Article

A Soluble Model for the Conflict between Lying and Truth-Telling

Eduardo V. M. Vieira [ID] **and José F. Fontanari*** [ID]

Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, São Carlos 13560-970, SP, Brazil; eduardov@usp.br
* Correspondence: fontanari@ifsc.usp.br

Abstract: Lying and truth-telling are conflicting behavioral strategies that pervade much of the lives of social animals and, as such, have always been topics of interest to both biology and philosophy. This age-old conflict is linked to one of the most serious threats facing society today, viz., the collapse of trustworthy sources of information. Here, we revisit this problem in the context of the two-choice sender–receiver game: the sender tosses a coin and reports the supposed outcome to the receiver, who must guess the true outcome of the toss. For the sender, the options are to lie or tell the truth, while for the receiver, the options are to believe or disbelieve the sender's account. We assume that social learning determines the strategy used by players and, in particular, that players tend to imitate successful individuals and thus change their strategies. Using the replicator equation formulation for infinite populations and stochastic simulations for finite populations, we find that when the sender benefits from the receiver's failure, the outcome of the game dynamics depends strongly on the choice of initial strategies. This sensitivity to the initial conditions may reflect the unpredictability of social systems whose members have antagonistic interests.

Keywords: replicator equation; game theory; evolution of lying; stochastic simulation

MSC: 91A05; 91A10; 91A28

1. Introduction

The conflict between lying and telling the truth is a topic of widespread interest that has laid the foundation for well-established fields of research in both biology and philosophy. In fact, a major theme of signaling theory, which is devoted to the study of communication between animals, is the evolution of fitness signals and their selection for "honesty" [1,2]. In philosophy, this conflict appears in ethics and was made famous by Kant's discussion of keeping promises [3], which concluded in the impossibility of a world in which everyone lies, and for the possibility of a world in which everyone tells the truth (see also [4,5]). The challenge to policymakers posed by today's epidemics of disinformation (i.e., misinformation with the explicit intent to mislead [6]) and its threat to epistemic security or knowledge safety [7–9] can also be traced back to the age-old conflict between lying and truth-telling. Indeed, without shared interests and beliefs, it will be very difficult to deal effectively with issues that threaten our world, such as climate change, a failure that will likely make our world literally impossible, not just in a philosophical sense. Game theory models can help guide and predict climate change negotiations and aid in determining the incentives for truth-telling and cooperation [10–13].

Not surprisingly, behavioral economists have proposed several theories and conducted several experiments to assess people's honesty [14]. In this paper, we consider the two-player sender–receiver game [15], where the players have two roles—sender and receiver—and each role has two strategies. For the sender, the options are to lie or tell the truth, while for the receiver, the options are to believe or disbelieve the sender's account. In total, there

are four strategies for each player. In a classic version of this game [16] (see [17,18] for a quantitative analysis), the sender rolls a die with six faces and informs the receiver about the result of the dice roll. At this point, the sender has the option of telling the truth or lying in her account. The receiver's goal is to guess the true result, so they have the option of believing or disbelieving the sender's report. Each player is informed of both her own and her opponent's possible payoffs if the receiver fails or succeeds in guessing the outcome of the dice roll. This introduces an emotional dimension to the game, as the sender may be selfish in the sense of maximizing her own payoff, but sensitive to the cost her lie imposes on the other side [15]. Unfortunately, this subjective dimension, which contributes to the player's reputation, is not easily incorporated into the mathematical formulation of the sender–receiver game (see, e.g., [19]).

Here, we analytically study a simplified version of the sender–receiver game in which the sender flips a coin instead of rolling a die, so that the receiver's success depends only on her ability to detect whether the sender is lying or not. We find that when the sender's and receiver's interests are opposed (i.e., when the sender benefits from the receiver's failure), the asymptotic solutions of the replicator equation [20] depend strongly on the choice of initial strategies. This dependence requires that we solve the full game dynamics to obtain the asymptotic solutions, which makes its study somewhat challenging. Nevertheless, we present analytical expressions for the phase space of the game and for the period of the periodic solutions. The sensitivity to initial conditions, and thus to perturbations in the dynamics, may reflect the unpredictability of social systems whose members have antagonistic interests. In this sense, we recall that the replicator equation describes the infinite population limit of the stochastic imitation game, where a player has a non-zero probability of switching to a strategy of a better-off opponent [21]. Hence, our results are relevant for the study of social systems [22,23].

The remainder of this paper is organized as follows. In Section 2, we describe the symmetric two-choice sender–receiver game and write the replicator equation for the frequencies of the four strategies in a well-mixed infinite population of players. In Section 3, we briefly describe the fixed-point solutions to the replicator equation. In Section 4, we study the periodic solutions of the replicator equation for a particular setting of parameters defining the deception game, where the sender is rewarded only if they deceive the receiver. As always, the receiver is rewarded if they guess the true outcome of the coin toss. We obtain analytical expressions for the phase-space trajectories and the period of the oscillations. In Section 5, we present finite population simulations of the stochastic imitation game, whose infinite population limit is described by the replicator equation formulation. In Section 6, we review our main results and discuss the trajectories in the phase space of truth tellers and believers. Finally, in Section 7 we present some concluding remarks.

2. The Two-Choice Sender–Receiver Game

The sender–receiver game is a game with two roles and two strategies for each role. The game is symmetric because a fair coin is tossed each round to decide which role—sender or receiver—is assigned to which player. The game is as follows. The sender tosses a coin and observes the result, which is then reported to the receiver. The twist here is that the sender has the option of lying (L) or telling the truth (T) about the outcome of the coin toss, and the receiver has the option of believing (B) or disbelieving (D) the sender's report. The payoff for both players depends on whether the receiver succeeds in guessing the correct outcome of the coin toss and on the intention of the sender. Table 1 shows the receiver's payoff matrix: they win $b_1 > 0$ if they succeed and lose $c_1 > 0$ if they fail to guess the true outcome of the coin toss, regardless of the sender's intentions. We note that the receiver can have success through the belief of a true report or the disbelief of a false report.

Table 1. Payoff matrix for the receiver player using the believe or disbelieve strategies to respond to the sender's report.

Strategy/Report	True	False
Believe	b_1	$-c_1$
Disbelieve	$-c_1$	b_1

The sender's payoff depends on the receiver's strategy, as shown in Table 2, where b_2 and c_2 are not necessarily positive. The sender's payoff is more complex than the receiver's, because while for the latter the question is whether the report received is true or not, for the former the question is whether the sender's intention was realized or not. Suppose the sender's intention is to deceive the receiver, so $b_2 > 0$ and $c_2 > 0$. Then, the sender succeeds whenever the receiver fails to guess the true outcome of the coin toss. This happens when the sender lies and the receiver believes the lie, or when the sender tells the truth and the receiver does not believe the truth. In both cases, the sender's intention (i.e., to have the receiver misjudge the coin toss) is realized, and so it is natural that the sender's reward is the same (i.e., b_2) in both cases. Of course, one may wonder why the sender would tell the truth if her intention was to deceive the receiver. This can be justified if the sender suspects that the receiver will not believe her report, so her intention is more likely to be fulfilled if she tells the truth. A similar reasoning follows if the sender's intention is to help the receiver guess the true outcome of the coin toss, so $b_2 < 0$ and $c_2 < 0$.

