Evolutionary Successful Strategies in a Transparent iterated Prisoner's Dilemma

Anton M. Unakafov[1,2,3,4,6,7](✉), Thomas Schultze[1,3], Igor Kagan[3,4],
Sebastian Moeller[1,3,4], Alexander Gail[1,3,4,5], Stefan Treue[1,3,4,5],
Stephan Eule[2,3,6,7], and Fred Wolf[2,3,5,6,7]

[1] Georg-Elias-Mueller-Institute of Psychology, University of Goettingen,
Gosslerstrasse 14, 37073 Göttingen, Germany
`schultze@psych.uni-goettingen.de`
[2] Max Planck Institute for Dynamics and Self-Organization,
Am Fassberg 17, 37077 Göttingen, Germany
`{anton,eule,fred}@nld.ds.mpg.de`
[3] Leibniz ScienceCampus Primate Cognition,
Kellnerweg 4, 37077 Göttingen, Germany
[4] German Primate Center - Leibniz Institute for Primate Research,
Kellnerweg 4, 37077 Göttingen, Germany
`{ikagan,smoeller}@dpz.eu, {agail,treue}@gwdg.de`
[5] Bernstein Center for Computational Neuroscience,
Am Fassberg 17, 37077 Göttingen, Germany
[6] Max Planck Institute for Experimental Medicine,
Hermann Rein Strasse 3, 37075 Göttingen, Germany
[7] Campus Institute for Dynamics of Biological Networks,
Hermann Rein Strasse 3, 37075 Göttingen, Germany

Abstract. A Transparent game is a game-theoretic setting that takes action visibility into account. In each round, depending on the relative timing of their actions, players have a certain probability to see their partner's choice before making their own decision. This probability is determined by the level of transparency. At the two extremes, a game with zero transparency is equivalent to the classical simultaneous game, and a game with maximal transparency corresponds to a sequential game. Despite the prevalence of intermediate transparency in many everyday interactions such scenarios have not been sufficiently studied. Here we consider a transparent iterated Prisoner's dilemma (iPD) and use evolutionary simulations to investigate how and why the success of various strategies changes with the level of transparency. We demonstrate that non-zero transparency greatly reduces the set of successful memory-one strategies compared to the simultaneous iPD. For low and moderate transparency the classical "Win - Stay, Lose - Shift" (WSLS) strategy is the only evolutionary successful strategy. For high transparency all strategies are evolutionary unstable in the sense that they can be easily counteracted, and, finally, for maximal transparency a novel "Leader-Follower" strategy outperforms WSLS. Our results provide a partial explanation for the fact that the strategies proposed for the simultaneous iPD are rarely observed in nature, where high levels of transparency are common.

1   Introduction

Game theory is widely used to account for strategic decision-making in rational agents. Classical game theory assumes that players act either sequentially or simultaneously. Yet, for social behavior these assumptions are often not fulfilled since humans and animals rarely act strictly simultaneously or sequentially. Instead, social agents try to observe their partners and adversaries, and use the others' behavior to adjust their own actions accordingly [1,2]. Recently a new game-theoretic setting of "transparent games" has been introduced, taking into account action visibility and providing a more realistic model of interactions under time constraints [3]. In transparent games, each player has a certain probability to observe the partner's choice before deciding on its own action. This probability is determined by the action times of the players. For instance, if Player 1 always acts well before Player 2, the probability to see the partner's choice is zero for Player 1 and one for Player 2 (corresponds to strictly sequential playing). If both players act approximately at the same time, on average, they have equal probabilities p_{see} to see each other's choices. This probability can range from $p_{\text{see}} = 0$ (players cannot take choices of each other into account; this case corresponds to the classical simultaneous game) to $p_{\text{see}} = 0.5$ (in every round typically one of the two players sees the choice of the partner and can adjust its own decision). The probability that neither player sees the choice of the partner is equal to $1 - 2p_{\text{see}}$.

One may expect that action visibility would increase cooperation in non-zero-sum games, such as the iterated Prisoner's dilemma (iPD). However, evolutionary simulations in [3] show that this is not necessarily the case. Evolutionary agents successfully establish cooperation in the iPD for low and moderate transparency by using the classic "Win - stay, lose - shift" (WSLS) strategy [4]. However, for high transparency cooperation drastically decreases (see Fig. 1), and the most frequent strategy is "Leader-Follower", which does not rely on mutual cooperation [3].

This unexpected drop of cooperation in an iPD with high transparency raises new questions. What explains the success of WSLS for the transparent iPD? Why does the Leader-Follower strategy become more frequent for high transparency? Is its success evolutionary stable or is this strategy just transiently successful? To answer these questions, we use evolutionary simulations, since analytic considerations for the transparent games require solving differential equations with many variables. Since most results for simultaneous and sequential versions of the iPD were obtained for strategies taking into account outcomes of only the last interaction ("memory-one strategies"), we also focus on such strategies. We show that the rather complex strategy dynamics associated with $p_{\text{see}} = 0$ (simultaneous iPD) [5] is greatly simplified for the transparent settings with $p_{\text{see}} > 0$. WSLS is the only non-transient strategy for $0 < p_{\text{see}} < 0.5$.
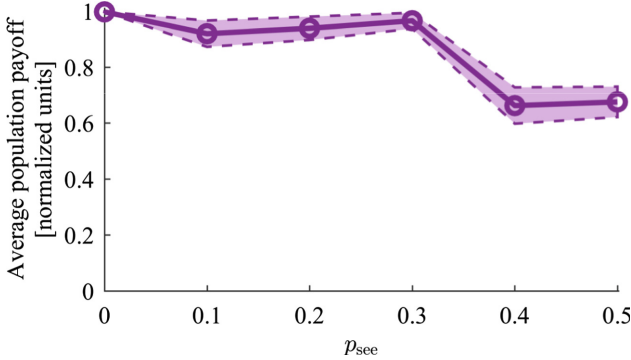
Fig. 1. Average fraction of maximal possible payoff (with 95% confidence interval) achieved by the evolutionary evolved population of players in the transparent iterated Prisoner's Dilemma. Payoff values are computed over 80 runs for probabilities $p_{see} = 0.0, 0.1, \ldots, 0.5$ of a player to see the choice of a partner. Maximal payoff corresponds to mutual cooperation. Stable cooperation was established for low and moderate transparencies, but it was disrupted for $p_{see} > 0.3$ resulting in a significant decrease of average payoff.

