show considerably different profiles for various values of $p_{\text{see}}$. Challengers, Turn-takers, and Leader-Followers succeeded for low, medium and high probabilities to see partner's choice, respectively.

| $p_{\text{see}}$ | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|---|
| Turn-taker | 37.5 | 41.2 | **37.5** | **37.5** | 24.9 | 0.0 |
| Challenger | **62.5** | **42.7** | 3.4 | 0.5 | 0.0 | 0.0 |
| Aggressive Challenger | 0.0 | 14.4 | 30.7 | 3.3 | 0.0 | 0.0 |
| Challenging Leader-Follower | 0.0 | 1.1 | 25.2 | 31.8 | 0.0 | 0.0 |
| Leader-Follower | 0.0 | 0.1 | 2.5 | 26.1 | **74.6** | **100.0** |
| Rare and transient strategies | 0.0 | 0.5 | 0.7 | 0.8 | 0.5 | 0.0 |

Table 2: **Frequencies of stable strategies in the Bach-or-Stravinsky game for different transparency levels.** The frequencies were computed over $10^9$ generations in 80 runs; strategies were counted as stable if they survived for more than 1000 generations after they emerged in the population. The frequency of the most successful strategy for each $p_{\text{see}}$ value is shown in **bold**.

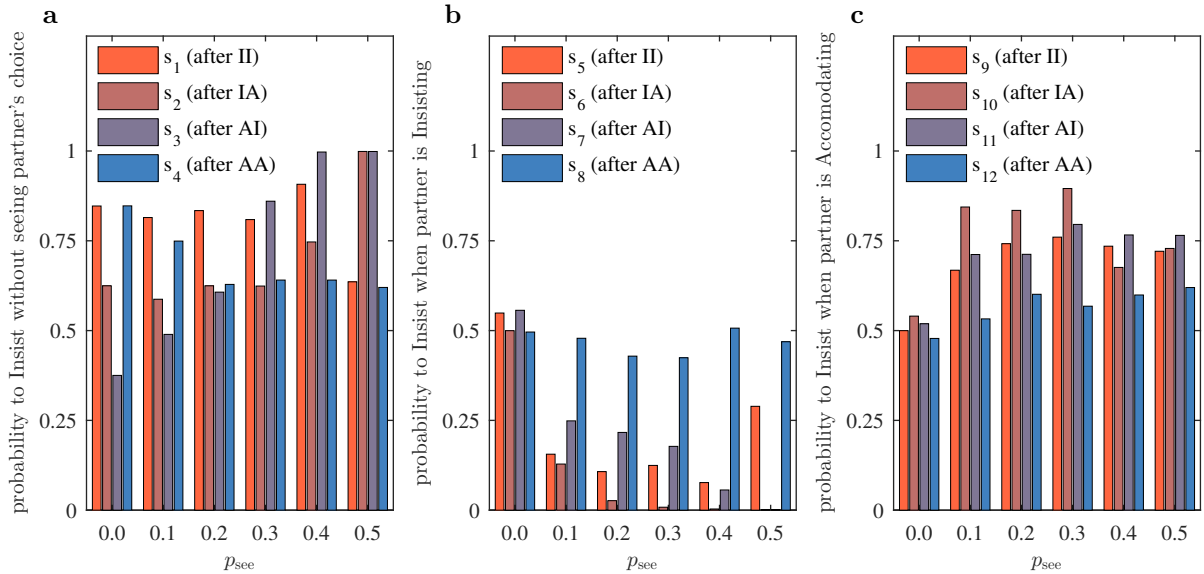



Figure 6: **iBoS strategies present in the final population.** Strategies are taken for the $10^9$-th generation and averaged over all runs. II, IA, AI and AA stand for the outcomes "both players Insisted", "self Insisted, partner Accommodated", "self Accommodated, partner Insisted" and "both Accommodated", respectively. **a** Strategy entries $s_1, \ldots, s_4$. The decrease of the $s_2/s_3$ ratio reflects the transition of the dominant strategy from challenging to turn-taking for $p_{\text{see}} = 0.0, \ldots, 0.4$. For $p_{\text{see}} = 0.5$ the dominance of the Leader-Follower strategy is indicated by $s_2 = s_3 = 1$. **b** Entries $s_5, \ldots, s_8$ are irrelevant for $p_{\text{see}} = 0$. Values of $s_6, s_7$ decrease as $p_{\text{see}}$ increases, indicating an enhancement of cooperation in iBoS with higher transparency ($s_8$ is almost irrelevant since mutual accommodation is a very rate event, and $s_5$ is irrelevant for a population of L-F players taking place for $p_{\text{see}} = 0.5$). **c** Entries $s_9, \ldots, s_{12}$ are irrelevant for $p_{\text{see}} = 0$. The decrease of the $s_{10}/s_{11}$ ratio for $p_{\text{see}} = 0.1, \ldots, 0.4$ reflects the transition of the dominant strategy from challenging to turn-taking.

To provide additional insight into the results of the iBoS simulations, we studied analytically how various strategies perform against each other (Fig. 7). As with the iPD, this analysis helps to understand why different strategies were successful at different transparency levels.

A change of behaviour for $p_{\text{see}} > 1/3$ is in line with our theoretical results indicating that for this transparency levels L-F is a Nash Equilibrium. Population dynamics for iBoS with a payoff different from the presented in Fig. 1b also depends on the Nash Equilibria of one-shot game, described by Proposition 4 in Methods section.

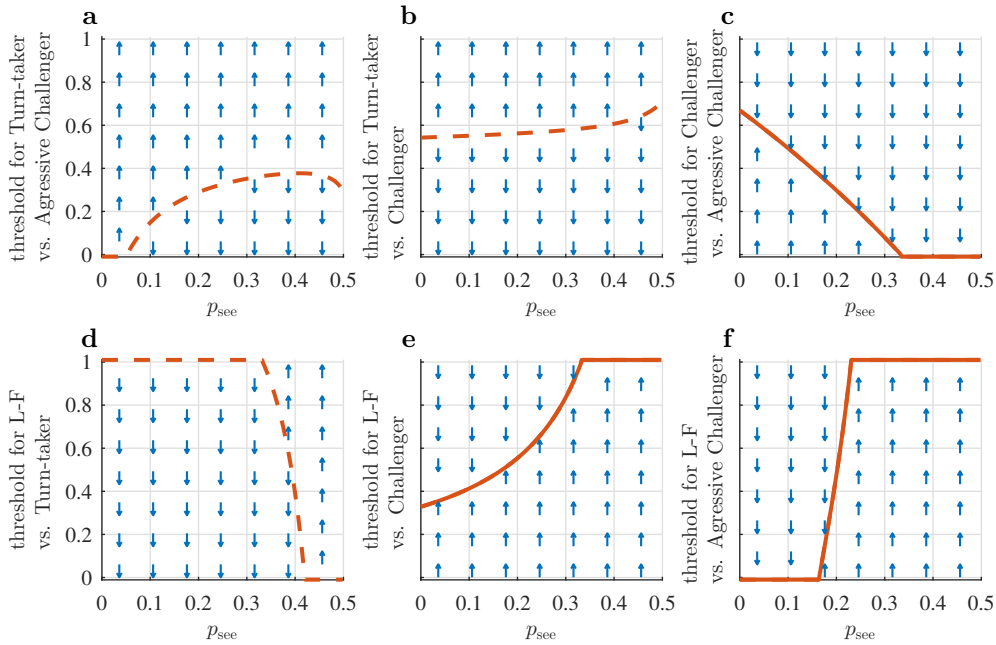



Figure 7: **Invasion thresholds computed for iBoS strategies playing against each other.** The thresholds are indicated by dashed lines when strategies are bistable, and by solid lines when strategies coexist (see "Methods" section). Arrows indicate whether the frequency $x_1(t)$ of the first strategy increases or decreases. In all strategies, 1 stands for 0.999 and 0 – for 0.001, the entries $s_9 = \ldots = s_{12} = 1$ are omitted. **a** Turn-taker $(q01q; 0000)$ with $q = 5/8$ for $p_{\text{see}} > 0$ outperforms Aggressive Challenger $(11\frac{1}{5}1; \frac{1}{2}\frac{1}{2}\frac{1}{5}\frac{1}{2})$, **b** but not Challenger $(\frac{9}{10}101; \frac{1}{2}\frac{1}{2}0\frac{1}{2})$. **c** Challenger can coexist with Aggressive Challenger for low transparency, but is dominated for $p_{\text{see}} > 1/3$. **d,e,f** Leader-Follower $(1111; 0000)$ clearly outperforms other strategies for $p_{\text{see}} > 1/3$.

In contrast to iPD, for iBoS high visibility results in a more effective cooperation, which is consistent with the notion that cooperation in the iBoS game rests on coordination (rather than trust in the good intentions of the partner). Indeed, for $p_{\text{see}} \geq 0.3$ non-cooperative Challengers no longer constituted the majority of the population. Note that for $p_{\text{see}} = 0.5$ cooperation thrives and is established much faster than for lower transparency (Fig. 3b) thanks to the emergence of Leader-Follower strategy.

## Discussion

In this paper, we introduced the concept of transparent games which integrates the visibility of the partner's actions into a game-theoretic settings. Specifically, we considered iterated dyadic games where players have probabilistic access to the information about the partner's choice in the current round. When reaction times for both players are equal on average, the probability $p_{\text{see}}$ of accessing this information can vary from $p_{\text{see}} = 0.0$ (corresponding to the canonical simultaneous games) to $p_{\text{see}} = 0.5$ (corresponding to sequential games with random order of choices. Note, however, that in the classic sequential games players have the same strategy when making their choice before and after the partner, while the transparent games framework allows players to use different sub-strategies in these cases). Our approach is similar to the continuous-time approach suggested in [32]. However, in that study a game is played continuously, without any rounds at all, while here we suppose that the game consists of clearly specified rounds, although the time within each round is continuous. This assumption seems to be natural, since many real world interactions and behaviours are episodic, have clear starting and end points, and hence are close to distinct rounds [6,8,33,34]. Transparent games to some degree resemble random games [35,36] since in both concepts the outcome of the game depends on a stochastic factor. However in random games randomness immediately affects the payoff, while in transparent games it determines the chance to learn the partner's choice. While this chance influences

the payoff of the players, the effect depends on their strategies, which is not the case in random games.