Table 2. Payoff matrix for the sender player using the lie or tell the truth strategies to inform the receiver about the coin toss outcome.

Strategy/Receiver	Believe	Disbelieve
Lie	b_2	$-c_2$
Tell the truth	$-c_2$	b_2

The two-choice sender–receiver game is a simplified version of the sender–receiver game introduced by Erat and Gneezy [16] (see also [17,18]), in which the sender rolls a six-faced die and reports the result to the receiver. The main difference with our scheme is that the receiver is not guaranteed to guess the correct outcome of the dice by disbelieving the sender if her report is false. Interestingly, an asymmetric version of the two-choice sender–receiver game describes a Batesian mimicry scenario involving the poisonous monarch butterflies, the non-poisonous viceroy butterflies that mimic the monarch butterflies, and the blue jay predators [4]: the butterflies always play the sender role, while the blue jays always play the receiver role (the sender and receiver payoffs must be properly chosen to describe this ecological scenario).

There are four possible strategies for each player, namely, (T, B), (T, D), (L, B), and (L, D), which we will refer to as strategies 1, 2, 3, and 4, respectively. The first component of the ordered pair refers to the strategy in the sender role: telling the truth (T) or lying (L). The second component refers to the strategy in the receiver role: believing (B) or disbelieving (D) the sender's report. Thus, the payoff for a player using strategy i against a player using strategy j is given, up to a factor $1/2$ which we will omit in this paper, by the entry M_{ij} of the matrix (see, e.g., [24])

$$M = \begin{bmatrix} b_1 - c_2 & b_1 + b_2 & -c_1 - c_2 & b_2 - c_1 \\ -c_1 - c_2 & b_2 - c_1 & b_1 - c_2 & b_1 + b_2 \\ b_1 + b_2 & b_1 - c_2 & b_2 - c_1 & -c_1 - c_2 \\ b_2 - c_1 & -c_1 - c_2 & b_1 + b_2 & b_1 - c_2 \end{bmatrix}. \tag{1}$$

In the traditional evolutionary game theory approach [25], the payoffs of players using a particular strategy determine the number of offspring they produce, so that the proportion of players using different strategies varies from generation to generation. Instead, we

assume that social learning determines the strategy used by players. In particular, we assume that players tend to imitate successful individuals and thus change their strategies. Explicitly, a randomly chosen player compares her average payoff with that of another randomly chosen player and adopts the strategy of the other player with a probability proportional to the payoff difference if it is positive. Otherwise, she keeps her strategy. In a remarkable contribution, Traulsen et al. [21] have shown that, for an infinite well-mixed population, the frequencies Π_i of the different strategies obey the replicator equation [20]

$$\dot{\Pi}_i = \Pi_i \left(\sum_j M_{ij} \Pi_j - \bar{M} \right), \tag{2}$$

for $i = 1, \dots, 4$ with

$$\sum_{i=1}^{4} \Pi_i = 1. \tag{3}$$

Here, the average payoff in the population $\bar{M} = \sum_i \Pi_i F_i$, where

$$F_i = \sum_{j=1}^{4} M_{ij} \Pi_j \tag{4}$$

is the expected payoff for strategy i, ensures that $\sum_{i=1}^{4} \dot{\Pi}_i = 0$. Using Equation (1), we can write explicitly the expected payoffs for the four strategies,

$$
\begin{aligned}
F_1 &= (b_1 - c_2)\Pi_1 + (b_2 + b_1)\Pi_2 - (c_1 + c_2)\Pi_3 + (b_2 - c_1)\Pi_4 & (5) \\
F_2 &= -(c_1 + c_2)\Pi_1 + (b_2 - c_1)\Pi_2 + (b_1 - c_2)\Pi_3 + (b_1 + b_2)\Pi_4 & (6) \\
F_3 &= (b_1 + b_2)\Pi_1 + (b_1 - c_2)\Pi_2 + (b_2 - c_1)\Pi_3 - (c_1 + c_2)\Pi_4 & (7) \\
F_4 &= (b_2 - c_1)\Pi_1 - (c_1 + c_2)\Pi_2 + (b_2 + b_1)\Pi_3 + (b_1 - c_2)\Pi_4, & (8)
\end{aligned}
$$

from where we obtain an explicit expression for the average payoff in the population,

$$\bar{M} = 2[b_1 + c_1 - b_2 - c_2] \left[\Pi_2 \Pi_3 - \Pi_1 \Pi_4 - \frac{1}{2}(\Pi_2 + \Pi_3) \right] + b_1 - c_2. \tag{9}$$

Next, we briefly discuss the fixed-point solutions of Equation (2) for the sake of completeness only, since our focus is on the parameter settings for which the solutions exhibit oscillatory behavior.

3. The Fixed-Point Solutions

The fixed-point solutions Π_i^* of Equation (2) are obtained by setting $\dot{\Pi}_i = 0$ for $i = 1, \dots, 4$. These conditions are satisfied if either $\Pi_i = 0$ or $F_i = \bar{M}$. The local stability of these solutions is determined by the linearization of Equation (2) in the neighborhood of the fixed points when one of the frequencies is eliminated using the constraint (3), so that the partial derivatives $\partial(F_i - \bar{M})/\partial\Pi_j$ can be defined [20]. Since these calculations are rather straightforward, we will only summarize the relevant results here.

- The two fixed points $\Pi_1^* = 1$, $\Pi_{i \neq 1}^* = 0$ and $\Pi_4^* = 1$, $\Pi_{i \neq 4}^* = 0$ are locally stable provided that $b_2 + c_2 < 0$, i.e., in the case where the sender benefits from the receiver correctly guessing the outcome of the coin toss.
- The fixed points $\Pi_2^* = 1$, $\Pi_{i \neq 2}^* = 0$ and $\Pi_3^* = 1$, $\Pi_{i \neq 3}^* = 0$ are always unstable.
- If $b_2 + c_2 \neq b_1 + c_1$, there exists a fixed point corresponding to the coexistence of two opposite strategies, namely, $\Pi_1^* = \Pi_4^* = 1/2$, $\Pi_2^* = \Pi_3^* = 0$ and $\Pi_2^* = \Pi_3^* = 1/2$, $\Pi_1^* = \Pi_4^* = 0$, but they are unstable.
- If $b_2 + c_2 = b_1 + c_1$, the fixed points corresponding to the coexistence of two opposite strategies, namely, $\Pi_4^* = 1 - \Pi_1^* < 1$, $\Pi_2^* = \Pi_3^* = 0$ and $\Pi_3^* = 1 - \Pi_2^* < 1$, $\Pi_1^* =$

$\Pi_4^* = 0$, are neutral with the values of Π_1^* and Π_2^* determined by the initial frequencies of the strategies.