For $p_{see} > 0.35$ all memory-one strategies are evolutionary unstable and replace each other in rapid succession, though WSLS and Leader-Follower are more stable than the others.

Our results complement the findings in [3] by providing a detailed analysis of the evolutionarily successful strategies for the transparent iPD. In particular, we emphasize the superiority of WSLS over other memory-one strategies for most transparency levels. Yet, the fact that WSLS and other strategies developed for the simultaneous iPD are unstable for an iPD with high transparency may partially explain why these classic strategies are rarely observed in nature [6,7], where high levels of transparency are prevalent [1].

2   Methods

2.1   Transparent Iterated Prisoner's Dilemma

Prisoner's dilemma [8] is perhaps the most studied game between two players. We refer to [9] for a review. The payoff matrix for this game is presented in Fig. 2. Apparently, the best choice for both players is mutual cooperation, but defection dominates cooperation (i.e., it yields a higher payoff for each individual player regardless of the partner's choice). Thus any self-interested player would defect (D), although mutual defection results in lowest possible joint payoff. This "paradox" of Prisoner's dilemma raises a question: under what conditions is it rational to cooperate (C)?

Player 2 \ Player 1	Cooperate (C)	Defect (D)
Cooperate (C)	R=3 R=3	T=5 S=0
Defect (D)	S=0 T=5	P=1 P=1

Fig. 2. Payoff matrix for Prisoner's Dilemma. In this game two players adopt roles of prisoners suspected of committing a crime. Each can either betray the other (Defect), or Cooperate 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 example, if Player 1 defects and Player 2 cooperates, then Player 1 gets payoff $T = 5$ and goes free, while Player 2 gets $S = 0$ and has to spend five years in prison. If both prisoners defect, each gets a slightly reduced sentence of four years ($P = 1$) Finally, with mutual cooperation between the prisoners there is only circumstantial evidence left, which is sufficient to sentence both prisoners to two years in prison ($R = 3$).

One possible answer is provided by playing the game repeatedly: in the iPD cooperation between self-interested players becomes plausible, since they can take into account past outcomes but also have to consider that they will play against each other again in future. When experience shows the partner to be trustworthy, cooperation in repeated game becomes a viable option. Traditionally each round of the iPD is considered to be a simultaneous game [8], where players make choices independently. In the recently suggested transparent iPD [3], both players have a certain chance to learn the current choice of the partner. Three cases are possible:

1. Player 1 sees the choice of Player 2 before making its own choice – with probability p_{see}^1.
2. Player 2 sees the choice of Player 1 before making its own choice – with probability p_{see}^2.
3. Neither of players knows the choice of the partner, probability of this case is $1 - p_{\text{see}}^1 - p_{\text{see}}^2$.

Note that $p_{\text{see}}^1 + p_{\text{see}}^2 \leq 1$ meaning that in each round only one of the players can see the partner's choice. In the transparent iPD it is natural to assume that both players act on average at the same time [3]. Indeed, though both players are interested in the partner's choice and would wait for the partner's action, there is always a time constraint preventing players from waiting indefinitely. If this time constraint is explicit and known to both players, in most cases they are motivated to act just before the end of the time allowed for the action, thus - nearly at the same time. Therefore for the rest of the paper we assume that $p_{\text{see}}^1 = p_{\text{see}}^2 = p_{\text{see}} \leq 0.5$ and call p_{see} the transparency level.

As it is usually done for the simultaneous iPD [5], we assume that players take into account outcomes of the previous game round, that is, use "memory-one strategies". Then a strategy of a player in a transparent iPD is represented

by a vector $\mathbf{s} = (s_k)_{k=1}^{12}$, where s_k are conditional probabilities to cooperate in 12 different situations. These depend on whether the player and the partner cooperated in the previous round, whether the player can see the current choice of the partner, and what the choice is if it is visible, see Table 1.

Table 1. Representation of strategies in transparent games. Each strategy entry s_i specifies probability of a player to cooperate depending on the outcome of previous round (the first action specifies the choice of the player, and the second the choice of the partner) and whether current partner's choice is visible.

Outcome of previous round	CC	CD	DC	DD
Current partner's choice unseen	s_1	s_2	s_3	s_4
Partner is cooperating (C)	s_5	s_6	s_7	s_8
Partner is defecting (D)	s_9	s_{10}	s_{11}	s_{12}

To find out what strategies are optimal depending on the transparency level p_{see}, we use evolutionary simulations described in the following subsection.

2.2   Evolutionary Dynamics of Memory-One Strategies in Transparent Games

To study evolutionary dynamics of the transparent iPD, we used the techniques developed for the simultaneous iPD in [4,5] with minor adaptation to account for the different strategy representation in transparent games, see also [3].

Consider a population consisting of transparent-iPD players from n "species" defined by their strategies $\mathbf{s}_i = (s_k^i)_{k=1}^{12}$ for $i = 1, 2, \ldots, n$. That is, any species i is a group of players sharing the same strategy $\mathbf{s}_i$, which they use when playing the iPD with a given transparency level $p_{\text{see}} \in [0.0, 0.5]$ against any partner. We assume that the population is infinitely large since a finite number of players results in complex stochastic effects [10], which we do not consider here. The population evolves in generations $t = 1, 2, \ldots$. In each generation every player plays infinitely many rounds of the transparent iPD against a partner assigned randomly according to the current composition of the population. Species getting higher average payoff than others reproduce more effectively, and their relative frequency $x_i(t)$ in the population increases in the next generation. Note that $\sum_{i=1}^{n(t)} x_i(t) = 1$ for any generation t.