The value of $p_{\text{see}}$ strongly affects the evolutionary success of strategies. In particular, in the Bach-or-Stravinsky game (iBoS) even moderate $p_{\text{see}}$ helps to establish cooperative turn-taking, while high $p_{\text{see}}$ brings about a new successful strategy, Leader-Follower (L-F). For the transparent iterated Prisoner's Dilemma (iPD) we have shown that for $p_{\text{see}} > 0$ the Generous tit-for-tat strategy is unsuccessful and Win–stay, lose–shift (WSLS) is an unquestionable evolutionary winner for $0 < p_{\text{see}} \leq 0.4$. However, WSLS is not evolutionary stable (see the caption of Fig. 4); our results indicate that in general there are no evolutionary stable strategies in the transparent iPD, which was already known to be the case for the simultaneous iPD [24]. Moreover, if reward for mutual cooperation $R \leq 3$, for high transparencies ($p_{\text{see}} \geq 0.4$) all strategies become quite unstable and cooperation is hard to establish (Fig. 5). Finally, for $p_{\text{see}} = 0.5$, L-F becomes successful in iPD and is more frequent than WSLS for $R \leq 3.2$. For such a payoff, mutual cooperation is not much more beneficial than the alternating unilateral defection resulting from the L-F strategy. It brings a payoff of $(S + T)/2 = 2.5$, but is generally less susceptible to exploitation by defecting strategies. This explains the abrupt drop of cooperation in the transparent iPD with $p_{\text{see}} \geq 0.4$ for $R = 3.0$ (Supplementary Fig. 4), while there is no such drop for $R > 3.2$ (Fig. 5). Note that $R > 3.2$ strongly promotes mutual cooperation over other options, therefore this case is slightly less interesting than the classic payoff matrix with $R = 3$.

Although resulting in a lower payoff than the explicit cooperation, L-F in a sense is also a cooperative strategy for iPD. While the choice of Leaders (defection) is entirely selfish, Followers "self-denyingly" cooperate with them. Importantly, the L-F strategy is not beneficial for some of the players using it in any finite perspective, which distinguishes this strategy from most cooperative strategies. Let us explain this point by comparing L-F with WSLS. In a game between two WSLS-players, neither benefits from unilaterally switching to defection even in a short term for $R \geq (T + P)/2$. While the defecting player gets $T = 5$ in the first round, its payoff in the next round is $P = 1$, which makes the average payoff over two rounds less than or equal to the reward for cooperation $R$. Thus for the iPD with standard payoff $R = 3 = (T + P)/2$ WSLS players do not benefit from defecting their WSLS-partners already for the two-round horizon. (Note that for $R < (T + P)/2$ defection is effective against WSLS, which explains the low frequency of WSLS for $R < 3$ in Fig. 5). Now, assume that we are playing the transparent iPD with $p_{\text{see}} = 0.5$ against a partner with a pure L-F strategy (0000;1111;1111) and have to choose between L-F and AllD strategies. In a single round using AllD is (strictly) better with probability $p = 1/2$ (probability of being a Follower). From the two-round-perspective using AllD is beneficial with $p = 1/4$ (the probability of being a Follower in both rounds). For $n = 6$ rounds, AllD is still better than L-F with $p = 7/64$ (the probability of being a Follower in 5 or 6 rounds out of 6, which results in an average payoff equal to either 5/6 or 0). In general, for any finite number $n$ of rounds, there is a risk to suffer from using the L-F strategy instead of AllD, and the probability of this is given by $\sum_{k=0}^{\lceil nP/T \rceil} C_k^n$, where $\lceil nP/T \rceil$ is the integer part of $nP/T$ and $C_k^n = \frac{n!}{k!(n-k)!}$ is a binomial coefficient. That is adhering to L-F is not beneficial for some of the L-F players in any finite horizon, which makes their behaviour in a sense altruistic. Our results for the transparent iPD demonstrate that such "altruistic-like" behaviour may evolve in a population even without immediate reciprocation. The inherently unequal payoff distribution among L-F players for a final number of rounds opens interesting perspectives for research, but is outside the scope of this manuscript.

The lack of stability in the transparent iPD turns the analysis of the strategy dynamics for this game into a non-trivial problem. Therefore we do not provide here an exhaustive description of strategies in iPD and contend ourselves with general observations and explanations. An in-depth analysis of strategy dynamics in the transparent iPD will be provided elsewhere as a separate, more technical paper.

Despite the clear differences between the two games, predominant strategies evolving in iPD and iBoS for various levels of transparency have some striking similarities. First of all, in both games, L-F appears to be the most successful strategy for high $p_{\text{see}}$ (although for iPD with $R \leq 3$ the share

of Leader-Followers in the population across all generations is only about 20%, other strategies are even less successful as most of them appear just transiently and rapidly replace each other). This prevalence of the L-F strategy can be explained as follows: in a group where the behaviour of each agent is visible to the others and can be correctly interpreted, group actions hinge upon agents initiating these actions. In both games these initiators are selfish, but see Supplementary Note 2 for an example of an "altruistic" leadership.

For low and moderate values of $p_{\text{see}}$ the similarities of the two games are less obvious. However, the Challenger strategy in iBoS follows the same principle of "Win – stay, lose – shift" as the predominant strategy WSLS in iPD, but with modified definitions of "win" and "lose". For Challenger winning is associated with any outcome better than the minimal payoff corresponding to the mutual accommodation. Indeed, Challenger accommodates until mutual accommodation takes place and then switches to insisting. Such behaviour is described as "modest WSLS" in [31, 37] and is in-line with the interpretation of the "Win – stay, lose – shift" principle observed in animals [38].

The third successful principle in the transparent iPD is "Tit-for-tat", embodied in Generous tit-for-tat (GTFT), TFT and Firm-but-fair (FbF) strategies. This principle also works in both games since turn-taking in iBoS is nothing else but giving tit for tat. In particular, the TFT and FbF strategies, which occur frequently in iPD for $p_{\text{see}} \geq 0.4$, are partially based on taking turns and are similar to the Turn-Taker strategy in iBoS. The same holds to a lesser extent for the GTFT strategy.

The success of specific strategies for different levels of $p_{\text{see}}$ makes sense if we understand $p_{\text{see}}$ as a species' ability to signal intentions and to interpret these signals when trying to coordinate (or compete). The higher $p_{\text{see}}$, the better (more probable) is the explicit coordination. This could mean that a high ability to explicitly coordinate actions leads to coordination based on observing the leader's behaviour. In contrast, moderate coordination ability results in some form of turn-taking, while low ability leads to simple strategies of WSLS-type. In fact, an agent utilizing the WSLS principle does not even need to comprehend the existence of the second player, since WSLS "embodies an almost reflex-like response to the pay-off" [23]. The ability to cooperate may also depend on the circumstances, for example, on the physical visibility of partner's actions. In a relatively clear situation, following the leader can be the best strategy. Moderate uncertainty requires some (implicit) rules of reciprocity embodied in turn-taking. High uncertainty makes coordination difficult or even impossible, and may result in a seemingly irrational "challenging behaviour" as we have shown for the transparent iBoS. However, when players can succeed without coordination (which was the case in iPD), high uncertainty about the other players' actions does not cause a problem.

By taking the visibility of the agents' actions into account, transparent games may offer a compelling theoretical explanation for a range of biological, sociological and psychological phenomena. One potential application of transparent games is related to experimental research on social interactions, including the emerging field of social neuroscience that seeks to uncover the neural basis of social signalling and decision-making using neuroimaging and electrophysiology in humans and animals [39–42]. So far, most studies have focused on sequential [43, 44] or simultaneous games [45]. One of the main challenges in this field is extending these studies to direct real-time interactions that would entail a broad spectrum of dynamic competitive and cooperative behaviours. In line with this, several recent studies also considered direct social interactions in humans and non-human primates [4–6, 34, 46–50] during dyadic games where players can monitor actions and outcomes of each other. Transparent games allow modelling the players' access to social cues, which is essential for the analysis of experimental data in the studies of this kind [22]. This might be especially useful when behaviour is explicitly compared between "simultaneous" and "transparent" game settings, as in [4, 6, 46, 50]. In particular, the enhanced cooperation in the transparent iBoS for high $p_{\text{see}}$ provides a theoretical explanation for the empirical observations in [6], where humans playing an iBoS-type game demonstrated a higher level of cooperation and a fairer payoff distribution when they were able to observe the actions of the partner while making their own choice. In view of the argument that true cooperation should benefit from enhanced communication [22], the transparent iBoS can in certain cases be a more suitable model for studying cooperation than the iPD (see also [51, 52] for a discussion of studying cooperation

by means of iBoS-type games).

In summary, transparent games provide a theoretically attractive link between classical concepts of simultaneous and sequential games, as well as a computational tool for modelling real-world interactions. This approach allows integrating work on sensorimotor decision-making under uncertainty with economic game theory. We thus expect that the transparent games framework will help to establish a deeper understanding of social behaviour in humans and animals.

# Methods

## Transparent games between two players

In this study, we focus on iterated two-player (dyadic) two-action games: in every round both players choose one of two possible actions and get a payoff depending on the mutual choice according to the payoff matrix (Fig. 1). A new game setting, *transparent game*, is defined by a payoff matrix and probabilities $p^i_{\text{see}}$ $(i = 1, 2)$ of Player $i$ to see the choice of the other player, $0 \leq p^1_{\text{see}}, p^2_{\text{see}} \leq 1$. Note that $p^1_{\text{see}} + p^2_{\text{see}} \leq 1$, and $(1 - p^1_{\text{see}} - p^2_{\text{see}})$ is the probability that neither of players knows the choice of the partner because they act sufficiently close in time so that neither players can infer the other's action prior to making their own choice. The probabilities $p^i_{\text{see}}$ can be computed from the distributions of reaction times for the two players, as shown in Supplementary Fig. 2 for reaction times modelled by exponentially modified Gaussian distribution [53, 54]. In this figure, reaction times for both players have the same mean, which results in symmetric distribution of reaction time differences (Supplementary Fig. 2b) and $p^1_{\text{see}} = p^2_{\text{see}} \leq 0.5$. Here we focus only on this case since for both games considered in this study, unequal mean reaction times provide a strong advantage to one of the players (see below). However, in general $p^1_{\text{see}} \neq p^2_{\text{see}}$.

To illustrate how transparent, simultaneous and sequential games differ, let us consider three scenarios for a Prisoner's Dilemma (PD):

1. If prisoners write their statements and put them into envelopes, this case is described by simultaneous PD.
2. If prisoners are questioned in the same room in a random or pre-defined order, one after another, this case is described by sequential PD.
3. Finally, in a case of a face-to-face interrogation where prisoners are allowed to answer the questions of prosecutors in any order (or even to talk simultaneously) the transparent PD comes into play. Here prisoners are able to monitor each other and interpret inclinations of the partner in order to adjust their own choice accordingly.