- There is no fixed point corresponding to the coexistence of the pairs of non-opposing strategies (e.g., $\Pi_1^* \neq 0$, $\Pi_2^* \neq 0$, $\Pi_3^* = \Pi_4^* = 0$).
- There is no fixed point corresponding to the coexistence of any three strategies (e.g., $\Pi_{i \neq 4}^* \neq 0$, $\Pi_4^* = 0$).
- Finally, if $b_2 + c_2 > 0$ and $b_2 + c_2 \neq b_1 + c_1$, there is a fixed point corresponding to the coexistence of all four strategies, namely, $\Pi_2^* = \Pi_3^* = 1/2 - \Pi_1^*$ and $\Pi_4^* = \Pi_1^*$, where Π_1^* is determined by the frequency values of the strategies at $t = 0$. Because of this strong dependence on the initial population, the coexistence fixed point is neutral, i.e., any perturbation to it will lead to a different fixed point. We will return to this scenario when we discuss the stochastic imitation dynamics in Section 5.

In the following, we consider the periodic solutions of Equation (2).

4. The Deceiving Game

Here, we consider a particular setting of the parameters of the payoff matrix M that allows an analytical solution of Equation (2) in a non-trivial scenario. More explicitly, we set $b_1 + c_1 = b_2 + c_2$ so that the average payoff in the population, given by Equation (9), becomes $\bar{M} = b_1 - c_2$, and thus does not depend on the frequencies of the players' strategies. If all four strategies are present in the initial population, all fixed points are unstable or neutral. An interesting realization of this scenario is the deceiving game where $b_1 = b_2$ and $c_1 = c_2$. In this scenario, the reward occurs when the player guesses the true outcome of the coin toss or prevents the other player from guessing. The cost occurs when the player guesses the toss wrong or the other player guesses the toss right. The following calculations apply to the general scenario $b_1 + c_1 = b_2 + c_2$, but since the results do not depend on the particular choices of b_1, b_2, c_1, and c_2, except for a trivial rescaling of time t, we will refer to this general scenario as the deceiving game.

Accordingly, Equation (2) is restated as

$$\frac{d \ln \Pi_1}{d\tau} = \Pi_2 - \Pi_3 \tag{10}$$

$$\frac{d \ln \Pi_2}{d\tau} = \Pi_4 - \Pi_1 \tag{11}$$

$$\frac{d \ln \Pi_3}{d\tau} = \Pi_1 - \Pi_4 \tag{12}$$

$$\frac{d \ln \Pi_4}{d\tau} = \Pi_3 - \Pi_2 \tag{13}$$

where $\tau = (b_1 + c_1)t$. Adding the Equations (11) and (12) results in

$$\Pi_2(\tau)\Pi_3(\tau) = \Pi_2(0)\Pi_3(0) \equiv A, \tag{14}$$

while adding the Equations (10) and (13) yields

$$\Pi_1(\tau)\Pi_4(\tau) = \Pi_1(0)\Pi_4(0) \equiv B. \tag{15}$$

Finally, using the constraint (3) we obtain an equation relating Π_1 and Π_2,

$$\Pi_1(\tau) + \Pi_2(\tau) + \frac{A}{\Pi_2(\tau)} + \frac{B}{\Pi_1(\tau)} = 1. \tag{16}$$

Note that since $\Pi_2(0) + \Pi_3(0) < 1$ and $\Pi_1(0) + \Pi_4(0) < 1$, we have $A < 1/4$ and $B < 1/4$. Equations (14)–(16) determine the closed trajectories in a four-dimensional phase space [26], as illustrated in Figure 1.
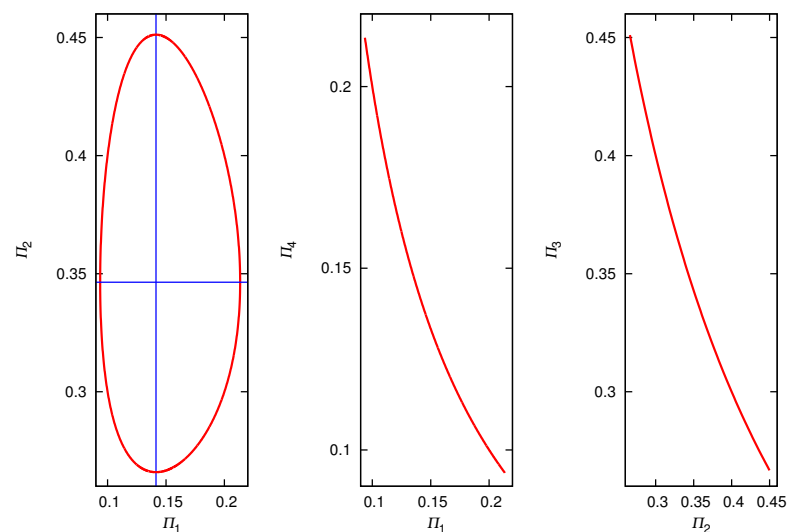
Figure 1. Projections of the trajectories of the phase space in the planes of the strategy frequencies. (**Left**) Π_1 vs. Π_2. (**Middle**) Π_1 vs. Π_4. (**Right**) Π_2 vs. Π_3. The trajectory in the left panel is counterclockwise. The parameters are $A = 3/25$ and $B = 1/50$. The domains of Π_1 and Π_4 are the interval $[0.0937, 0.213]$, while the domains of Π_2 and Π_3 are the interval $[0.266, 0.451]$. The thin horizontal and vertical blue lines are $\Pi_2 = \sqrt{A} \approx 0.346$ and $\Pi_1 = \sqrt{B} \approx 0.141$, respectively.

Using Equation (16) (see also the left panel of Figure 1), we can derive the domain of Π_1 by looking at the values of the strategy frequencies for which $d\Pi_2/d\Pi_1 \to \infty$. This divergence happens at $\Pi_2 = \sqrt{A}$, and by inserting this value in Equation (16), we obtain a quadratic equation for the extreme values of Π_1,

$$\Pi_1^2 - \Pi_1(1 - 2\sqrt{A}) + B = 0. \tag{17}$$

Denoting the two roots of this equation by Π_1^+ and Π_1^-, we have $\Pi_1^+ \Pi_1^- = B$, and so these roots yield the extremes values of Π_4 as well (see Equation (15) and middle panel of Figure 1). Explicitly,

$$\Pi_1^+ = \frac{1}{2} - \sqrt{A} + \left[(\frac{1}{2} - \sqrt{A})^2 - B \right]^{\frac{1}{2}}, \tag{18}$$

$$\Pi_1^- = \frac{1}{2} - \sqrt{A} - \left[(\frac{1}{2} - \sqrt{A})^2 - B \right]^{\frac{1}{2}}, \tag{19}$$

so that the domain of Π_1 is the interval $\left[\Pi_1^-, \Pi_1^+ \right]$ since $A < 1/4$.