Specifically, the evolutionary success of species i is encoded by its fitness $f_i(t)$, computed as the average payoff for a player from this species when playing against the current population:

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

where E_{ij} is the expected payoff of species i playing against species j. If $f_i(t)$ is higher 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 equation [5]:

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

It remains to compute the expected payoff E_{ij} for a strategy $\mathbf{s}_i = (s_k^i)_{k=1}^{12}$ against $\mathbf{s}_j = (s_k^j)_{k=1}^{12}$. For this, consider Players 1 and 2 from species i and j, respectively. Since both players use memory-one strategies, their choices in every round of the game depend only on their mutual choices in the previous round. This allows to describe the game dynamics by a Markov chain with four states being the mutual choices of the two players (CC, CD, DC and DD), and transition matrix given by

$$M = (1 - 2p_{\text{see}})M_0 + p_{\text{see}}M_1 + p_{\text{see}}M_2, \tag{2}$$

with matrices M_0, M_1 and M_2 describing 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.

$$M_0 = \begin{pmatrix} s_1^i s_1^j & s_1^i(1-s_1^j) & (1-s_1^i)s_1^j & (1-s_1^i)(1-s_1^j) \\ s_2^i s_3^j & s_2^i(1-s_3^j) & (1-s_2^i)s_3^j & (1-s_2^i)(1-s_3^j) \\ s_3^i s_2^j & s_3^i(1-s_2^j) & (1-s_3^i)s_2^j & (1-s_3^i)(1-s_2^j) \\ s_4^i s_4^j & s_4^i(1-s_4^j) & (1-s_4^i)s_4^j & (1-s_4^i)(1-s_4^j) \end{pmatrix}, \tag{3}$$

$$M_1 = \begin{pmatrix} s_5^i s_1^j & s_9^i(1-s_1^j) & (1-s_5^i)s_1^j & (1-s_9^i)(1-s_1^j) \\ s_6^i s_3^j & s_{10}^i(1-s_3^j) & (1-s_6^i)s_3^j & (1-s_{10}^i)(1-s_3^j) \\ s_7^i s_2^j & s_{11}^i(1-s_2^j) & (1-s_7^i)s_2^j & (1-s_{11}^i)(1-s_2^j) \\ s_8^i s_4^j & s_{12}^i(1-s_4^j) & (1-s_8^i)s_4^j & (1-s_{12}^i)(1-s_4^j) \end{pmatrix}, \tag{4}$$

$$M_2 = \begin{pmatrix} s_1^i s_5^j & s_1^i(1-s_5^j) & (1-s_1^i)s_9^j & (1-s_1^i)(1-s_9^j) \\ s_2^i s_7^j & s_2^i(1-s_7^j) & (1-s_2^i)s_{11}^j & (1-s_2^i)(1-s_{11}^j) \\ s_3^i s_6^j & s_3^i(1-s_6^j) & (1-s_3^i)s_{10}^j & (1-s_3^i)(1-s_{10}^j) \\ s_4^i s_8^j & s_4^i(1-s_8^j) & (1-s_4^i)s_{12}^j & (1-s_4^i)(1-s_{12}^j) \end{pmatrix}. \tag{5}$$

Then we can represent the expected payoff by the following formula

$$E_{ij} = y_{\text{R}}R + y_{\text{S}}S + y_{\text{T}}T + y_{\text{P}}P, \tag{6}$$

where R, S, T, P are the entries of the payoff matrix ($R = 3, S = 0, T = 5, P = 1$ for the standard iPD, see Fig. 2), and $y_{\text{R}}, y_{\text{S}}, y_{\text{T}}, y_{\text{P}}$ 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 [5]:

$$(y_{\text{R}}, y_{\text{S}}, y_{\text{T}}, y_{\text{P}}) = (y_{\text{R}}, y_{\text{S}}, y_{\text{T}}, y_{\text{P}})M. \tag{7}$$

To guarantee the existence of the eigenvector (y_R, y_S, y_T, y_P), strategy entries for any species i should satisfy the following double inequality

$$0 < s_k^i < 1. \tag{8}$$

for all $k = 1, 2, \ldots, 12$. Therefore, it is common to introduce a minimal possible error ε in the strategies such that $\varepsilon \leq s_k^i \leq 1 - \varepsilon$. This also allows accounting for error-proneness of players (so-called "trembling hand" effect [4, 11]).

Equation (6) allows to compute the expected payoff E_{ij} for all strategies i, j in the population. The dynamics of the population is entirely described by the matrix E. This dynamics is relatively simple for two species, when only four cases are possible [10]:

– *Dominance*: one species is unconditionally more fit than the other and replaces it in the population. Strategy i dominates j when $E_{ii} > E_{ji}$ and $E_{ij} > E_{jj}$.
– *Bistability*: either species can take over the whole population depending on the initial relative frequencies $x_i(1)$, $x_j(1)$ and the threshold x^*, given by the (non-stable) equilibrium frequency of species i

$$x^* = \frac{E_{jj} - E_{ij}}{E_{ii} - E_{ij} - E_{ji} + E_{jj}}. \tag{9}$$

If $x_i(1) > x^*$, then species i takes over the population, otherwise species j wins. Bistability occurs when $E_{ii} > E_{ji}$ and $E_{ij} < E_{jj}$.
– *Coexistence*: the population is composed of two species with the asymptotic frequencies $x_i = x^*$, $x_j = 1 - x^*$, where x^* is a stable equilibrium given by (9). This case takes place when $E_{ii} < E_{ji}$ and $E_{ij} > E_{jj}$.
– Finally, *neutrality* takes place when $E_{ii} = E_{ji}$ and $E_{ij} = E_{jj}$, meaning that the frequencies of the species do not change.

We used the analysis of two-strategy dynamics for a pairwise comparison of strategies. However, already for a population consisting of $n = 3$ species, the dynamics can be rather complex, since there are 33 possible types of dynamics [12]. Analytic considerations for $n > 3$ are even more complicated [13] and are usually replaced by evolutionary simulations [5]. Details of our simulations are provided in Sect. 2.3.

2.3 Evolutionary Simulations

For the evolutionary simulations we adopted methods suggested in [4,5]. We studied the populations of transparent iPD players for various transparency levels p_{see}, and for each level we ran multiple evolutionary simulations as described below.