While the transparent setting can be used both in zero-sum and non-zero-sum games, here we concentrate on the latter class where players can cooperate to increase their joint payoff. For the purposes of this work, we define cooperation simply as joint actions towards mutually beneficially outcomes. In various areas more specific definitions of cooperation are used (see, for example, [8, 22] for a discussion of cooperation in animals). We consider the transparent versions of two classic games, the PD and the Bach-or-Stravinsky game (BoS). We have selected PD and BoS as representatives of two distinct types of symmetric non-zero-sum games [19, 20]: maximal joint payoff is awarded when players select the **same** action (cooperate) in PD, but **complementary** actions in BoS (one insists, and the other accommodates). The games of PD type are known as *synchronization* games; other examples of synchronization games include Stag Hunt and Game of Chicken [20]. Games such as BoS with two optimal mutual choices are called *alternation* games [19, 20]; as one of these choices is more beneficial for Player 1, and the other for Player 2, to achieve fair cooperation players should alternate between these two states.

Another important difference between the two considered games is that in BoS it is always better to act before the partner, while in PD it is mostly preferable to act after the partner. Indeed, in BoS

the player acting first has good chances to get the maximal payoff of $S = 4$ by insisting: when the second player knows that the partner insists, it is better to accommodate and get a payoff of $T = 3$, than to insist and get $R = 2$. In PD, however, defection is less beneficial if it can be discovered by the opponent and acted upon (for details, see Subsection "One-shot transparent Prisoner's Dilemma with unequal reaction times" below). Therefore, in PD most players prefer acting later: defectors to have a better chance of getting $T = 5$ for a successful defection, and cooperators to make sure that the partners are not defecting them. The only exception from this rule is the Leader-Follower strategy, but as we show in Supplementary Note 1 this special case does not change the overall situation for the simulations. Therefore, the optimal behaviour in PD is generally to wait as long as possible, while in BoS a player should act as quickly as possible. Consequently, when the time for making choice is bounded from below and from above, evolution in these games favours species with marginal mean reaction times: maximal allowed reaction time in PD and minimal allowed reaction time in BoS. Species with different behaviour are easily invaded. Therefore we assumed in all simulations that the reaction times have a constant and equal mean. We also assumed that reaction times for all species have an equal non-zero variance and that the difference of the reaction time distributions of two species is always symmetric (see Supplementary Fig. 2). This results in $p^i_{\text{see}}$ being the same for all species, thus all players have equal chances to see the choices of each other.

## Analysis of one-shot transparent games

Consider a one-shot transparent game between Player 1 and Player 2 having strategies $\mathbf{s}_1 = (s^1_1; s^1_2; s^1_3)$ and $\mathbf{s}_2 = (s^2_1; s^2_2; s^2_3)$, and probabilities to see the choice of the partner $p^1_{\text{see}}$ and $p^2_{\text{see}}$, respectively. An expected payoff for Player 1 is given by

$$\begin{aligned} \mathrm{E}(\mathbf{s}_1, \mathbf{s}_2) = {} & (1 - p^1_{\text{see}} - p^2_{\text{see}})(s^1_1 s^2_1 R + s^1_1(1 - s^2_1)S + (1 - s^1_1)s^2_1 T + (1 - s^1_1)(1 - s^2_1)P) \\ & + p^2_{\text{see}} s^1_1 (s^2_2 R + (1 - s^2_2)S) + p^2_{\text{see}}(1 - s^1_1)(s^2_3 T + (1 - s^2_3)P) \\ & + p^1_{\text{see}} s^2_1 (s^1_2 R + (1 - s^1_2)T) + p^1_{\text{see}}(1 - s^2_1)(s^1_3 S + (1 - s^1_3)P), \end{aligned} \tag{1}$$

where the first line describes the case when neither player sees partner's choice, the second line describes the case when Player 2 sees the action of Player 1, and the third – when Player 1 sees the action of Player 2.

Let us provide two definitions that will be used throughout this section.
**Definition 1.** Strategies $\mathbf{s}_1$ and $\mathbf{s}_2$ are said to form a *Nash Equilibrium* if neither player would benefit from unilaterally switching to another strategy, that is $\mathrm{E}(\mathbf{s}_1, \mathbf{s}_2) \geq \mathrm{E}(\mathbf{s}'_1, \mathbf{s}_2)$ and $\mathrm{E}(\mathbf{s}_2, \mathbf{s}_1) \geq \mathrm{E}(\mathbf{s}'_2, \mathbf{s}_1)$ for any alternative strategies $\mathbf{s}'_1$ and $\mathbf{s}'_2$ of Players 1 and 2, respectively.
**Definition 2.** Let us denote $\mathrm{E}_{ij} = \mathrm{E}(\mathbf{s}_i, \mathbf{s}_j)$. Strategy $\mathbf{s}_1$ is said to *dominate* strategy $\mathbf{s}_2$ if using $\mathbf{s}_1$ would give better outcome for both players, that is $\mathrm{E}_{11} \geq \mathrm{E}_{21}$ and $\mathrm{E}_{12} \geq \mathrm{E}_{22}$. If both inequalities are strict, $\mathbf{s}_1$ *strongly dominates* $\mathbf{s}_2$. Strategies $\mathbf{s}_1$ and $\mathbf{s}_2$ are said to be *bistable* when $\mathrm{E}_{11} > \mathrm{E}_{21}$ and $\mathrm{E}_{12} < \mathrm{E}_{22}$. Strategies $\mathbf{s}_1$ and $\mathbf{s}_2$ *co-exist* when $\mathrm{E}_{11} < \mathrm{E}_{21}$ and $\mathrm{E}_{12} > \mathrm{E}_{22}$.

Some intuition on these notions is provided below in subsection "Evolutionary dynamics of two strategies". We refer to [24] for details.

For the sake of simplicity, we assume for the rest of this section that $0 < p^1_{\text{see}}, p^2_{\text{see}} < 1$, otherwise the game is equivalent to the classic sequential or simultaneous game. First we consider the one-shot transparent Prisoner's dilemma (PD), and then – Bach-or-Stravinsky (BoS) game.

**One-shot transparent Prisoner's Dilemma with equal reaction times.** Here we assume that $p^1_{\text{see}} = p^2_{\text{see}} = p_{\text{see}}$ to simplify the discussion. Similar to the classic one-shot PD, in the transparent PD all Nash Equilibria (NE) correspond to mutual defection. To show this we make an important observation: in the one-shot PD it is never profitable to cooperate when seeing the partner's choice.

**Lemma 1.** *In one-shot transparent PD with $p_{\text{see}} > 0$ any strategy $(s_1; s_2; s_3)$ is dominated by strategies $(s_1; s_2; 0)$ and $(s_1; 0; s_3)$. The dominance of $(s_1; s_2; 0)$ is strong when $s_1 < 1$, the dominance of $(s_1; 0; s_3)$ is strong when $s_1 > 0$.*

*Proof.* The lemma follows immediately from (1). Since in PD $R < T$, expected payoff E of strategy $(s_1; s_2; s_3)$ is maximized when $s_2 = 0$. Similarly, from $S < P$ it follows that the payoff is maximized for $s_3 = 0$. □

Now we can describe the NE strategies in transparent PD:

**Proposition 2.** *In one-shot transparent PD all the Nash equilibria are comprised by pairs of strategies $(0; x; 0)$ with $0 \leq x \leq 1$ and*

$$x \leq \frac{1 - p_{\text{see}}}{p_{\text{see}}} \frac{P - S}{R - S}. \tag{2}$$

*Proof.* First we show that for any $x, y$ satisfying (2), strategies $(0; x; 0)$ and $(0; y; 0)$ form a Nash equilibrium. Assume that there exists a strategy $(s_1; s_2; s_3)$, which provides a better payoff against $(0; x; 0)$ than $(0; y; 0)$. According to Lemma 1, expected payoff of a strategy $(s_1; 0; 0)$ is not less than the payoff of $(s_1; s_2; s_3)$. Now it remains to find the value of $s_1$ maximizing the expected payoff E of $(s_1; 0; 0)$. From (1) we have:

$$\begin{aligned} \mathrm{E} &= (1 - 2p_{\text{see}})(s_1 S + (1 - s_1)P) + p_{\text{see}} s_1 (xR + (1 - x)S) + p_{\text{see}}(1 - s_1)P + p_{\text{see}} P \\ &= P + s_1 (p_{\text{see}} x(R - S) - (1 - p_{\text{see}})(P - S)). \end{aligned}$$

Thus the expected payoff is maximized by $s_1 = 0$ if inequality (2) holds and by $s_1 = 1$ otherwise. In the former case the strategy $(s_1; 0; 0)$ results in the same payoff $P$ as the strategy $(0; y; 0)$, which proves that a pair of strategies $(0; x; 0)$, $(0; y; 0)$ is an NE. If (2) does not hold, strategy $(0; x; 0)$ is not an NE, since switching to $(1; 0; 0)$ results in a better payoff.

Let us show that there are no further NE. Indeed, according to Lemma 1 if an alternative NE exists, it can only consist of strategies $(1; 0; z)$ or $(u; 0; 0)$ with $0 \leq z \leq 1$ and $0 < u < 1$. In both cases switching to unconditional defection is preferable, which finishes the proof. □

The one-shot transparent PD has two important differences from the classic game. First, defective strategies dominate the cooperative strategy $(1; 1; 0)$ only for $p_{\text{see}} < \frac{T-R}{T-P}$. Indeed, when both players stick to $(1; 1, 0)$, their payoff is equal to $R$, while when switching to $(0; 0; 0)$ strategy, a player gets $p_{\text{see}} P + (1 - p_{\text{see}})T$. However, $(1; 1, 0)$ is dominated by a strategy $(1; 0; 0)$ that cooperates when it does not see the choice of the partner and defects otherwise. This strategy, in turn is dominated by $(0; 0; 0)$.

Second, in transparent PD unconditional defection $(0; 0; 0)$ is not evolutionary stable as players can switch to $(0; x; 0)$ with $x > 0$ retaining the same payoff. This, together with Proposition 3 below, makes possible a kind of evolutionary cycle: $(1; 0; 0) \to (0; 0; 0) \leftrightarrow (0; x; 0) \to (1; 1; 0), (1; 0; 0) \to (1; 0; 0)$. In summary, although transparency does not allow cooperation to persist when evolution is governed by deterministic dynamics, it would increase chances of cooperators for the stochastic dynamics in a finite population.