A similar analysis of the values of the strategy frequencies for which $d\Pi_2/d\Pi_1 = 0$ allows us to show that the domain of Π_2 is the interval $\left[\Pi_2^-, \Pi_2^+ \right]$, where

$$\Pi_2^\pm = \frac{1}{2} - \sqrt{B} \pm \left[(\frac{1}{2} - \sqrt{B})^2 - A \right]^{\frac{1}{2}} \tag{20}$$

since $B < 1/4$.

Finally, note that the roots $\Pi_1^\pm$ and $\Pi_2^\pm$ are always real (and positive), since $\sqrt{A} + \sqrt{B} < 1/2$, which is easy to prove by maximizing $\sqrt{A} + \sqrt{B}$ with the constraint $\sum_i^4 \Pi_i(0) = 1$.

A closed trajectory in the four-dimensional space of strategy frequencies implies periodic solutions in τ for Π_i in Equations (10)–(13). The initial conditions, $\Pi_i(0)$, determine the constants A and B in Equations (14)–(16) and thus the phase trajectory in Figure 1. The corresponding periodic solutions are shown in Figure 2.
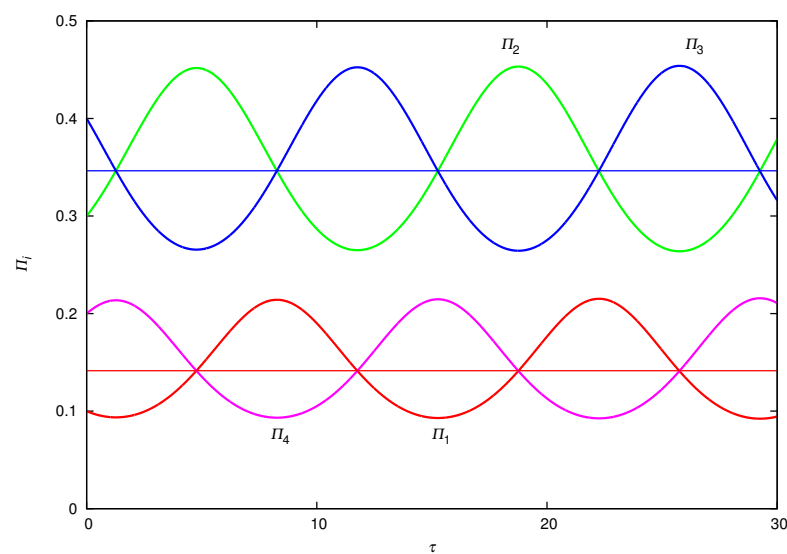
Figure 2. Periodic solutions for the strategy frequencies $\Pi_i(\tau)$ for the initial conditions $\Pi_1(0) = 0.1$, $\Pi_2(0) = 0.3$, $\Pi_3(0) = 0.4$ and $\Pi_4(0) = 0.2$, corresponding to $A = 3/25$ and $B = 1/50$. The thin horizontal blue and red lines are $\sqrt{A} \approx 0.346$ and $\sqrt{B} \approx 0.141$, respectively.

We now focus on estimating the period T of the periodic solutions of Equations (10)–(13). These equations can be reduced to a single differential equation for Π_1 using the Equations (14)–(16),

$$
\begin{aligned}
\frac{d\Pi_1}{d\tau} &= \pm\sqrt{(B + \Pi_1^2 - \Pi_1)^2 - 4A\Pi_1^2} \\
&= \pm\sqrt{(\Pi_1 - \Pi_1^l)(\Pi_1 - \Pi_1^-)(\Pi_1 - \Pi_1^+)(\Pi_1 - \Pi_1^u)}
\end{aligned} \tag{21}
$$

where

$$
\Pi_1^u = \frac{1}{2} + \sqrt{A} + \left[\left(\frac{1}{2} + \sqrt{A}\right)^2 - B\right]^{\frac{1}{2}}, \tag{22}
$$

$$
\Pi_1^l = \frac{1}{2} + \sqrt{A} - \left[\left(\frac{1}{2} + \sqrt{A}\right)^2 - B\right]^{\frac{1}{2}}, \tag{23}
$$

and Π_1^+ and Π_1^- are given by Equations (18) and (19), respectively. Note that $\Pi_1^u > \Pi_1^+$ and $\Pi_1^l < \Pi_1^-$ so that the expression under the square root in Equation (21) is always positive in the domain of Π_1, i.e., for $\Pi_1 \in [\Pi_1^-, \Pi_1^+]$. In Equation (21), the $+$ sign is selected when Π_1 increases from Π_1^- to Π_1^+ and the $-$ sign when Π_1 decreases from Π_1^+ to Π_1^-. Thus, the period is given by

$$
T = 2\int_{\Pi_1^-}^{\Pi_1^+} \frac{d\Pi_1}{\sqrt{(\Pi_1 - \Pi_1^+)(\Pi_1 - \Pi_1^-)(\Pi_1 - \Pi_1^u)(\Pi_1 - \Pi_1^l)}}. \tag{24}
$$

Although the integrand diverges at both integration limits, the improper integral can be easily evaluated numerically [27]. Figure 3 summarizes the results of the numerical solution of the elliptic integral that appears in Equation (24) in the case where B is fixed and A increases from 0 to $(1/2 - \sqrt{B})^2$.
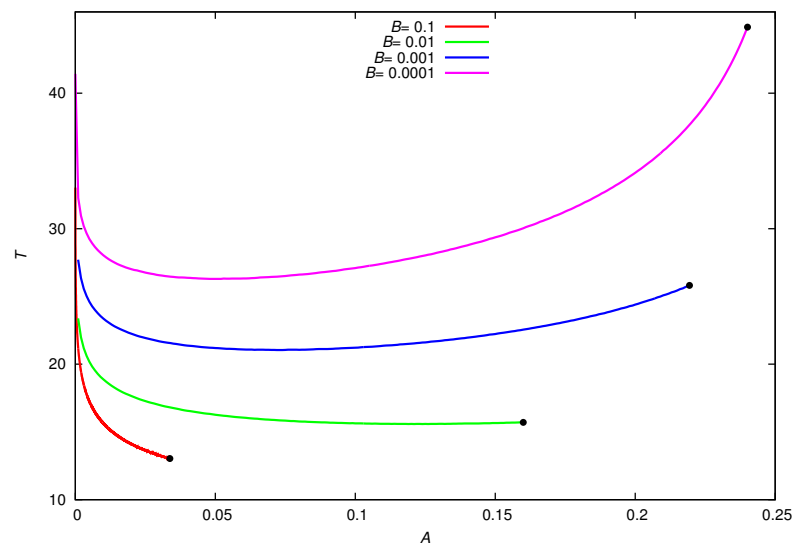
Figure 3. Period T of the solutions of the Equations (10)–(13) as function of $A \equiv \Pi_2(0)\Pi_3(0)$ for $B \equiv \Pi_1(0)\Pi_4(0) = 0.1, 0.01, 0.001$, and 0.0001 as indicated. The filled circles indicate the analytical estimate (30) for the period of small-amplitude oscillations.