In each run, the population evolved as described in Sect. 2.2, thus the frequencies of species $x_i(t)$ changed according to (1). When $x_i(t) < \chi$, the species i was assumed to die out and was removed from the population; its fraction $x_i(t)$ was distributed among the remaining species proportional to their share in the population. We followed [4,5] in taking $\chi = 0.001$.

In each run a new species could enter the population with probability 0.01, thus new species emerged on average every 100 generations. Following [4], probabilities s_k^i for the new species were randomly drawn from the distribution with U-shaped probability density function (10), favoring probability values around 0 and 1:

$$\rho(y) = \pi\big(y(1-y)\big)^{-1/2} \tag{10}$$

for $y \in (0,1)$. We required $s_k^i \in [\varepsilon, 1-\varepsilon]$ with $\varepsilon = 0.001$ as suggested in [4,5] to satisfy inequality (8). Initial frequencies of the introduced species were set to 1.1χ with $\chi = 0.001$ [4].

The most important strategy entries, especially for low transparency, are s_1, s_2, s_3, s_4 (probability to cooperate when partner's choice is unknown), since players use one of them with probability $1 - p_{\text{see}} > 0.5$. Therefore these entries converge to optimal values quite fast and their values in evolutionary successful strategies are most precise, while convergence of other entries to optimal values may take longer. Since values of s_1, s_2, s_3, s_4 for evolutionary successful strategies were described in [3], here we limited their variability to allow faster convergence of the remaining entries $s_5, \dots, s_{12}$ to the optimal values. For this we rounded strategy entries s_1, s_2, s_3, s_4 so that they had the values from the set

$$s_1^i, s_2^i, s_3^i, s_4^i \in \{\varepsilon, \frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, 1-\varepsilon\}. \tag{11}$$

For the iPD with the payoff matrix shown in Fig. 2 commonly discussed strategies are formed by the values from the set $\{\varepsilon, \frac{1}{3}, \frac{2}{3}, 1-\varepsilon\}$ [4,5,14], but in 11 we also consider intermediate values to achieve better discretization.

We carried out evolutionary simulations with random and with pre-defined initial compositions of the population. A random initial population consisted of $n(1) = 5$ species with equal frequencies $x_1(1) = \dots = x_5(1) = 0.2$ and random strategies. In this case we traced 10^9 generations in each run. We also considered three pre-defined initial populations consisting of a single species with one of the following strategies:

– Win – stay, lose – shift (WSLS): $\mathbf{s} = (1, 0, 0, 1; 1, 0, 0, 1; 0, 0, 0, 0)$;
– Generous tit-for-tat (GTFT): $\mathbf{s} = (1, \frac{1}{3}, 1, \frac{1}{3}; 1, \frac{1}{3}, 1, \frac{1}{3}; 0, 0, 0, 0)$;
– Leader-Follower (L-F):

$$\mathbf{s} = (\frac{1}{3}, 0, 0, 0; \frac{2}{3}, 0, 0, 0; 1, 1, 1, \frac{1}{5}) \tag{12}$$

These strategies were selected as most evolutionary successful in simulations with random initial population for various transparency levels p_{see} (see Sect. 3 for details). For the pre-defined initial populations we traced 10^8 generations in each run since it was not necessary to ensure stabilization of the population dynamics (which was the case for the random initial population). For each of the four initial compositions of the population described above, we performed 80 runs for each value of $p_{\text{see}} = 0.0, 0.1, \dots, 0.5$. Additional 80 runs of simulations were performed for transparency levels $p_{\text{see}} = 0.26, 0.28, \dots, 0.50$, for the initial WSLS and L-F populations.

2.4 Measuring the Strategies' Evolutionary Success

To quantify the evolutionary success of a strategy we computed the average of its relative frequency $x_i(t)$ across all generations t and across all runs. Since the strategies in the evolutionary simulations were generated randomly, some of the observed strategies might not entirely converge to the theoretical optimum. Therefore, we used a coarse-grained description of strategies as suggested in [3] with the following notation: symbol 0 for $s_k^i \leq 0.1$, symbol 1 for $s_k^i \geq 0.9$, symbol * is used as a wildcard character to denote an arbitrary probability. We characterized as Always Defect (AllD) all strategies 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 behavior is not specified. Similarly, the generalized representations of other strategies were as follows

- Win – stay, lose – shift (WSLS): (100c; 1***; ****) with $c \geq 2/3$;
- Tit-for-tat (TFT): (1010; 1***; ****);
- Generous tit-for-tat (GTFT): (1a1c; 1***; ****), where $0.1 < a, c < 0.9$;
- Generous WSLS (GWSLS): (1abc; 1***; ****), where $c \geq 2/3$, $a, b < 2/3$ and either $a > 0.1$ or $c > 0.1$;
- Firm-but-fair (FbF): (101c; 1***; ****), where $0.1 < c < 0.9$;
- Leader-Follower (L-F): (*00c; ****; *11d), where $c < 1/3$ and $d < 2/3$.

3 Results

Frequencies of strategies for various initial compositions of the population are presented in Fig. 3. In all cases, the fraction of the described strategies drops down for $p_{\text{see}} > 0.3$. This happens due to the fact that for higher transparencies population dynamics often enters a "chaotic" mode, where many transient strategies replace each other in a rapid succession [3]. Each of these transient strategies has a low relative frequency, while taken together they constitute a considerable fraction of the population, especially for high p_{see}.

Regardless of the initial population composition, "Win-stay, lose-shift" (WSLS) strategy is a clear winner for $0 < p_{\text{see}} < 0.4$. The theoretically optimal form of this strategy is (1001; 1001; 0000), however in simulations it may appear as (1001; 1**1; 000*), where the entries marked by the wildcard character * can take a value from 0 to 1 (see Fig. 5).

For $p_{\text{see}} \geq 0.1$ WSLS can be only replaced in the population by a strategy (1001; d***; *00*) with $0 \leq d < 1$ (Fig. 6a–c). In [3] this strategy was called treacherous WSLS since it behaves like WSLS when not seeing the choice of the partner, and defects when seeing that partner cooperates. Treacherous WSLS has low payoff when playing with itself and is easily replaced by other strategies, but it dominates WSLS for any $p_{\text{see}} > 0$.