**Proposition 3.** *In transparent PD strategies $(1; 0; 0)$ and $(0; x; 0)$ have the following relations:*

1. *if condition (2) and the following condition*

$$x \leq \frac{1}{p_{\text{see}}} - 2 + \frac{P - S}{T - R} \tag{3}$$

   *are satisfied, then $(0; x; 0)$ dominates $(1; 0; 0)$;*

2. *if neither (2) nor (3) are satisfied, then $(1; 0; 0)$ dominates $(0; x; 0)$;*

3. *if (2) is satisfied but (3) is not, then the two strategies coexist;*

*4. if* (3) *is satisfied but* (2) *is not, then the two strategies are bistable.*

*Proof.* We prove only the first statement since the proof of the others is almost the same.

Let Player 1 use strategy $(1;0;0)$ and Player 2 – strategy $(0;x;0)$. To prove that $(0;x;0)$ dominates $(1;0;0)$ we need to show that Player 2 has no incentive to switch to $(1;0;0)$ and that Player 1, on the contrary, would get higher payoff if using $(0;x;0)$. The latter statement follows from Proposition 3. To show that the former also takes place we simply write down expected payoffs $\mathrm{E}_1$ and $\mathrm{E}_2$ of strategies $(1;0;0)$ and $(0;x;0)$ when playing against $(1;0;0)$:

$$\begin{aligned}\mathrm{E}_{11} &= (1-2p_{\text{see}})R + p_{\text{see}}T + p_{\text{see}}S = R + p_{\text{see}}(T-R) + p_{\text{see}}(S-R),\\ \mathrm{E}_{21} &= (1-2p_{\text{see}})T + p_{\text{see}}(xR + (1-x)T) + p_{\text{see}}P = T - p_{\text{see}}(T-R) - xp_{\text{see}}(T-R).\end{aligned}$$

Now it can be easily seen that $\mathrm{E}_{11} \leq \mathrm{E}_{21}$ holds whenever inequality (3) is satisfied. □

**One-shot transparent Prisoner's Dilemma with unequal reaction times.** Here we consider the case when players have unequal probabilities to see partner's choice. We focus on a simple example showing why waiting is generally beneficial in the transparent iPD. Assume that all players in population act as quickly as they can, but cooperation takes on average longer than defection. Assume further that a player preparing to cooperate may see the partner defecting and then it is still possible for the player to change decision and defect. Finally let us consider only pure strategies that is $s_1; s_2; s_3 \in \{0,1\}$. The question now is, which strategy would win in this case.

From Lemma 1, we know that is sufficient to consider two strategies: "cooperators" $\mathbf{s}_1 = (1;0;0)$ and "defectors" $\mathbf{s}_2 = (0;0;0)$ since they dominate all other strategies. Note that the probability $p^{12}_{\text{see}}$ of cooperative players to see the choice of defectors is higher than the probability $p^{21}_{\text{see}}$ of defectors to see the choice of cooperators, resulting in $0 \leq p^{21}_{\text{see}} < p^{12}_{\text{see}} \leq 1$. Probabilities of a player to see the choice of another player with the same strategy is below 0.5 (since these probabilities are equal for both players and the sum of these probabilities is below 1), therefore it holds $0 \leq p^{11}_{\text{see}}, p^{22}_{\text{see}} \leq 0.5$.

Then the expected payoff matrix for these two strategies in the one-shot transparent PD is given by

$$\begin{aligned}\mathbf{E} &= \begin{pmatrix} (1-2p^{11}_{\text{see}})R + p^{11}_{\text{see}}(S+T) & p^{12}_{\text{see}}P + (1-p^{12}_{\text{see}})S \\ p^{12}_{\text{see}}P + (1-p^{12}_{\text{see}})T & P \end{pmatrix}\\ &= \begin{pmatrix} R - p^{11}_{\text{see}}(2R-S-T) & p^{12}_{\text{see}}P + (1-p^{12}_{\text{see}})S \\ T - p^{12}_{\text{see}}(T-P) & P \end{pmatrix},\end{aligned}$$

Since $S + p^{12}_{\text{see}}(P-S) < P$ for $p^{12}_{\text{see}} < 1$, three variants are possible:

1. cooperative strategy $\mathbf{s}_1$ dominates for $p^{12}_{\text{see}} = 1$;
2. $\mathbf{s}_1$ and $\mathbf{s}_2$ are bistable for $\mathrm{E}_{11} > \mathrm{E}_{12}$, that is for

$$p^{12}_{\text{see}} > \frac{T-R}{T-P} + \frac{2R+S-T}{T-P}p^{11}_{\text{see}}; \tag{4}$$

3. defecting strategy $\mathbf{s}_2$ dominates otherwise.

For the standard Prisoner's Dilemma payoff matrix (Figure 1, main text), inequality (4) turns into $p^{12}_{\text{see}} > \frac{1}{2} + \frac{1}{4}p^{11}_{\text{see}}$. Since $p^{11}_{\text{see}} < 0.5$, cooperative strategy $\mathbf{s}_1$ acting with a delay has a chance to win over defectors if it can see their actions with probability $p^{12}_{\text{see}} > 5/8$. This example demonstrates that cooperation can survive in one-shot Prisoner's dilemma under certain (artificial) assumptions. More importantly, this example shows the importance of seeing partner's choice in transparent Prisoner's Dilemma in general, illustrating the incentive of players to wait for partner's action.

**One-shot transparent Bach-or-Stravinsky game.** Recall [55] that in the classic one-shot BoS game there are three Nash Equilibria: two pure (Player 1 insists, Player 2 accommodates, or vice verse) and one mixed (each player insists with probability $\frac{S-P}{S+T-P-R}$). The latter NE is weak and suboptimal compared to the pure NE; yet it is fair in the sense that both players receive the same payoff. The Nash equilibria for the transparent BoS game are specified by the following proposition.

**Proposition 4.** *Consider one-shot transparent BoS between Players 1 and 2 with probabilities to see the choice of the partner $p^{12}_{\text{see}}$ and $p^{21}_{\text{see}}$, respectively. Let $p^{12}_{\text{see}} \le p^{21}_{\text{see}}$, then this game has the following pure strategy NE.*

1. *Player 1 uses strategy $(0;0;1)$, Player 2 uses strategy $(1;0;1)$ – for*

$$\frac{p^{21}_{\text{see}}}{1-p^{12}_{\text{see}}} \le \frac{T-R}{S-R}; \tag{5}$$

2. *Player 1 uses strategy $(1;0;1)$, Player 2 uses strategy $(0;0;1)$ – for*

$$\frac{p^{12}_{\text{see}}}{1-p^{21}_{\text{see}}} \le \frac{T-R}{S-R} \tag{6}$$

   *(note that this inequality holds automatically if (5) holds);*

3. *Both players use strategy $(1;0;1)$ – when (6) is not satisfied.*

*Additionally, if inequality (5) is satisfied, there is also a mixed-strategy NE: Player $i$ uses strategy $(x_i;0;1)$ with*

$$x_1 = \frac{(1-p^{21}_{\text{see}})(S-P) - p^{12}_{\text{see}}(T-P)}{(1-p^{12}_{\text{see}}-p^{21}_{\text{see}})(T+S-P-R)}, \quad x_2 = \frac{(1-p^{12}_{\text{see}})(S-P) - p^{21}_{\text{see}}(T-P)}{(1-p^{12}_{\text{see}}-p^{21}_{\text{see}})(T+S-P-R)}. \tag{7}$$

*Thus when (5) holds, there are two pure-strategy and one mixed-strategy NE. Otherwise there is only one pure-strategy NE: Player 1 uses strategy $(1;0;1)$, Player 2 uses strategy $(0;0;1)$ when (6) holds, and both Players use $(1;0;1)$ when (6) does not hold.*

To prove the Proposition, we need two lemmas. First, similar to the Prisoner's dilemma, for the transparent BoS we have:

**Lemma 5.** *In one-shot transparent BoS any strategy $(s_1;s_2;s_3)$ is dominated by strategies $(s_1;s_2;1)$ and $(s_1;0;s_3)$. The dominance of $(s_1;s_2;1)$ is strong when $s_1 < 1$, the dominance of $(s_1;0;s_3)$ is strong when $s_1 > 0$.*

The proof is identical to the proof of Lemma 1.

**Lemma 6.** *In one-shot transparent BoS, when Player 1 uses strategy $(1;0;1)$, the best response for Player 2 is to use strategy $(0;0;1)$ for $\frac{p^{12}_{\text{see}}}{1-p^{21}_{\text{see}}} \le \frac{T-R}{S-R}$ and to use $(1;0;1)$ otherwise.*

*Proof.* By Lemma 5 the best response for Player 2 is a strategy $(s_1;0;1)$ with $0 \le s_1 \le 1$. When Player 2 uses this strategy against $(1;0;1)$, the expected payoff of Player 2 is given by

$$\begin{aligned} \mathrm{E}_{21} &= (1-p^{12}_{\text{see}}-p^{21}_{\text{see}})(s_1R+(1-s_1)T) + p^{12}_{\text{see}}(s_1S+(1-s_1)T) + p^{21}_{\text{see}}T \\ &= T + s_1(p^{12}_{\text{see}}(S-R) + p^{21}_{\text{see}}(T-R) - (T-R)). \end{aligned}$$

Thus the payoff of Player 2 depends linearly on the value of $s_1$ and is maximized by $s_1 = 0$ if

$$p^{12}_{\text{see}}(S-R) - (1-p^{21}_{\text{see}})(T-R) < 0 \tag{8}$$

and by $s_1 = 1$ otherwise. Inequality (8) is equivalent to (6), which completes the proof. □

Using Lemmas 5 and 6, we can now compute NE for the one-shot transparent BoS:

*Proof.* Pure strategy NEs are obtained immediately from Lemma 6. To compute the mixed-strategy NE, recall that Player 1 achieves it when the expected payoff obtained by Player 2 for insisting and accommodating is equal:

$$(1 - p^{12}_{\text{see}} - p^{21}_{\text{see}})(x_1 R + (1 - x_1)S) + p^{12}_{\text{see}} S = (1 - p^{12}_{\text{see}} - p^{21}_{\text{see}})(x_1 T + (1 - x_1)P) + p^{12}_{\text{see}} T.$$

By computing $x_1$ from this equation and applying the same argument for Player 2, we get the strategy entries given in (7). □

**Corollary 7.** *Consider one-shot transparent BoS with $S = 4, T = 3, R = 2, P = 1$, where both players have equal probabilities $p_{\text{see}}$ to see the choice of the partner. In this game there are three NE for $p_{\text{see}} < 1/3$: (a) Player 1 uses strategy $(1; 0; 1)$, Player 2 uses strategy $(0; 0; 1)$; (b) vice versa; (c) both players use strategy $(x; 0; 1)$, with $x = \frac{3}{4} + \frac{p_{\text{see}}}{4-8p_{\text{see}}}$. For $p_{\text{see}} \geq 1/3$, $(1; 0; 1)$ is the only NE.*

## Analysis of iterated transparent games

For the analysis of iterated games we use the techniques described in [23,24,56]. Since most of results for simultaneous and sequential iPD were obtained for strategies taking into account outcomes of the last interaction ("memory-one strategies"), here we also focus on memory-one strategies. Note that considering multiple previous round results in very complex strategies. To overcome this, one can, for instance, use pure strategies (see, for instance, [20]), but we reserve this possibility for future research.