The period T diverges for both extremes $A \to 0$ or $B \to 0$, since in these cases the dynamics are attracted by fixed points, as we will see later. Now, we offer an analytical estimate for the period of the periodic solutions of very small amplitude. These solutions occur when $\Pi_1^+ \to \Pi_1^-$, i.e., for $\sqrt{A} + \sqrt{B} \approx 1/2$, which gives $\Pi_1 \approx 1/2 - \sqrt{A} = \sqrt{B}$ (see Equations (18) and (19)). In phase space, the small-amplitude periodic solutions are represented by closed trajectories in the neighborhood of the neutral fixed point $\Pi_1^* = \Pi_4^* = \sqrt{B}$, $\Pi_2^* = \Pi_3^* = \sqrt{A}$ (see left panel of Figure 1). In fact, setting $\Pi_1 = \Pi_1^* + \epsilon_1$, $\Pi_2 = \Pi_2^* + \epsilon_2$, $\Pi_3 = \Pi_3^* + \epsilon_3$ and $\Pi_4 = \Pi_4^* + \epsilon_4$, and keeping only the linear terms in ϵ_i, we rewrite Equations (10)–(13) as

$$\frac{d\epsilon_1}{d\tau} = (\epsilon_2 - \epsilon_3)\sqrt{B} \tag{25}$$

$$\frac{d\epsilon_2}{d\tau} = (\epsilon_4 - \epsilon_1)\sqrt{A} \tag{26}$$

$$\frac{d\epsilon_3}{d\tau} = (\epsilon_1 - \epsilon_4)\sqrt{A} \tag{27}$$

$$\frac{d\epsilon_4}{d\tau} = (\epsilon_3 - \epsilon_2)\sqrt{B}, \tag{28}$$

which implies $\epsilon_1 + \epsilon_4 = K_{14}$ and $\epsilon_2 + \epsilon_3 = K_{23}$ where K_{14} and K_{23} are constants such that $K_{14} + K_{23} = 0$. Eliminating all variables in favor of ϵ_1 yields the nonhomogeneous harmonic oscillator equation

$$\frac{d^2\epsilon_1}{d\tau^2} + 4\sqrt{AB}\epsilon_1 = 2\sqrt{AB}K_{14} \tag{29}$$

from where we obtain the period

$$T = \pi(AB)^{-\frac{1}{4}}. \tag{30}$$

We emphasize that this equation holds only for $\sqrt{A} + \sqrt{B} = 1/2$, and Figure 3 shows that its estimate of T is in perfect agreement with the numerical evaluation of the elliptic integral that appears in Equation (24).

As noted, the above results hold for the case where all four strategies are present in the initial population, i.e., $\Pi_i(0) \neq 0$ for $i = 1, \dots, 4$. Now, we consider the case where some

strategies are missing at $t = 0$ and thus at all times $t > 0$. If $\Pi_2(0) = \Pi_3(0) = 0$, then the Equations (10) and (13) give $\Pi_1(\tau) = \Pi_1(0)$ and $\Pi_4(\tau) = \Pi_4(0)$, respectively. Similarly, $\Pi_1(0) = \Pi_4(0) = 0$ implies that $\Pi_2(\tau) = \Pi_2(0)$ and $\Pi_3(\tau) = \Pi_3(0)$. In other words, the replicator dynamics freezes at the initial frequencies.

A more interesting situation is when only one of the strategies is missing from the initial population, say $\Pi_3(0) = 0$. Setting $A = 0$ in Equation (16) we restate Equation (10) as

$$\frac{d\Pi_1}{d\tau} = \Pi_1 - \Pi_1^2 - B = \frac{1}{4} - B - \left(\Pi_1 - \frac{1}{2}\right)^2. \tag{31}$$

Although we can easily solve this equation for any τ, we will consider only the asymptotic solution $\Pi_1^* = \lim_{\tau \to \infty} \Pi_1(\tau)$ obtained by setting $d\Pi_1/d\tau = 0$ in Equation (31). Recalling that $\Pi_4^* = B/\Pi_1^*$ and $B < 1/4$ we find

$$\Pi_1^* = \frac{1}{2} + \sqrt{\frac{1}{4} - B} \tag{32}$$

$$\Pi_4^* = \frac{1}{2} - \sqrt{\frac{1}{4} - B}. \tag{33}$$

Moreover, since $\Pi_1^* + \Pi_4^* = 1$, we have $\Pi_2^* = 0$. In this scenario (i.e., for $\Pi_3(0) = 0$), Equations (10) and (13) indicate that Π_1 increases and Π_4 decreases with τ. In fact, the short-lived presence of strategy 2, i.e., (T, D) telling the truth and not believing, increases the payoff of strategy 1, i.e., (T, B) telling the truth and believing, but decreases the payoff of strategy 4, i.e., (L, D), which always fails to predict the true outcome of the coin toss announced by the players using the other strategies. Figure 4 illustrates the time evolution of the non-zero strategy frequencies in this scenario. Note that the fixed point is neutral: an arbitrary perturbation to Π_1^*, Π_2^*, and Π_4^* leads to different value of B and thus to a different fixed point.
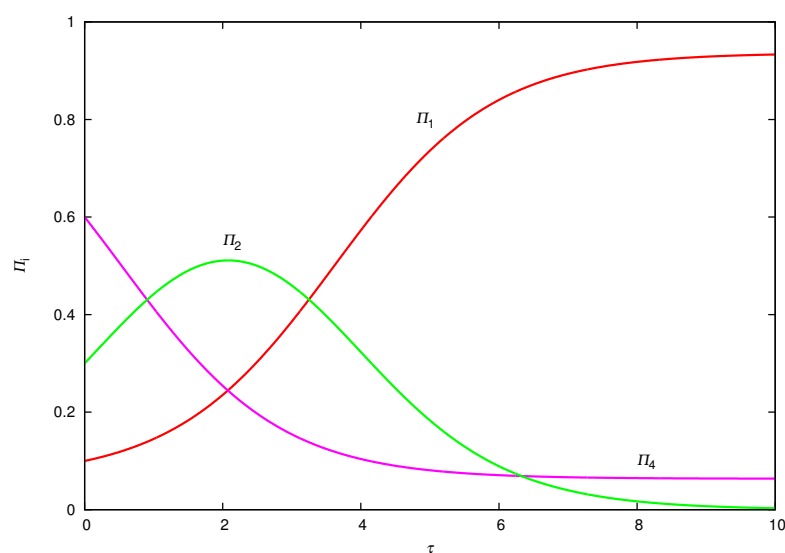

Figure 4. Time evolution of the strategy frequencies $\Pi_i(\tau)$ for the initial conditions $\Pi_1(0) = 0.1$, $\Pi_2(0) = 0.3$, $\Pi_3(0) = 0$ and $\Pi_4(0) = 0.6$, corresponding to $A = 0$ and $B = 0.06$. The fixed points are $\Pi_1^* \approx 0.936$, $\Pi_2^* = 0$, and $\Pi_4^* \approx 0.064$.