The predominance of WSLS as the single frequent non-transient strategy can be challenged only for $p_{\text{see}} \approx 0$ (minimal transparency, nearly-simultaneous iPD) by GTFT and for $p_{\text{see}} > 0.4$ (maximal transparency, nearly-sequential iPD) by L-F. Below we discuss both these cases in detail.
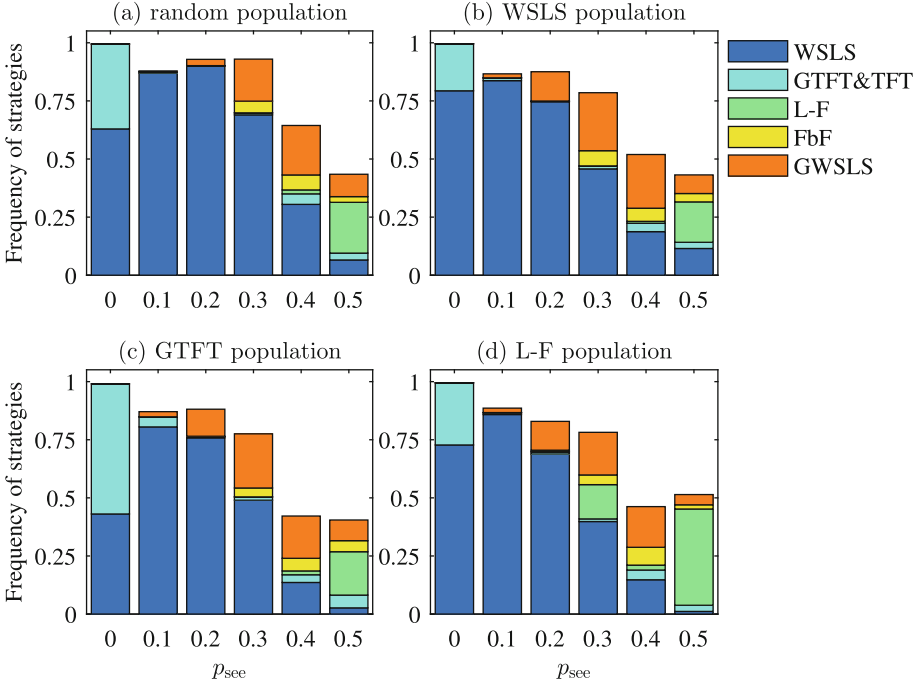
Fig. 3. Average relative frequencies of strategies occurred in the population for its various initial compositions: (a) random; constituted by (b) WSLS, (c) GTFT and (d) L-F strategies. Other strategies are transient (i.e. persist in the population only for a relatively short time). Results for all initial populations are quite similar, yet a few differences are noteworthy. First, the frequencies of WSLS and GTFT depend on the initial composition of the population for $p_{see} = 0$, but not for higher transparency. When one of these strategies was the initial strategy of the population, it is much more abundant than for the random initial composition. Second, when L-F is the initial strategy of the population, it is has considerably higher frequency for $p_{see} = 0.3$ and $p_{see} = 0.5$, but not for $p_{see} = 0.4$. Since for $p_{see} = 0.3$ L-F has a noticeable frequency only when forming the initial population, for this transparency L-F can persist but cannot take over the population when starting with a low frequency (see Fig. 4b). Note that such non-monotonicity in strategy dynamics is not altogether unexpected, since the dynamics depends non-linearly on the transparency level [3].

For minimal transparency the WSLS frequency slightly drops (Fig. 3a) since in this case population of WSLS players can be invaded by AllD. Let $\mathbf{s}_1 = (1 - \varepsilon, \varepsilon, \varepsilon, 1 - \varepsilon; 1 - \varepsilon, \varepsilon, \varepsilon, 1 - \varepsilon; \varepsilon, \varepsilon, \varepsilon, \varepsilon)$ (WSLS), $\mathbf{s}_2 = (\varepsilon, \varepsilon, \varepsilon, \varepsilon; \varepsilon, \varepsilon, \varepsilon, \varepsilon; \varepsilon, \varepsilon, \varepsilon, \varepsilon)$ (AllD). Estimating matrix of expected payoff by (6), we see that

$$\mathrm{E}_{11} \approx 3 - 5\varepsilon, \ \mathrm{E}_{12} \approx \frac{1 + 7\varepsilon}{2} + \frac{1 - 7\varepsilon}{2}p_{see}, \ \mathrm{E}_{21} \approx 3 - \frac{3}{2}\varepsilon - p_{see}, \ \mathrm{E}_{22} \approx 1 + 3\varepsilon.$$

Thus for $p_{see} \leq 3.5\varepsilon$ WSLS is dominated by AllD and for $p_{see} > 3.5\varepsilon$ the two strategies are bistable. AllD invades WSLS in the former case unconditionally
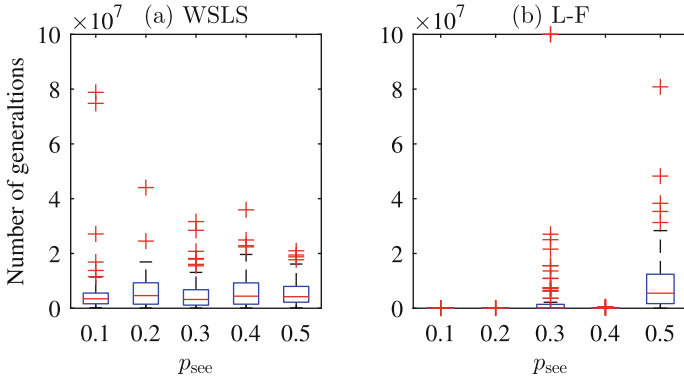
Fig. 4. Number of generations for that initial strategy of the population remains the most frequent for the initial population constituted by (a) WSLS and (b) L-F players. The central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol. The higher the number of generations, the longer the initial strategy persists in the population. While persistence of WSLS is relatively stable, L-F can persist only for $p_{see} = 0.3$ and $p_{see} = 0.5$.

and in the latter case if its equilibrium frequency given by (9) is sufficiently low. Namely, if AllD with initial fraction 1.1χ is introduced to the population playing WSLS, for $p_{see} < \frac{7\varepsilon + 1.1\chi(1-8\varepsilon)}{2-1.1\chi}$ AllD invades the population. In this case, unconditional defectors can take over the population for a considerable number of generations, and strategies different from WSLS (TFT and GTFT) should be introduced to the population to restore cooperation. These strategies can resist WSLS invasion, therefore once WSLS dies out it may take many generations to re-establish itself.