Consider an infinite population of players evolving in generations. For any generation $t = 1, 2, \ldots$ the population consists of $n(t)$ "species" defined by their strategies $\mathbf{s}_i = (s^i_k)^{12}_{k=1}$ and their frequencies $x_i(t)$ in the population, $\sum\limits_{i=1}^{n(t)} x_i(t) = 1$. Besides, the probability of a player from species $i$ to see the choice of a partner from species $j$ is given by $p^{ij}_{\text{see}} \in [0, 1]$ (in our case $p^{ij}_{\text{see}} = p_{\text{see}}$ for all species $i$ and $j$, but in this section we use the general notation).

Consider a player from species $i$ playing an infinitely long iterated game against a player from species $j$. Since both players use memory-one strategies, this game can be formalized as a Markov chain with states being the mutual choices of the two players and a transition matrix $M$ given by

$$M = (1 - p^{ij}_{\text{see}} - p^{ji}_{\text{see}})M_0 + p^{ij}_{\text{see}} M_1 + p^{ji}_{\text{see}} M_2, \tag{9}$$

where the matrices $M_0$, $M_1$ and $M_2$ describe the cases when neither player sees the choice of the partner, Player 1 sees the choice of the partner before making own choice, and Player 2 sees the choice of the partner, respectively. These matrices are given by

$$M_0 = \begin{pmatrix} s^i_1 s^j_1 & s^i_1(1 - s^j_1) & (1 - s^i_1)s^j_1 & (1 - s^i_1)(1 - s^j_1) \\ s^i_2 s^j_3 & s^i_2(1 - s^j_3) & (1 - s^i_2)s^j_3 & (1 - s^i_2)(1 - s^j_3) \\ s^i_3 s^j_2 & s^i_3(1 - s^j_2) & (1 - s^i_3)s^j_2 & (1 - s^i_3)(1 - s^j_2) \\ s^i_4 s^j_4 & s^i_4(1 - s^j_4) & (1 - s^i_4)s^j_4 & (1 - s^i_4)(1 - s^j_4) \end{pmatrix},$$

$$M_1 = \begin{pmatrix} s^i_5 s^j_1 & s^i_9(1 - s^j_1) & (1 - s^i_5)s^j_1 & (1 - s^i_9)(1 - s^j_1) \\ s^i_6 s^j_3 & s^i_{10}(1 - s^j_3) & (1 - s^i_6)s^j_3 & (1 - s^i_{10})(1 - s^j_3) \\ s^i_7 s^j_2 & s^i_{11}(1 - s^j_2) & (1 - s^i_7)s^j_2 & (1 - s^i_{11})(1 - s^j_2) \\ s^i_8 s^j_4 & s^i_{12}(1 - s^j_4) & (1 - s^i_8)s^j_4 & (1 - s^i_{12})(1 - s^j_4) \end{pmatrix},$$

$$M_2 = \begin{pmatrix} s^i_1 s^j_5 & s^i_1(1 - s^j_5) & (1 - s^i_1)s^j_9 & (1 - s^i_1)(1 - s^j_9) \\ s^i_2 s^j_7 & s^i_2(1 - s^j_7) & (1 - s^i_2)s^j_{11} & (1 - s^i_2)(1 - s^j_{11}) \\ s^i_3 s^j_6 & s^i_3(1 - s^j_6) & (1 - s^i_3)s^j_{10} & (1 - s^i_3)(1 - s^j_{10}) \\ s^i_4 s^j_8 & s^i_4(1 - s^j_8) & (1 - s^i_4)s^j_{12} & (1 - s^i_4)(1 - s^j_{12}) \end{pmatrix}.$$

The gain of species $i$ when playing against species $j$ is given by the expected payoff $\mathrm{E}_{ij}$, defined by

$$\mathrm{E}_{ij} = y_R R + y_S S + y_T T + y_P P, \tag{10}$$

where $R, S, T, P$ are the entries of the payoff matrix ($R = 3, S = 0, T = 5, P = 1$ for standard iPD and $R = 2, S = 4, T = 3, P = 1$ for iBoS, see Fig. 1), and $y_R, y_S, y_T, y_P$ represent the probabilities of getting to the states associated with the corresponding payoffs by playing $\mathbf{s}_i$ against $\mathbf{s}_j$. This vector is computed as a unique left-hand eigenvector of matrix $M$ associated with eigenvalue one [24]:

$$(y_R, y_S, y_T, y_P) = (y_R, y_S, y_T, y_P)M.$$

The evolutionary success of species $i$ is encoded by its fitness $f_i(t)$: if species $i$ has higher fitness than the average fitness of the population $\overline{f}(t) = \sum_{i=1}^{n(t)} x_i(t) f_i(t)$, then $x_i(t)$ increases with time, otherwise $x_i(t)$ decreases and the species is dying out. This evolutionary process is formalized by the replicator dynamics equation, which in discrete time takes the form

$$x_i(t+1) = x_i(t) + \frac{f_i(t) - \overline{f}(t)}{\overline{f}(t)} x_i(t) = \frac{f_i(t)}{\overline{f}(t)} x_i(t). \tag{11}$$

The fitness $f_i(t)$ is computed as the average payoff for a player of species $i$ when playing against the current population:

$$f_i(t) = \sum_{j=1}^{n(t)} x_j(t) \mathrm{E}_{ij},$$

where $\mathrm{E}_{ij}$ is given by (10).

**Evolutionary dynamics of two strategies.** To provide an example of evolutionary dynamics and introduce some useful notation, we consider a population consisting of two species playing iPD with strategies: $\mathbf{s}_1 = (1, 0, 0, 1;\ 1, 0, 0, 1;\ 0, 0, 0, 0)$, $\mathbf{s}_2 = (0, 0, 0, 0;\ 0, 0, 0, 0;\ 0, 0, 0, 0)$ (recall that we write 0 instead of $\varepsilon$ and 1 instead of $1 - \varepsilon$) and initial conditions $x_1(1) = x_2(1) = 0.5$. That is, the first species plays the "Win–stay, lose–shift" (WSLS) strategy, and the second species always defects (uses the AllD strategy). We set $p^{11}_{\text{see}} = p^{12}_{\text{see}} = p^{21}_{\text{see}} = p^{22}_{\text{see}} = p_{\text{see}}$. Note that since $p^{11}_{\text{see}}, p^{22}_{\text{see}} \leq 0.5$ and $p^{12}_{\text{see}} + p^{21}_{\text{see}} \leq 1$, it holds $p_{\text{see}} \leq 0.5$. Given $p_{\text{see}}$ we can compute a transition matrix of the game using (9) and then calculate the expected payoffs for all possible pairs of players $ij$ using (10). For instance, for $p_{\text{see}} = 0$ we have

$$\mathrm{E}_{11} = 2.995, \mathrm{E}_{12} = 0.504, \mathrm{E}_{21} = 2.999, \mathrm{E}_{22} = 1.003.$$

This means that a player from the WSLS-species on average gets a payoff $\mathrm{E}_{11} = 2.995$ when playing against a conspecific partner, and only $\mathrm{E}_{12} = 0.504$, when playing against an AllD-player. The fitness for each species is given by

$$f_1(t) = x_1(t)\mathrm{E}_{11} + x_2(t)\mathrm{E}_{12} = 2.995x_1(t) + 0.504x_2(t),$$
$$f_2(t) = x_1(t)\mathrm{E}_{21} + x_2(t)\mathrm{E}_{22} = 2.999x_1(t) + 1.003x_2(t).$$

Since $f_2(t) > f_1(t)$ for any $0 < x_1(t), x_2(t) < 1$, the AllD-players take over the whole population after several generations. Dynamics of the species frequencies $x_i(t)$ computed using (11) shows that this is indeed the case (Fig. 8a). Note that since $\mathrm{E}_{21} > \mathrm{E}_{11}$ and $\mathrm{E}_{22} > \mathrm{E}_{12}$, AllD is garanteed to win over WSLS for any initial frequency of WSLS-players $x_1(1)$. In this case one says that AllD *dominates* WSLS and can *invade* it for any $x_1(1)$.

As we increase $p_{\text{see}}$, the population dynamics changes. While for $p_{\text{see}} = 0.2$ AllD still takes over the population, for $p_{\text{see}} = 0.4$ WSLS wins (Fig. 8a). This can be explained by computing the expected payoff for $p_{\text{see}} = 0.4$:

$$\mathrm{E}_{11} = 2.995, \mathrm{E}_{12} = 0.628, \mathrm{E}_{21} = 2.500, \mathrm{E}_{22} = 1.003.$$

Hence $f_1(t) > f_2(t)$ for $0 \leq x_2(t) \leq 0.5 \leq x_1(t) \leq 0$, which explains the observed dynamics. Note that here $\mathrm{E}_{11} > \mathrm{E}_{21}$, while $\mathrm{E}_{12} < \mathrm{E}_{22}$, that is a conspecific partner wins more than a partner from

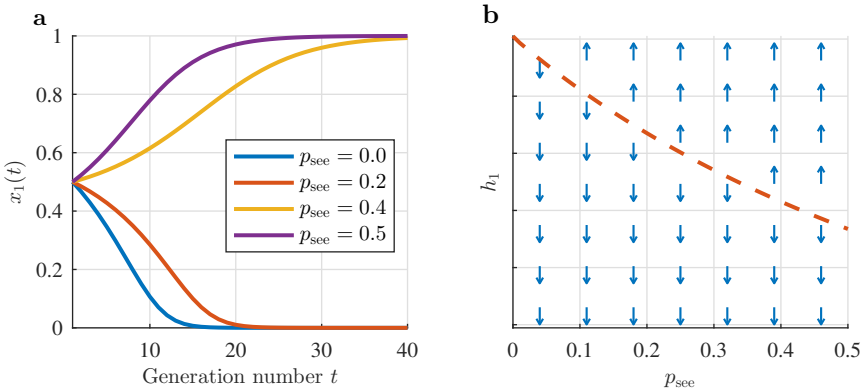



Figure 8: **Evolutionary dynamics of iPD-population consisting of species with WSLS and AllD strategies.** **a** Initially, both species have the same frequency, but after 40 generations the fraction of WSLS-players $x_1(t)$ converges to 0 for probabilities to see partner's choice $p_{\text{see}} = 0.0, 0.2$ and to 1 for $p_{\text{see}} = 0.4, 0.5$. **b** This is due to the decrease of the invasion threshold $h_1$ for WSLS: while $h_1 = 1$ for $p_{\text{see}} = 0$ (AllD dominates WSLS and the fraction of WSLS-players unconditionally decreases), AllD and WSLS are bistable for $p_{\text{see}} > 0$ and WSLS wins whenever $x_1(t) > h_1$. Arrows indicate whether frequency $x_1(t)$ of WSLS increases or decreases. Interestingly, $h_1 = 0.5$ holds for $p_{\text{see}} \approx 1/3$, which corresponds to the maximal uncertainty since the three cases ("Player 1 knows the choice of Player 2 before making its own choice"; "Player 2 knows the choice of Player 1 before making its own choice"; "Neither of players knows the choice of the partner") have equal probabilities.