A similar analysis can be performed by considering different missing strategies. For example, if strategy 2 is missing, i.e., $\Pi_2(0) = \Pi_2(\tau) = 0$, then the above results hold if we exchange the labels of strategies 1 and 4.

5. Finite Population Simulations

As noted above, the formulation of the replicator equation corresponds to the infinite population size limit of a (finite) population of individuals who change their strategies by comparing their payoff performance with that of their peers [21]. This social learning process is of utmost importance as it has been claimed that imitation is the fabric of human society [28], a point nicely summarized by the phrase "Imitative learning acts like a synapse, allowing information to leap the gap from one creature to another" [29]. In addition, imitation of more successful individuals has inspired search algorithms [30,31] where social learning replaces the biological processes of selection and recombination of traditional evolutionary algorithms [32]. In this sense, the replicator equation formalism becomes an essential tool for modeling the dynamics of social systems [33]. To highlight the surprising link between social learning and the replicator equation [21], we present here finite population simulations of the two-choice sender–receiver game where social learning is explicitly considered in the interactions of the players.

In particular, we consider the two-choice sender–receiver game among N players who interact pairwise in a well-mixed population, i.e., each player interacts with the other $N - 1$ players in the population. The strategies $i = 1, \ldots, 4$ of the initial population are given by the proportions $\Pi_i(0)$. Note that due to the strong dependence on the initial conditions, we cannot randomly assign the strategies to the players in such a way as to obtain the proportions $\Pi_i(0)$ on average: the proportion of strategy i among the N players must perfectly match $\Pi_i(0)$ in order to reproduce the analytical results of the previous section. In each round of the game, which comprises $N(N - 1)$ interactions, we evaluate the average payoff of each player $a = 1, \ldots, N$ and denote it by w_a. Then, we pick a player at random, say player b, and compare her average payoff to the payoff of another randomly chosen player, say player a. Player b switches to the strategy of player a with probability

$$p = \frac{w_a - w_b}{\Delta w_{\max}} \tag{34}$$

where $\Delta w_{\max}$ is the maximum possible average payoff difference that guarantees $p \leq 1$. For example, for the deceiving game we have $\Delta w_{\max} = b_1 + b_2 + c_1 + c_2 = 2(b_1 + c_1)$. If $w_a < w_b$, then player b keeps her strategy. If player b changes strategy, the average payoffs of all individuals in the population are recalculated. This completes the time step δt of the imitation dynamics and so the time is updated to $t = t + \delta t$, where $\delta t = 1/(N \Delta w_{\max})$ [21]. In the seminal paper [21], the probability of switching strategies has an additional term that does not depend on the average rewards of the players. If that term is omitted, we obtain Equation (34). Our implementation of the stochastic simulations of the imitation dynamics is standard in the evolutionary game literature (see, e.g., [34,35]). We note that the replicator equation can be derived as the large population limit of a variety of rules for individual learning [36]. This procedure can be rigorously justified and exact estimates of the deviation of the replicator dynamics from a finite population dynamics are available [37].

Randomness enters our simulations in two situations. The first situation is when the two players a and b are randomly chosen and their average payoffs w_a and w_b are compared. The second situation occurs if $w_a > w_b$, where a uniformly distributed random number in the unit interval u is generated and compared to the probability p given in Equation (34): if $u < p$, player b switches to the strategy of player a.

Figure 5 shows the results of four runs of the stochastic imitation dynamics for the case $b_2 + c_2 > 0$ and $b_2 + c_2 \neq b_1 + c_1$ mentioned in Section 3, where the coexistence fixed point $\Pi_1^* = \Pi_4^*$ and $\Pi_2^* = \Pi_3^*$ depends on the initial frequencies of the strategies. There is very good agreement between the deterministic predictions of the replicator equation and the runs of the stochastic imitation dynamics. In this case, the average payoff in the population, Equation (9), depends on the frequencies of the strategies and so there is no way to obtain the coexistence fixed point other than to solve Equation (2) numerically.
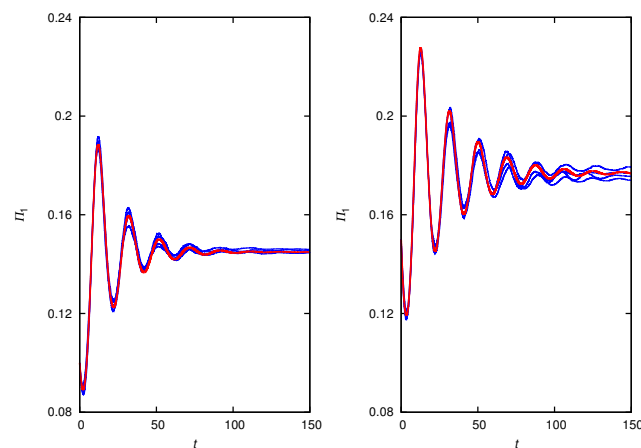
Figure 5. Time evolution of the strategy frequency Π_1 for N = 20,000 and different initial conditions. (**Left**) $\Pi_1(0) = 0.1$, $\Pi_2(0) = 0.3$, $\Pi_3(0) = 0.4$, and $\Pi_4(0) = 0.2$, which leads to $\Pi_1^* = \Pi_4^* \approx 0.145$ and $\Pi_2^* = \Pi_3^* \approx 0.355$. (**Right**) $\Pi_1(0) = 0.15$, $\Pi_2(0) = 0.25$, $\Pi_3(0) = 0.4$, and $\Pi_4(0) = 0.2$, which leads to $\Pi_1^* = \Pi_4^* \approx 0.177$ and $\Pi_2^* = \Pi_3^* \approx 0.323$. The parameters of the payoff matrix are $b_1 = 1$, $b_2 = 0.5$, and $c_1 = c_2 = 0$. The thick red curves are the solutions of the replicator equation, and the thin blue curves are runs of the stochastic imitation dynamics.