In our simulations, $\varepsilon = 0.001$, $\chi = 0.001$, so WSLS can resist AllD-invasion for $p_{see} > 0.0041$. In this case WSLS is dominated only by treacherous WSLS. This weak transient strategy is easily replaced by others, which allows WSLS to reappear in the population relatively quickly.

The decrease of WSLS frequency for high transparency (Figs. 3 and 7) is caused by two factors. First, as one can see from Fig. 5, the higher p_{see} the more precisely WSLS should correspond to its theoretically optimal profile. Indeed, for $p_{see} = 0.1$ most entries of WSLS have relatively high standard deviations, meaning that in this case a successful strategy can follow the WSLS principle in a rather general fashion. Meanwhile, for $p_{see} = 0.5$ most entries have low variability, meaning that successful WSLS variants can have only slight deviations from the optimal profile. The closer a strategy must be to the theoretical WSLS profile in order to be successful, the lower the probability that such a strategy is by chance introduced in the population, which results in lower frequency of WSLS for higher p_{see}. Note that the frequency of WSLS would decrease even if we understand WSLS in the most general sense and consider GWSLS as a part
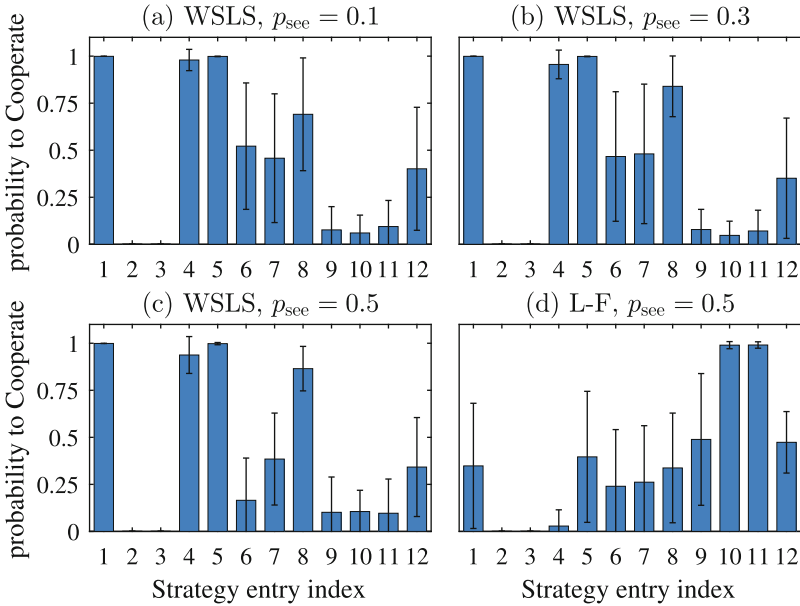
Fig. 5. Average strategy profiles with standard deviations for the winning strategies in a transparent iPD: (a) for WSLS at $p_{\mathrm{see}} = 0.1$, (b) $p_{\mathrm{see}} = 0.3$, (c) $p_{\mathrm{see}} = 0.5$ and for L-F at $p_{\mathrm{see}} = 0.5$ (d). Each strategy profile is represented by 12 entries characterizing probability of a player to cooperate. Namely, $s_1, \ldots, s_4$ are probabilities to cooperate when current partner's choice is unknown and the outcome of the previous round was "both cooperated", "self cooperated, partner defected", "self defected, partner cooperated" and "both defected", respectively. Similarly, $s_5, \ldots, s_8$ and $s_9, \ldots, s_{12}$ are probabilities of a player to cooperate seeing that current partner's choice is to cooperate and to defect, respectively. Standard deviations show the difference in variability for various entries. For instance in (a) some strategy entries are constant s_1, s_2, s_3, s_5, some vary only slightly s_4, s_9, s_{10}, s_{11}, and other entries are almost random s_6, s_7, s_8, s_{12}.

of it (Fig. 8). Second, the higher p_{see} is, the faster and easier the treacherous WSLS takes over the population, since it has higher chances to defect WSLS. This also means that a strategy that only roughly resembles the treacherous WSLS profile can be successful against WSLS for high p_{see} (see Fig. 6a–c).

It remains to explain the success of L-F for high transparency. Consider the effect of p_{see} on the frequencies of WSLS and L-F in more detail. Figure 8 shows how these frequencies vary for $p_{\mathrm{see}} \in [0.26, 0.50]$ taken with a step of 0.02. for the initial population consisting either of WSLS or L-F players. Note that L-F is only successful for maximal transparency $p_{\mathrm{see}} = 0.5$ (Fig. 8a), although for some transparency levels $p_{\mathrm{see}} < 0.5$, L-F can remain in the population once introduced (Fig. 8b).
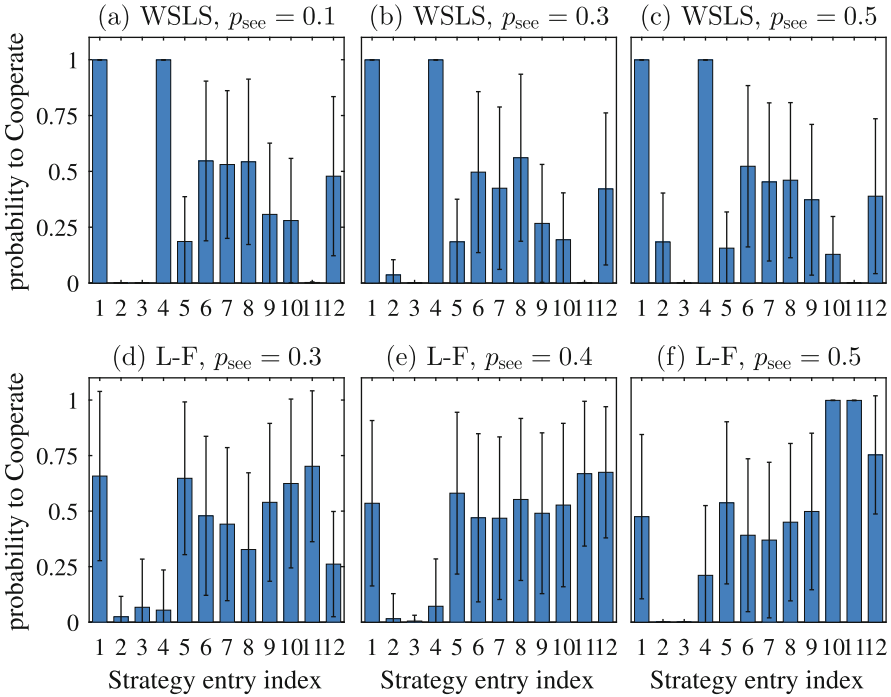
Fig. 6. Average strategy profiles with standard deviations for the strategies replacing initial strategies in simulations with pre-defined initial population: (a)–(c) strategies replacing WSLS for $p_{see} = 0.1$, $p_{see} = 0.3$ and $p_{see} = 0.5$, respectively; (d)–(f) strategies replacing L-F for $p_{see} = 0.3$, $p_{see} = 0.4$ and $p_{see} = 0.5$, respectively.