a different species when playing against WSLS- and AllD-players alike. In this case one says that WSLS and AllD are *bistable* and there is an unstable equilibrium fraction of WSLS players given by

$$h_1 = \frac{\mathrm{E}_{22} - \mathrm{E}_{12}}{\mathrm{E}_{11} - \mathrm{E}_{12} - \mathrm{E}_{21} + \mathrm{E}_{22}}. \tag{12}$$

We call $h_i$ an *invasion threshold* for species $i$, since it takes over the whole population for $x_i(t) > h_i$, but dies out for $x_i(t) < h_i$. To illustrate this concept, we plot in Fig. 8b the invasion threshold $h_1$ for WSLS species playing against AllD as a function of $p_{\text{see}}$.

One more possible type of two-species dynamics is *coexistence*, which takes place when $\mathrm{E}_{11} < \mathrm{E}_{21}$, $\mathrm{E}_{12} > \mathrm{E}_{22}$, that is when playing against a player from any of the species is less beneficial for a conspecific partner than for a partner from a different species. In this case the fraction of a species given by (12) corresponds to a stable equilibrium meaning that the frequency of the first species $x_1(t)$ increases for $x_1(t) < h_1$, but decreases for $x_1(t) > h_1$.

**Evolutionary simulations for transparent games.** Theoretical analysis of the strategies in repeated transparent games is complicated due to the many dimensions of the strategy space, which motivates using of evolutionary simulations. For this we adopt the methods described in [23,24]. Each run of simulations starts with five species having equal initial frequencies: $n(1) = 5$, $x_1(1) = \ldots = x_5(1) = 0.2$. Following [23], probabilities $s_k^i$ with $k = 1, \ldots, 12$ for these species are randomly drawn from the distribution with U-shaped probability density, favouring probability values around 0 and 1:

$$\rho(y) = \pi\big(y(1-y)\big)^{-1/2} \tag{13}$$

for $y \in (0, 1)$. Additionally, we require $s_k^i \in [\varepsilon, 1-\varepsilon]$, where $\varepsilon = 0.001$ accounts for the minimal possible error in the strategies [23]. The frequencies of strategies $x_i(t)$ change according to the replicator dynamics equation (11). If $x_i(t) < \chi$, the species is assumed to die out and is removed from the population (share $x_i(t)$ is distributed proportionally among the remaining species); we follow [23, 24] in taking $\chi = 0.001$. Occasionally (every 100 generations on average to avoid strong synchronization), new species are entered in the population. The strategies for the new species are drawn from (13) and the initial frequencies are set to $x_i(t_0) = 1.1\chi$ [23].

## Data availability

The empirical datasets generated during the current study and the source code used for this are available from the corresponding author on reasonable request.

## Acknowledgements

We acknowledge funding from the Ministry for Science and Education of Lower Saxony and the Volkswagen Foundation through the program "Niedersächsisches Vorab". Additional support was provided by the Leibniz Association through funding for the Leibniz ScienceCampus Primate Cognition and the Max Planck Society.

## Contributions

A.U. conceived the study and performed simulations with the help and advice of S.E. and F.W.; T.S., I.K. and S.M. contributed to the interpretation of the results. All authors contributed to writing the manuscript.

## Competing interests

The authors declare no competing financial interests.

## References

[1] Axelrod, R. On six advances in cooperation theory. *Analyse & Kritik* **22**, 130–151 (2000).

[2] Nowak, M. & Sigmund, K. The alternating prisoner's dilemma. *Journal of theoretical Biology* **168**, 219–226 (1994).

[3] Dugatkin, L., Mesterton-Gibbonsand, M. & Houston, A. Beyond the Prisoner's dilemma: Toward models to discriminate among mechanisms of cooperation in nature. *Trends in ecology & evolution* **7**, 202–205 (1992).

[4] Brosnan, S., Wilson, B. J. & Beran, M. Old world monkeys are more similar to humans than new world monkeys when playing a coordination game. *Proceedings of the Royal Society of London B: Biological Sciences* **279**, 1522–1530 (2012).

[5] Duguid, S., Wyman, E., Bullinger, A., Herfurth-Majstorovic, K. & Tomasello, M. Coordination strategies of chimpanzees and human children in a stag hunt game. *Proceedings of the Royal Society of London B: Biological Sciences* **281**, 20141973 (2014).

[6] Hawkins, R. & Goldstone, R. The formation of social conventions in real-time environments. *PlOS ONE* **11**, e0151670 (2016).

[7] Vaziri-Pashkam, M., Cormiea, S. & Nakayama, K. Predicting actions from subtle preparatory movements. *Cognition* **168**, 65–75 (2017).

[8] Silk, J. The strategic dynamics of cooperation in primate groups. *Advances in the Study of Behavior* **37**, 1–41 (2007).

[9] Sueur, C. & Petit, O. Signals use by leaders in macaca tonkeana and macaca mulatta: group-mate recruitment and behaviour monitoring. *Animal cognition* **13**, 239–248 (2010).

[10] King, A. & Sueur, C. Where next? Group coordination and collective decision making by primates. *International Journal of Primatology* **32**, 1245–1267 (2011).

[11] Fichtel, C., Zucchini, W. & Hilgartner, R. Out of sight but not out of mind? behavioral coordination in red-tailed sportive lemurs (lepilemur ruficaudatus). *International journal of primatology* **32**, 1383–1396 (2011).

[12] Strandburg-Peshkin, A., Papageorgiou, D., Crofoot, M. & Farine, D. Inferring influence and leadership in moving animal groups. *Philosophical Transactions B: Biological Sciences* **373**, 2170006 (2017).

[13] Chiappori, P.-A., Levitt, S. & Groseclose, T. Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review* **92**, 1138–1151 (2002).

[14] Palacios-Huerta, I. Professionals play minimax. *The Review of Economic Studies* **70**, 395–415 (2003).

[15] Bar-Eli, M., Azar, O., Ritov, I., Keidar-Levin, Y. & Schein, G. Action bias among elite soccer goalkeepers: The case of penalty kicks. *Journal of economic psychology* **28**, 606–621 (2007).

[16] Axelrod, R. & Hamilton, W. The evolution of cooperation. *Science* **211**, 13–90 (1981).

[17] Rapoport, A. Exploiter, Leader, Hero, and Martyr: the four archetypes of the 2× 2 game. *Systems Research and Behavioral Science* **12**, 81–84 (1967).

[18] Colman, A. *Game theory and its applications: In the social and biological sciences. 2nd ed.* (London: Routledge, 2005).

[19] Helbing, D., Schönhof, M., Stark, H.-U. & Hołyst, J. How individuals learn to take turns: Emergence of alternating cooperation in a congestion game and the prisoner's dilemma. *Advances in Complex Systems* **8**, 87–116 (2005).

[20] Colman, A. & Browning, L. Evolution of cooperative turn-taking. *Evolutionary Ecology Research* **11**, 949–963 (2009).

[21] Noë, R. A veto game played by baboons: a challenge to the use of the Prisoner's Dilemma as a paradigm for reciprocity and cooperation. *Animal Behaviour* **39**, 78–90 (1990).

[22] Noë, R. Cooperation experiments: coordination through communication versus acting apart together. *Animal Behaviour* **71**, 1–18 (2006).

[23] Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* **364**, 56–58 (1993).

[24] Nowak, M. *Evolutionary dynamics* (Harvard University Press, 2006).

[25] Tomasello, M., Melis, A., Tennie, C., Wyman, E. & Herrmann, E. Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current anthropology* **53**, 673–692 (2012).

[26] Traulsen, A. & Hauert, C. Stochastic evolutionary game dynamics. *Reviews of nonlinear dynamics and complexity* **2**, 25–61 (2009).

[27] Selten, R. Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory* **4**, 25–55 (1975).

[28] Frean, M. The prisoner's dilemma without synchrony. *Proceedings of the Royal Society of London B: Biological Sciences* **257**, 75–79 (1994).

[29] Zagorsky, B., Reiter, J., Chatterjee, K. & Nowak, M. Forgiver triumphs in alternating Prisoner's Dilemma. *PlOS ONE* **8**, e80814 (2013).

[30] Friedman, D. Evolutionary games in economics. *Econometrica: Journal of the Econometric Society* 637–666 (1991).

[31] Posch, M., Pichler, A. & Sigmund, K. The efficiency of adapting aspiration levels. *Proceedings of the Royal Society of London B: Biological Sciences* **266**, 1427–1435 (1999).

[32] Friedman, D. & Oprea, R. A continuous dilemma. *The American Economic Review* **102**, 337–363 (2012).

[33] Melis, A., Grocke, P., Kalbitz, J. & Tomasello, M. One for you, one for me: Humans' unique turn-taking skills. *Psychological science* **27**, 987–996 (2016).

[34] Sánchez-Amaro, A., Duguid, S., Call, J. & Tomasello, M. Chimpanzees, bonobos and children successfully coordinate in conflict situations. In *Proc. R. Soc. B*, vol. 284, 20170259 (The Royal Society, 2017).

[35] Bárány, I., Vempala, S. & Vetta, A. Nash equilibria in random games. *Random Structures & Algorithms* **31**, 391–405 (2007).

[36] Arieli, I. & Babichenko, Y. Random extensive form games. *Journal of Economic Theory* **166**, 517–535 (2016).

[37] Posch, M. Win–Stay, Lose–Shift Strategies for Repeated Games—Memory Length, Aspiration Levels and Noise. *Journal of theoretical biology* **198**, 183–195 (1999).

[38] Clements, K. & Stephens, D. Testing models of non-kin cooperation: mutualism and the prisoner's dilemma. *Animal Behaviour* **50**, 527–535 (1995).