Now, we return to the setting of parameters that characterize the deceiving game, i.e., $b_1 + c_1 = b_2 + c_2$, but consider initial conditions that cause the frequencies of the pair of completely opposite strategies Π_1 and Π_4 to take on values very close to zero. The results are shown in Figure 6. The deterministic results of the replicator equation agree very well with the four runs of the stochastic imitation dynamics until the number of players using strategy 1 becomes less than 10, when each run begins to follow different trajectories due to the strong effects of stochasticity on this population of players. In fact, for one of the runs shown in Figure 6, strategy 4 becomes extinct at $t = 27.3$, leading to the subsequent extinction of strategy 1 at $t = 42.6$. After the extinction of strategy 4, the stochastic dynamics of the remaining strategies follow a pattern similar to that shown in Figure 4. Of course, if we follow the stochastic dynamics further in time, all runs will lead to the extinction of strategies 1 and 4. This extinction can be delayed by increasing the size of the population N, and it can be avoided altogether in the infinite population limit.
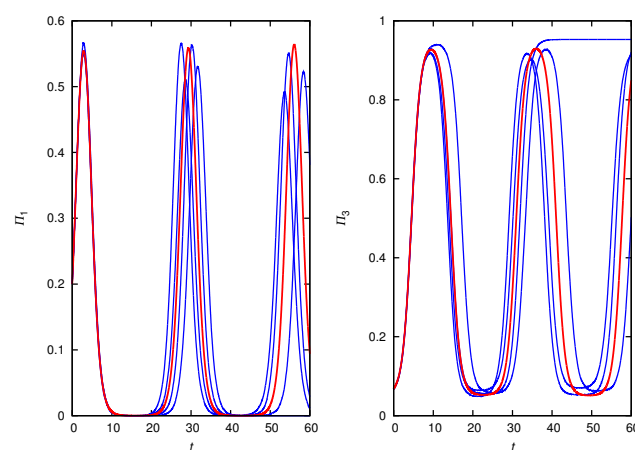

Figure 6. Periodic solutions for the strategy frequencies of the deceiving game for N = 20,000. (**Left**) Π_1 vs. t. (**Right**) Π_3 vs. t. The initial conditions are $\Pi_1(0) = 0.2$, $\Pi_2(0) \approx 0.731$, $\Pi_3(0) \approx 0.068$ and $\Pi_4(0) = 0.0005$ which corresponds to $A = 0.05$ and $B = 0.0001$. The parameters of the payoff matrix are $b_1 = b_2 = 1$, and $c_1 = c_2 = 0$. The thick red curves are the solutions of the replicator equation and the thin blue curves are runs of the stochastic imitation dynamics.

6. Discussion

Our study of the two-choice sender–receiver game offers some interesting caveats to Kant's conclusion that a world in which everyone lies simply cannot exist, while a world in which everyone tells the truth is possible [3,4]. Of course, in the binary world considered here, where disbelieving a lie is equivalent to believing a truth, there is complete symmetry between the strategies (T, B), i.e., always tell the truth and believe the sender's report, and (L, D), i.e., always lie and disbelieve the sender's report. In this sense, our results show that a world consisting only of liars and unbelievers or only of truth tellers and believers is possible, provided that the sender has no intention to deceive the receiver, which amounts to satisfying the condition $b_2 + c_2 < 0$ for the parameters that determine the sender's payoff (see Table 2).

A less optimistic and perhaps more realistic scenario is characterized by the condition $b_1 + c_1 = b_2 + c_2 > 0$ and is called the deceiving game. It describes the situation where the sender's intentions do not coincide with the receiver's interests. In this case, our results show a dynamic coexistence among the four possible strategies of the players. An instructive way to visualize this conclusion is through the phase plane, which shows the proportion of truth tellers $\Pi_T = \Pi_1 + \Pi_2$ and believers $\Pi_B = \Pi_1 + \Pi_3$ in the population at a given time. In words, Π_T is the proportion of players using either strategy (T, B) or strategy (T, D), while Π_B is the proportion of players using either strategy (T, B) or strategy (L, B). Figure 7 shows the phase planes for the two different initial conditions we used to generate Figures 2 and 6 in the infinite population limit of the deceiving game. Essentially, these results show that the increase in the proportion of believers Π_B is accompanied by an increase in the proportion of liars (i.e., $1 - \Pi_T$) until it no longer makes sense to believe and Π_B begins to decrease. Since in our binary world, the way to deceive a non-believer is to tell the truth, the decrease in Π_B is accompanied by the increase in Π_T. Note that the closed trajectories are centered on the neutral fixed point $\Pi_T^* = \Pi_1^* + \Pi_2^* = \sqrt{B} + \sqrt{A} = 1/2$, and $\Pi_B^* = \Pi_1^* + \Pi_3^* = \sqrt{B} + \sqrt{A} = 1/2$.
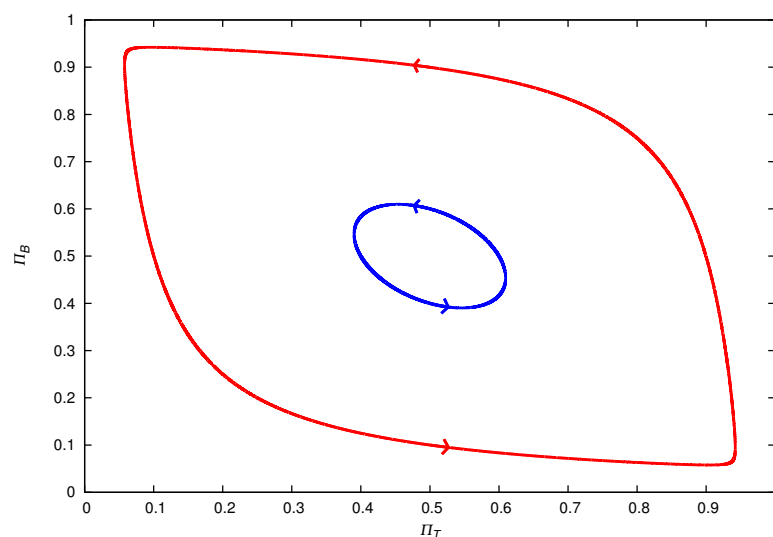

Figure 7. Trajectories in the phase plane showing the frequency of truth tellers Π_T and the frequency of believers Π_B for $A = 3/25$ and $B = 1/50$ (blue curve) and $A = 0.05$ and $B = 0.0001$ (red curve) in an infinite population. The parameters A and B are determined by the initial conditions according to Equations (14) and (15), respectively. The trajectories are counterclockwise, as indicated.

7. Conclusions

A pervasive aspect of our study of the two-choice sender–receiver game when the sender and receiver interests are antagonistic (i.e., $b_2 + c_2 > 0$) is the important role played by the frequencies of the strategies in the initial population. This makes the analytical study

of this game rather challenging, as only a thorough solution of the replicator Equation (2) provides information about the time asymptotic behavior of the game. For the deceiving game, where there is symmetry in the rewards and costs of the sender and receiver (i.e., $b_1 + c_1 = b_2 + c_2 > 0$), the solutions of the replicator equation are periodic and the system is conservative, so that the dynamics always return to the initial setting of the strategy frequencies. This means that the imitation dynamics are not powerful enough to permanently alter the initial strategy frequencies. When this symmetry is broken (i.e., $b_1 + c_1 \neq b_2 + c_2 > 0$), the dynamics converges to fixed points determined by the initial conditions. The robust influence of player history on the outcome of the two-choice sender–receiver game is a non-trivial result of our study and it may reflect the unpredictability of social systems whose members have antagonistic interests. It would be interesting to see if this conclusion holds true in more realistic scenarios where the game takes place within free-forming or casual groups so that the well-mixed population assumption is relaxed [38,39].