The success of L-F for maximal transparency can be partially explained by the fact that a variant of this strategy given by Eq. (12) can be replaced in the population only by other variants of the L-F strategy (Fig. 6d). Although these variants are not evolutionary stable and are eventually replaced by other strategies (Fig. 3d), this indicates a good evolutionary potential of L-F for $p_{see} = 0.5$. One particularly interesting modification of L-F is a strategy with a profile (1001; 1001; 1111). As it combines the features of L-F and WSLS we term it "WSLS-like L-F". This strategy dominates normal L-F and easily replaces it in the population. But contrary to the normal L-F, WSLS-like L-F is highly unstable since it is dominated by many strategies including WSLS, treacherous WSLS and AllD. Therefore, WSLS-like L-F stays in population just for a few generations and then is replaced by other strategies (similar to the treacherous WSLS).
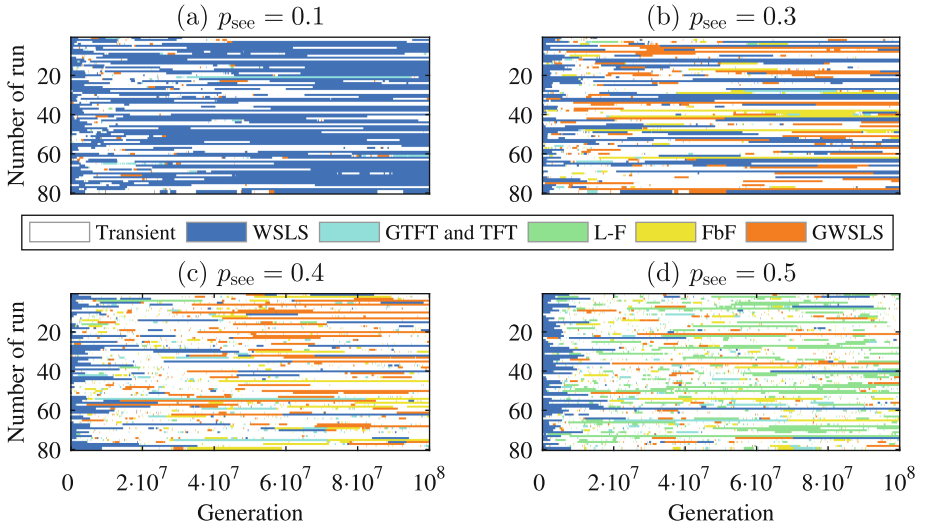
Fig. 7. Dynamics of the most frequent strategy of the population in various runs for initial WSLS-population. (a) For $p_{see} = 0.1$ WSLS is predominant in most runs. (b) For $p_{see} = 0.3$ WSLS frequency wanes, while other TFT, L-F, FbF and GWSLS become more abundant. Fraction of various transient strategies also increases and (c) for $p_{see} = 0.4$ they become most frequent. (d) For $p_{see} = 0.5$ WSLS rarely reestablishes itself and the only frequent strategy is L-F. Population mostly is in a chaotic mode with transient strategies replacing each other.
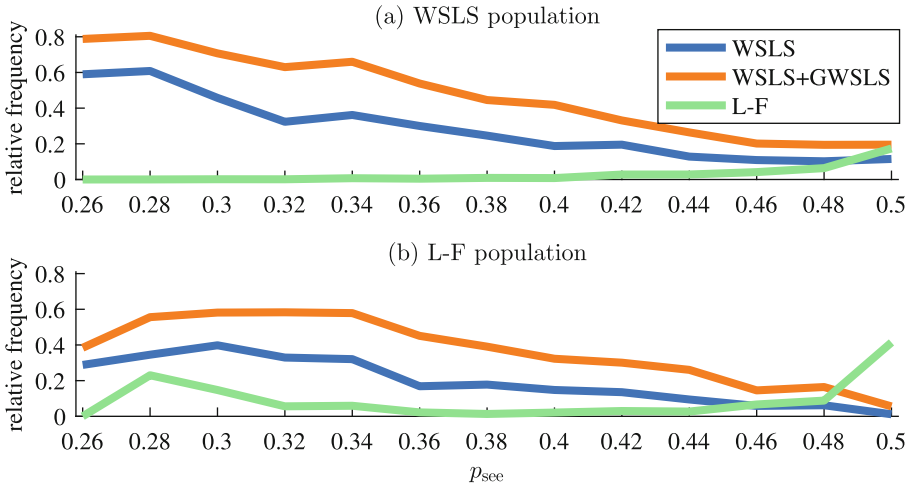

Fig. 8. Average relative frequencies of WSLS, WSLS with GWSLS, and L-F strategies for the initial populations constituted by (a) WSLS player and (b) L-F players. Note that in both cases WSLS frequency decreases considerably for $p_{see} > 0.35$.