[39] Chang, S. An emerging field of primate social neurophysiology: Current developments. *eNeuro* **4**, ENEURO–0295–17 (2017).

[40] Ruff, C. & Fehr, E. The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience* **15**, 549 (2014).

[41] Platt, M., Seyfarth, R. & Cheney, D. Adaptations for social cognition in the primate brain. *Phil. Trans. R. Soc. B* **371**, 20150096 (2016).

[42] Tremblay, S., Sharika, K. & Platt, M. Social decision-making and the brain: A comparative perspective. *Trends in cognitive sciences* **21**, 265–276 (2017).

[43] Ballesta, S. & Duhamel, J.-R. Rudimentary empathy in macaques' social decision-making. *Proceedings of the National Academy of Sciences* **112**, 15516–15521 (2015).

[44] Báez-Mendoza, R. & Schultz, W. Performance error-related activity in monkey striatum during social interactions. *Scientific Reports* **6**, 37199 (2016).

[45] Haroush, K. & Williams, Z. Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell* **160**, 1233–1245 (2015).

[46] Bullinger, A., Wyman, E., Melis, A. & Tomasello, M. Coordination of chimpanzees (pan troglodytes) in a stag hunt game. *International Journal of Primatology* **32**, 1296–1310 (2011).

[47] Brosnan, S. *et al.* Responses to the assurance game in monkeys, apes, and humans using equivalent procedures. *Proceedings of the National Academy of Sciences* **108**, 3442–3447 (2011).

[48] Visco-Comandini, F. *et al.* Do non-human primates cooperate? evidences of motor coordination during a joint action task in macaque monkeys. *Cortex* **70**, 115–127 (2015).

[49] Sánchez-Amaro, A., Duguid, S., Call, J. & Tomasello, M. Chimpanzees coordinate in a snowdrift game. *Animal Behaviour* **116**, 61 – 74 (2016). URL `http://www.sciencedirect.com/science/article/pii/S0003347216300045`.

[50] Brosnan, S. *et al.* Human and monkey responses in a symmetric game of conflict with asymmetric equilibria. *Journal of Economic Behavior & Organization* **142**, 293–306 (2017).

[51] King, A., Johnson, D. & Van Vugt, M. The origins and evolution of leadership. *Current biology* **19**, R911–R916 (2009).

[52] Devaine, M., Hollard, G. & Daunizeau, J. Theory of mind: did evolution fool us? *PloS One* **9**, e87619 (2014).

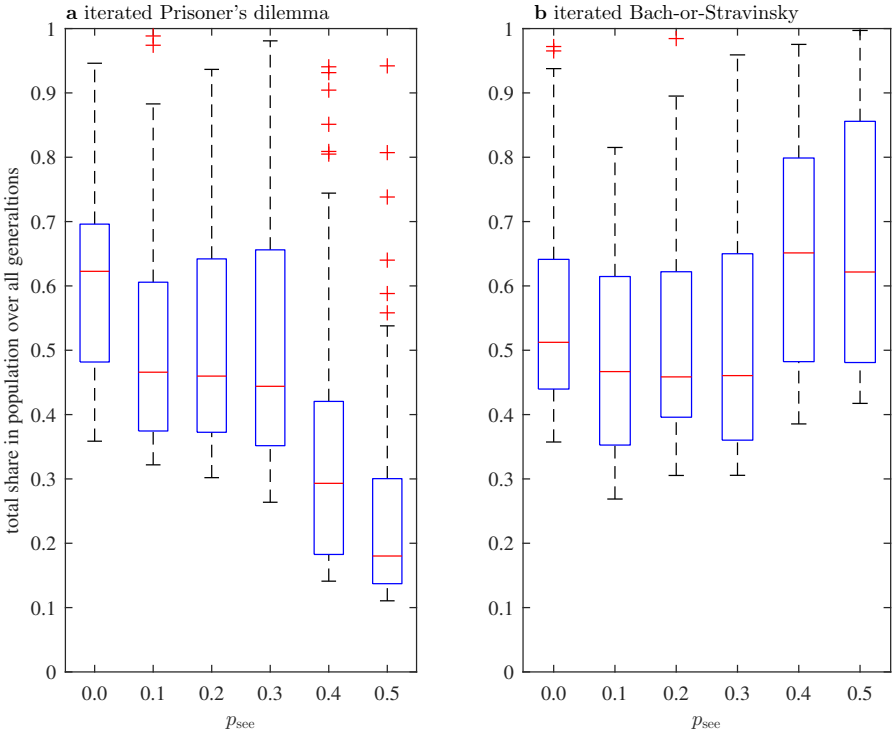
[53] Luce, R. *Response times: Their role in inferring elementary mental organization* (Oxford University Press, 1986).

[54] Ratcliff, R. Methods for dealing with reaction time outliers. *Psychological bulletin* **114**, 510 (1993).

[55] Rapoport, A. *Two-person game theory* (NY: Dover Publications Inc, 1973).

[56] Akin, E. The iterated prisoner's dilemma: good strategies and their dynamics. In *Ergodic Theory, Advances in Dynamical Systems*, 77–107 (2016).

## Supplementary Figures

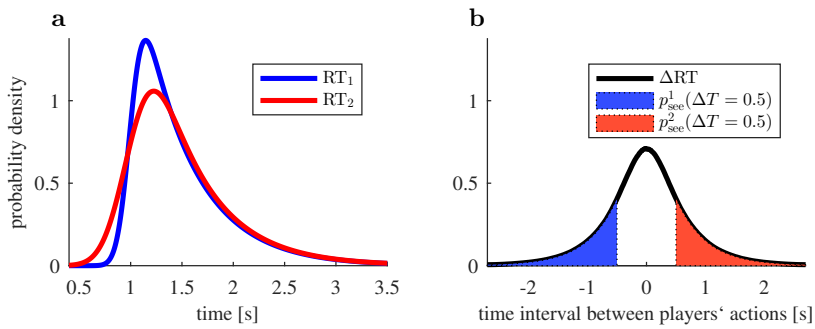



Supplementary Figure 1: **Distributions of total shares in the population over all generations for 80 most persistent species over the 80 runs of evolutionary simulations a** for iterated Prisoner's Dilemma (iPD) and **b** for iterated Bach-or-Stravinsky game (iBoS). The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol. The higher total shares of the species are, the more stable the dynamics in the population is. While stability varies with transparency for both games, the drop of stability in iPD for $p_{\text{see}} > 0.3$ is especially noticeable. Indeed, in highly transparent iPD any strategy is sufficiently "predictable", which allows a best-response strategy to replace it in a population. Such best-response strategies can be generally weak and short-living, see for example treacherous WSLS described in Figure 4 (main text). Note that stability increases considerably for $p_{\text{see}} > 0.3$ in iBoS, which reflects the fact that Leader-Follower strategy becomes evolutionary stable for high transparency.

## Supplementary Note 1

In the Methods section we argue that evolution favours equal reaction times both in iPD and iBoS, since the optimal behaviour in iPD is to wait as long as possible, and in iBoS – to act as quickly as possible. However, for iPD there is a notable exception: the Leader-Follower (L-F) strategy is better of when acting fast and exposing it's choice to the partner. Consider, for instance, a population consisting of L-F players from two species, the first acting fast and the second waiting. In all inter-species interactions, players from the first species have an upper hand since they take the role of Leaders, maximizing own payoff. Thus the first species dominates the second and finally takes over the population. The question then is, whether this contradiction to the general rule for the transparent iPD (to wait as long as possible) changes the simulation results?

Additional simulations show that this is not the case. We have used the same evolutionary simulations as before with one modification. Instead of using for all species a fixed probability to see the partner's choice $p_{\text{see}}$, we computed this probability for each pair of species as shown in Supplementary Fig. 2: from the reaction times (RT) modelled by exponentially modified Gaussian distributions and from the visibility threshold $\Delta T$.




Supplementary Figure 2: **Distributions of reaction times (RT) of the players determine their probability to see the partner's choice. a** RT distributions for two players modelled by exponentially modified Gaussian distribution. **b** Distribution of RT difference $\Delta\text{RT} = \text{RT}_2 - \text{RT}_1$ and probabilities to see the partner's choice given by $p^{12}_{\text{see}} = \Pr(\Delta\text{RT} < -\Delta T)$ (Player 1 knows the choice of Player 2 before making own choice, the blue area) and $p^{21}_{\text{see}} = \Pr(\Delta\text{RT} > \Delta T)$ (vice versa, the red area), where $\Delta T$ is a time interval required for a player to interpret and act on the partner's choice.

Exponentially modified Gaussian distribution has three parameters: mean of Gaussian component $\mu$, standard deviation of Gaussian component $\sigma$ and relaxation time of exponential component $\tau$. For each species a random mean reaction time $\mu$ was selected from the set $\{2.0, 2.1, \ldots, 3.0\}$. Since we were mainly interested in the influence of the species' mean RT on the results, we set other parameters to constants: $\sigma = 0.1$ and $\tau = 0.5$.

For each two species $i$ and $j$ we computed probabilities to see partner's choice as follows:

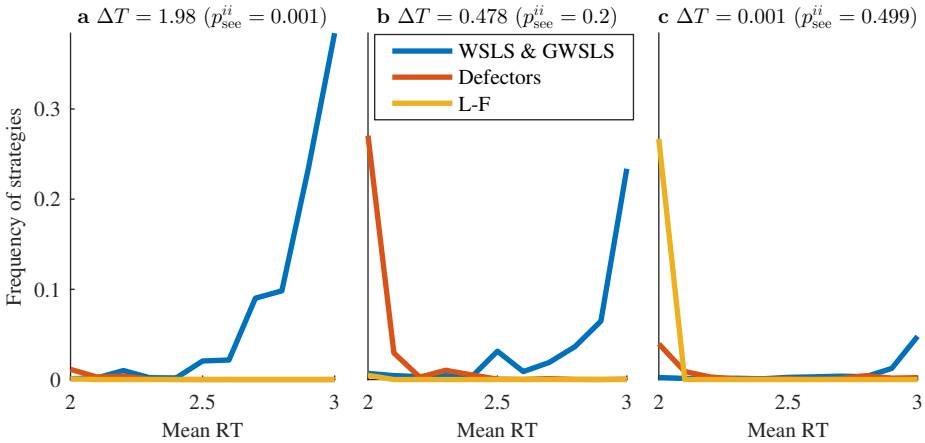
1. Using exponentially modified Gaussian distribution, we generated for each species samples of reaction times $\text{RT}_{i,k}$, $\text{RT}_{j,k}$ for $k = 1, 2, \ldots, K$ with $K = 10^6$.
2. We computed reaction time differences between species $i$ and $j$ by $\Delta\text{RT}_k = \text{RT}_{j,k} - \text{RT}_{i,k}$.
3. We estimated probabilities to see partner's choice by

$$p^{ij}_{\text{see}} = \frac{1}{K}\sharp\{k = 1, \ldots, K \mid \Delta\text{RT}_k < -\Delta T\},$$
$$p^{ji}_{\text{see}} = \frac{1}{K}\sharp\{k = 1, \ldots, K \mid \Delta\text{RT}_k > \Delta T\},$$

where $\sharp A$ stands for the number of elements in the set $A$.