Author Contributions: Conceptualization, J.F.F.; formal analysis, E.V.M.V. and J.F.F.; investigation, E.V.M.V.; methodology, J.F.F.; supervision, J.F.F.; validation, E.V.M.V.; writing—original draft, J.F.F.; writing—review and editing, E.V.M.V. and J.F.F. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data will be made available on request.

References

1. Maynard Smith, J.; Harper, D. *Animal Signals*; Oxford University Press: Oxford, UK, 2003.
2. Zahavi, A. Mate selection—A selection for a handicap. *J. Theor. Biol.* **1975**, *53*, 205–214. [CrossRef]
3. Kant, I. *Groundwork of the Metaphysics of Morals*; Cambridge University Press: Cambridge, UK, 2012.
4. Sober, E. The primacy of truth-telling and the evolution of lying. In *From a Biological Point of View: Essays in Evolutionary Philosophy*; Sober, E., Ed.; Cambridge University Press: Cambridge, UK, 1994; pp. 71–92.
5. Fontanari, J.F. Kant's Modal Asymmetry between Truth-Telling and Lying Revisited. *Symmetry* **2023**, *15*, 555. [CrossRef]
6. Fallis, D. What Is Disinformation? *Libr. Trends* **2015**, *63*, 401–426. [CrossRef]
7. Seger, E.; Avin, S.; Pearson, G.; Briers, M.; Heigeartaigh, S.O.; Bacon, H. *Tackling Threats to Informed Decision-Making in Democratic Societies*; The Alan Turing Institute: London, UK, 2020.
8. Del Vicario, M.; Bessi, A.; Zollo, F.; Petroni, F.; Scala, A.; Caldarelli, G.; Stanley, H.E.; Quattrociocchi, W. The spreading of misinformation online. *Proc. Natl Acad. Sci. USA* **2016**, *113*, 554–559. [CrossRef]
9. Tórtura, H.A.; Fontanari, J.F. The synergy between two threats: Disinformation and COVID-19. *Math. Models Methods Appl. Sci.* **2022**, *32*, 2077–2097.
10. Forgó, F.; Fülöp, J.; Prill, M. Game theoretic models for climate change negotiations. *Eur. J. Oper. Res.* **2005**, *160*, 252–267. [CrossRef]
11. Kopec, M. Game Theory and the Self-Fulfilling Climate Tragedy. *Environ. Values* **2017**, *26*, 203–221. [CrossRef]
12. Chander, P. *Game Theory and Climate Change*; Columbia University Press: New York, NY, USA, 2018.
13. Mond, D. Game theory and climate change. In *The Impacts of Climate Change*; Letcher, T.M., Ed.; Elsevier: Amsterdam, The Netherlands, 2021; pp. 437–451.
14. Ariely, D. *The Honest Truth About Dishonesty: How We Lie to Everyone-Especially Ourselves*; Harper Perennial: New York, NY, USA, 2013.
15. Gneezy, U. Deception: The role of consequences. *Am. Econ. Rev.* **2005**, *95*, 384–394. [CrossRef]
16. Erat, S.; Gneezy, U. White lies. *Manag. Sci.* **2012**, *58*, 723–733. [CrossRef]
17. Capraro, V.; Perc, M.; Vilone, D. The evolution of lying in well-mixed populations. *J. R. Soc. Interface* **2019**, *16*, 20190211. [CrossRef] [PubMed]
18. Capraro, V.; Perc, M.; Vilone, D. Lying on networks: The role of structure and topology in promoting honesty. *Phys. Rev. E* **2020**, *101*, 032305. [CrossRef] [PubMed]
19. Xia, C.; Wang, J.; Perc, M.; Wang, Z. Reputation and reciprocity. *Phys. Life Rev.* **2023**, *46*, 8–45. [CrossRef] [PubMed]

20. Hofbauer, J.; Sigmund, K. *Evolutionary Games and Population Dynamics*; Cambridge University Press: Cambridge, UK, 1998.
21. Traulsen, A.; Claussen, J.C.; Hauert, C. Coevolutionary Dynamics: From Finite to Infinite Populations. *Phys. Rev. Lett.* **2005**, *95*, 238701. [CrossRef]
22. Dolfin, M.; Leonida, L.; Outada, N. Modeling human behavior in economics and social science. *Phys. Life Rev.* **2017**, *22–23*, 1–21. [CrossRef] [PubMed]
23. Bellomo, N.; Esfahanian, M.; Secchini, V.; Terna, P. What is life? Active particles tools towards behavioral dynamics in social-biology and economics. *Phys. Life Rev.* **2022**, *43*, 189–207. [CrossRef] [PubMed]
24. De Silva, H.; Sigmund, K. Public Good Games with Incentives: The Role of Reputation. In *Games, Groups, and the Global Good*; Levin, S.A., Ed.; Springer: New York, NY, USA, 2009; pp. 85–103.
25. Maynard Smith, J. *Evolution and the Theory of Games*; Cambridge University Press: Cambridge, UK, 1982.
26. Murray, J.D. *Mathematical Biology: I. An Introduction*; Springer: New York, NY, USA, 2007.
27. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. *Numerical Recipes in Fortran: The Art of Scientific Computing*; Cambridge University Press: Cambridge, UK, 1992.
28. Blackmore, S. *The Meme Machine*; Oxford University Press: Oxford, UK, 2000.
29. Bloom, H. *Global Brain: The Evolution of Mass Mind from the Big Bang to the 21st Century*; Wiley: New York, NY, USA, 2001.
30. Kennedy, J. Thinking is social: Experiments with the adaptive culture model. *J. Conflict. Res.* **1998**, *42*, 56–76. [CrossRef]
31. Fontanari, J.F. Imitative Learning as a Connector of Collective Brains. *PLoS ONE* **2014**, *9*, e110517. [CrossRef]
32. Goldberg, D.E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley: Reading, PA, USA, 1989.
33. Nowak, M.A. *Evolutionary Dynamics: Exploring the Equations of Life*; Belknap Press: New York, NY, USA, 2006.
34. Zheng, D.F.; Yin, H.P.; Chan, C.H.; Hui, P.M. Cooperative behavior in a model of evolutionary snowdrift games with N-person interactions. *Europhys. Lett.* **2007**, *80*, 18002. [CrossRef]
35. Meloni, S.; Buscarino, A.; Fortuna, L.; Frasca, M.; Gómez-Gardeñes, J.; Latora, V.; Moreno, Y. Effects of mobility in a population of prisoner's dilemma players. *Phys. Rev. E* **2009**, *79*, 067101. [CrossRef]
36. Sandholm, W.H. *Population Games and Evolutionary Dynamics*; MIT Press: Cambridge, UK, 2010.
37. Kolokoltsov, V. The evolutionary game of pressure (or interference), resistance and collaboration. *Math. Oper. Res.* **2017**, *42*, 915–944. [CrossRef]
38. Fontanari, J.F. Stochastic Simulations of Casual Groups. *Mathematics* **2023**, *11*, 2152. [CrossRef]
39. Sumpter, D.J.T. *Collective Animal Behavior*; Princeton University Press: Princeton, NJ, USA, 2010.