4   Conclusion

Here we aimed to explain the success of various strategies in the transparent iterated Prisoner's dilemma [3]. Our main findings are:

– Win – stay, lose – shift (WSLS) and Generous tit-for-tat (GTFT) are two predominant strategies for $p_{see} \approx 0$. This case of low transparency is very close to the classical simultaneous Prisoner's dilemma in which agents make choices without knowing the respective choice of the other agent.
– For most non-zero transparency levels ($0 < p_{see} < 0.5$), WSLS is the only effective memory-one strategy. However, WSLS is not evolutionary stable, since it can be counteracted by its modified version, termed "treacherous WSLS", which defects when seeing the partner's choice.
– For $p_{see} \approx 0.5$ (maximal transparency) a second strategy, Leader-Follower (L-F), becomes predominant. When two players use this strategy, one of them defects and the other cooperates, leading to a kind of "turn-taking" behavior. However, L-F is also not evolutionary stable since it evolves towards highly unstable WSLS-like variants.
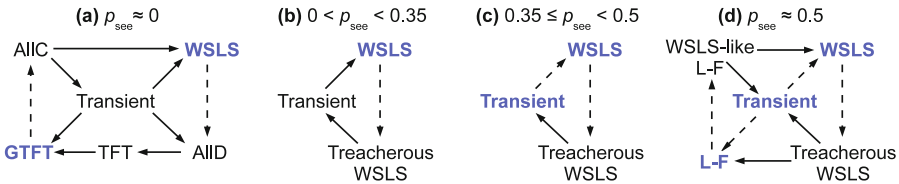

Fig. 9. Strategy dynamics for various values of p_{see}. Solid and dashed arrows indicate likely and less likely directions of dynamics, respectively. Strategies that can persist in the population for high number of generations (compared to other strategies) are shown in bold blue font. We group all strategies that cannot persist in the population and do not have special importance for the dynamics as "transient". Four cases of dynamics can be distinguished. (a) For $p_{see} \approx 0$ (for $p_{see} < 0.004$ in our simulations) the game resembles the classic simultaneous iPD. In this case strategy dynamics is relatively complex [5]. The population oscillates between two predominant strategies, WSLS and GTFT, and the transient states are relatively short. (b) As we increase p_{see}, two important changes take place. First, GTFT becomes ineffective against other strategies, including WSLS, L-F and AllC. Second, for $p_{see} > 0.004$ AllD cannot invade a population of WSLS players, and the only strategy that can "dethrone" WSLS is treacherous WSLS. These changes result in greatly simplified dynamics with WSLS being the only non-transient strategy. It is occasionally invaded by treacherous WSLS, which controls the population only for a short time and is then replaced by other strategies. (c) For high transparency ($0.35 < p_{see} < 0.5$, with threshold value 0.35 selected based on dynamics of strategy frequencies shown in Fig. 8), WSLS is still predominant, but it becomes difficult for this strategy to take over the population. Therefore, the population spends most of the time in a "chaotic" state with various transient strategies quickly replacing each other. (d) Finally, for maximal transparency ($p_{see} \approx 0.5$), L-F becomes the second predominant strategy. Yet, L-F does not have stable control of the population since some modifications of L-F can be invaded by non-L-F strategies. As a result, transient strategies still control the population for most of the time. (Color figure online)

Figure 9 summarizes our results schematically representing the strategy dynamics for various transparency levels p_{see}.

Overall our results provide an important extension to previous studies that have used evolutionary simulations to suggest and test strategies in classic simultaneous and sequential games. By concentrating on transparent games frequently encountered in real life but rarely investigated, we provide a partial explanation for the fact that the strategies developed for the simultaneous iPD are rarely observed in nature [6, 7], where high levels of transparency are common [1].

References

1. Dugatkin, L., Mesterton-Gibbonsand, M., Houston, A.: Beyond the Prisoner's dilemma: toward models to discriminate among mechanisms of cooperation in nature. Trends Ecol. Evol. **7**(6), 202–205 (1992)
2. Vaziri-Pashkam, M., Cormiea, S., Nakayama, K.: Predicting actions from subtle preparatory movements. Cognition **168**, 65–75 (2017)
3. Unakafov, A.M., Schultze, T., Kagan, I., Moeller, S., Eule, S., Wolf, F.: Emergence and suppression of cooperation by action visibility in transparent games (2018). https://doi.org/10.1101/314500
4. Nowak, M., Sigmund, K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's dilemma game. Nature **364**(6432), 56–58 (1993)
5. Nowak, M.: Evolutionary Dynamics. Harvard University Press, Cambridge (2006)
6. 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. Anim. Behav. **39**(1), 78–90 (1990)
7. Noë, R.: Cooperation experiments: coordination through communication versus acting apart together. Anim. Behav. **71**(1), 1–18 (2006)
8. Axelrod, R., Hamilton, W.: The evolution of cooperation. Science **211**, 13–90 (1981)
9. Doebeli, M., Hauert, C.: Models of cooperation based on the Prisoner's dilemma and the snowdrift game. Ecol. Lett. **8**(7), 748–766 (2005)
10. Traulsen, A., Hauert, C.: Stochastic evolutionary game dynamics. Rev. Nonlinear Dyn. Complexity **2**, 25–61 (2009)
11. Selten, R.: Reexamination of the perfectness concept for equilibrium points in extensive games. Int. J. Game Theor. **4**(1), 25–55 (1975)
12. Hofbauer, J., Sigmund, K.: Evolutionary game dynamics. Bull. Am. Math. Soc. **40**(4), 479–519 (2003)
13. Zeeman, E.C.: Population dynamics from game theory. In: Nitecki, Z., Robinson, C. (eds.) Global Theory of Dynamical Systems. LNM, vol. 819, pp. 471–497. Springer, Heidelberg (1980). https://doi.org/10.1007/BFb0087009
14. Nowak, M., Sigmund, K.: The alternating Prisoner's dilemma. J. Theor. Biol. **168**(2), 219–226 (1994)