We performed three series of evolutionary simulations for $\Delta T = 1.98,\ 0.478,\ 0.001$. These values were selected so that for any species $i$ probability $p^{ii}_{\text{see}}$ was equal to 0.001, 0.2 and 0.499, respectively. Each series consisted of 80 runs of evolutionary simulations, we traced $10^9$ generations in each run. Except the way the values of $p^{ij}_{\text{see}}$ were computed, the simulations were as described in the main text of the manuscript

As expected, results were similar to those with equal RT but more noisy since additional species variability increases the number of generations necessary for the population to reach the equilibrium state. In Supplementary Fig. 3a, for low (but non-zero) transparency WSLS wins with a total relative frequency above 85% (without GWSLS), but as transparency increases the share of WSLS drops down. On the contrary, the Leader-Follower strategy has best performance for high transparency with a relative frequency 27% (Supplementary Fig. 3c). Note that all successful species have marginal RT: WSLS-players mostly have maximal reaction times, while L-F-players have minimal reaction times.



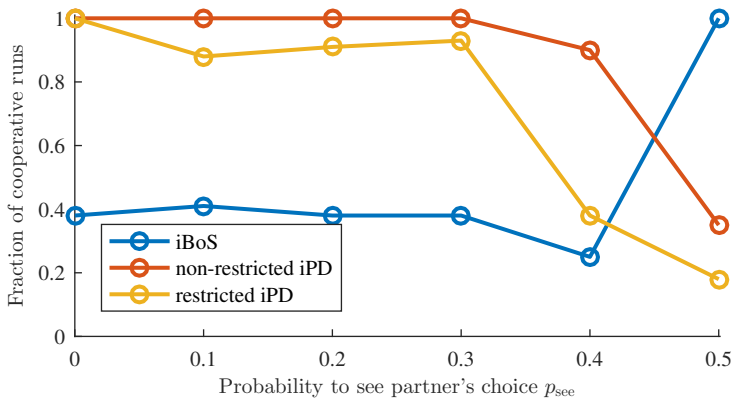

Supplementary Figure 3: **Fractions of the most frequent stable strategies in transparent iPD with unequal reaction times (RT).** RT were modelled by exponentially modified Gaussian distributions with $\mu$ randomly selected from the set $\{2.0, 2.1, \ldots, 3.0\}$, $\sigma = 0.1$ and $\tau = 0.5$. WSLS is considered here together with GWSLS, they have a strategy profile ($1abc$;1***;****) with $a, b < 2/3$, $c \geq 2/3$. We characterized as L-F all strategies with a profile (*00$b$;****;*11$c$), where $b < 1/3$ and $c < 2/3$. Finally, we considered a strategy as defecting if it has entries $s_4, s_{12} < 0.2$, $s_1, s_2, s_3 < 1/3$ and $s_8 < 2/3$. **a** For low transparencies WSLS is predominate and WSLS-players clearly prefer waiting over fast action. **b** For moderate transparencies population is controlled either by the waiting WSLS players or by the fast-acting defectors, though the latter are successful only since many strategies may have $s^i_9, s^i_{10}, s^i_{11}, s^i_{12} > 0$, resulting in cooperation with apparent defectors. **c** For high transparencies Leader-Follower outperforms defecting strategies. Note that in all cases species with marginal RT prevail and the observed strategy frequencies are similar to those for equal RT.

The only principal difference from the simulations with fixed $p_{\text{see}}$ takes place for moderate transparencies, in particular, for $\Delta T = 0.478$ when probability to see the partner's choice in intra-species interactions is given by $p^{ii}_{\text{see}} = 0.2$. Supplementary Fig. 3b shows that in this case defecting strategies have an unexpectedly high relative frequency. However, this seems to be an artefact caused by the fact that for the most species added to the population strategy entries $s^i_9, s^i_{10}, s^i_{11}, s^i_{12} > 0$ (meaning that players may cooperate even seeing that partner defects). Playing against a fast-acting defectors, these species take the role of Followers and become an easy prey. Indeed, if a defecting strategy has $\mu_i = 2$, its opponent with $\mu_j = 2.5$ sees the choice of the defector with probability $p^{ji}_{\text{see}} > 0.5$, and an opponent with $\mu_j = 3$ – with probability $p^{ji}_{\text{see}} > 0.8$. In this case probabilities $s^i_9, \ldots, s^i_{12}$ are much more important than for the case when RT are equal and these entries are used only with probability $p^{ii}_{\text{see}} = 0.2$. Fast-acting defecting strategies can be only counteracted by TFT-like strategies with $s^i_9, s^i_{10}, s^i_{11}, s^i_{12} \approx 0$. Note that the L-F strategy is not successful against defecting strategies in this case, since L-F can only survive for high $p^{ii}_{\text{see}}$.

## Supplementary Note 2

Here we introduce a variant of transparent iterated Prisoner's dilemma (iPD) with a restricted strategy space. Note that in iPD a rational player in most cases would not cooperate seeing that partner defects. The only notable exception is the Leader-Follower strategy. In general, one can see in Figure 2 (main text) that probabilities $s^i_9, \ldots, s^i_{12}$ to cooperate seeing that partner defects are quite low, especially for $p_{\text{see}} < 0.4$ (note that this takes place despite of the fact that defection for $p_{\text{see}} < 0.4$ is rare, meaning that entries $s^i_9, \ldots, s^i_{12}$ are not very important for the strategy success).

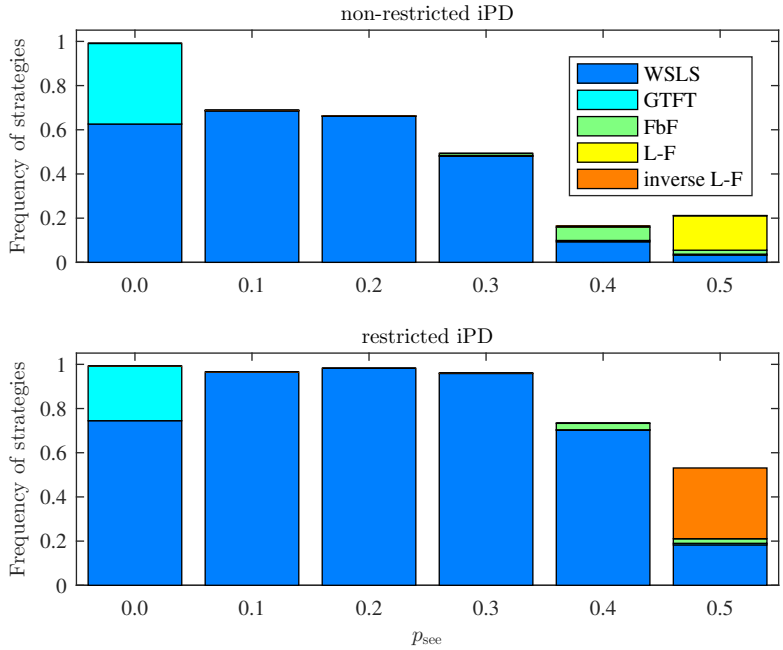
Assuming that cooperation with a defecting partner is unnatural, we can set $s^i_9 = \ldots = s^i_{12} = 0$. A question then is, whether such priors change the dynamics of the iPD-strategies. Supplementary Fig. 4 shows that restricting strategy space results in the same drop of cooperation as in the non-restricted iPD.




Supplementary Figure 4: **Fraction of runs for which cooperation was established in iterated Bach-or-Stravinsky game (iBoS) and in two types of iterated Prisoner's Dilemma (iPD).** We assumed that cooperation was established in the population if the average payoff was above $0.9 \cdot 3$ for iPD and above $0.95 \cdot 3.5$ for iBoS (90% and 95% of maximal possible value). In iPD (for both non-restricted and restricted strategy space) seeing the partner's choice adversely affects cooperation as it increases the temptation to exploit the partner. In iBoS, "evolution" results in more cooperative agents when they have a higher probability of seeing the partner's choice as this helps them to coordinate. The small drop in cooperation for iBoS at $p_{\text{see}} = 0.4$ is caused by a transition from turn-taking to leader-following.

There is however, one difference: Supplementary Fig. 5 shows that for high $p_{\text{see}}$ an "inverse Leader-Follower" strategy (inverse L-F) emerges instead of Leader-Follower introduced for the non-restricted iPD. Inverse L-F is theoretically represented by $\mathbf{s} = (1110; 0000; 0000)$, that is the player cooperates when it does not see the choice of the partner and defects otherwise. In the simultaneous iPD ($p_{\text{see}} = 0$) L-F behaves as unconditional cooperator and is easily beaten, but it becomes predominant in restricted settings for $p_{\text{see}} = 0.5$. Note that L-F is an extension of the strategy $(1; 0; 0)$, which plays a special role in one-shot PD (see "Methods" section). However, memory provides to inverse L-F an important advantage: it can distinguish unconditional defectors AllD from conspecifics. Resistance to AllD is achieved by defecting after mutual defection ($s_4 = 0$).

Spread of inverse L-F in the restricted iPD for high transparency illustrates pervasiveness of "Leader-Follower" principle. It also shows that the role of initiators can vary: in some cases, these agents reap special benefits, but in other cases they also carry the burden. Although counter-intuitive at first glance, the cooperativeness of Leaders in the L-F strategy corresponds to the behaviour of individuals that agree to do a necessary but risky or unpleasant job without immediate benefit. Examples include volunteering in human societies and acting as sentries in animal groups.




Supplementary Figure 5: **Fractions of WSLS, GTFT, FbF and L-F strategies in transparent iPD with non-restricted and restricted strategy space.** The frequencies were computed over $10^9$ generations in 80 runs; strategies were counted as stable if they survived for more than 1000 generations after they emerged in the population. Note that the striking similarities between two scenarios. The main differences include the lower stability in the non-restricted iPD and emergence of inverse L-F instead of L-F for $p_{\text{see}} = 0.5$ in restricted iPD. We classified as inverse L-F all strategies with profile (*11*; *00*; 0000) since behaviour after mutual cooperation or mutual defection is only relevant when L-F is playing against another strategy, and success for different types of behaviour depends on the composition of the population.