Supplementary Information for Intuition, deliberation, and the evolution of cooperation

Adam Bear, David G. Rand

Contents

1 Strategy space and payoff function

Our main model considers agents playing 1-shot Prisoner's Dilemma (PD) games or PDs with reciprocal consequences (modeled using the framework of infinitely repeated games); and responding using either a generalized intuition S_i or paying a cost d^* (stochastically sampled from the interval $[0, d]$) to deliberate and tailor their strategy such that they use strategy S_r if the game is repeated and S_1 if the game is 1-shot. In each interaction, agents choose either the cooperative strategy tit-for-tat (TFT) or the non-cooperative strategy always defect (ALLD). Importantly, as we demonstrate below in Section 6, our results are not specific to TFT and ALLD playing repeated PDs, but instead generalize to a wide range of coordination games.

An agent's strategy profile is specified by four variables: their (i) probability of intuitively playing TFT S_i, (ii) probability of playing TFT when they deliberate and face a repeated game S_r, (iii) probability of playing TFT (i.e. cooperating) when they deliberate and face a 1-shot game S_1, and (iv) maximum acceptable cost of deliberation T. Since we stipulate that the cost of deliberation is sampled uniformly from the interval $[0, d]$, an agent with threshold T deliberates with probability $\frac{T}{d}$ and on average pays a cost $\frac{T}{2}$ when deliberating. We use a uniform distribution for simplicity, but more realistic cost distributions should not change our results qualitatively.

Here we specify the expected payoff $\pi(x, y)$ of an agent with strategy profile $x = [S_i, S_1, S_r, T]$ playing against an agent with strategy profile $y = [S_i', S_1', S_r', T']$. To do so, we calculate agent x's expected payoff from playing infinitely repeated PDs with probability p and 1-shot PDs with $1-p$, over the cases in which (i) both agents deliberate (probability $\frac{TT'}{d^2}$), (ii) agent x deliberates and agent y decides intuitively (probability $\frac{T}{d}(1-\frac{T'}{d})$), (iii) agent x decides intuitively and agent y deliberates (probability $(1-\frac{T}{d})\frac{T'}{d}$), and (iv) both agents decide intuitively (probability $(1-\frac{T}{d})(1-\frac{T'}{d})$):

$$\pi(x,y) = \frac{TT'}{d^2}(\pi_{DD} - \frac{T}{2}) + \frac{T}{d}(1-\frac{T'}{d})(\pi_{DI} - \frac{T}{2}) + (1-\frac{T}{d})\frac{T'}{d}\pi_{ID} + (1-\frac{T}{d})(1-\frac{T'}{d})\pi_{II}$$

where π_{DD} is agent x's expected payoff when both agents deliberate, π_{DI} is agent x's expected payoff when agent x deliberates and agent y uses intuition, and so on.

These expected payoffs are calculated based on the payoff tables for 1-shot and repeated PDs. In 1-shot games, TFT cooperates and pays a cost c to give a benefit b to the partner, while ALLD defects and does nothing. Thus, the payoff table for the 1-shot games is given by

1-shot PD Payoffs

	TFT	**ALLD**
TFT	$b-c$	$-c$
ALLD	b	0

where the row player's payoff is shown.

To make payoffs in an infinitely repeated game comparable to those of a 1-shot game, we use the average payoff per round. Here, two TFT agents cooperate with each other in every round and earn average payoffs per round of $b-c$, while two ALLD agents defect every round, earning 0.

Thus these payoffs are the same as the 1-shot PD. When a TFT agent and an ALLD agent meet, however, the outcome differs from the 1-shot game, because the TFT agent cooperates only on the first round, and then defects in every subsequent round. Because the interaction is modeled as being infinitely repeated, the first round (where TFT cooperates) contributes only a negligible amount to the average payoff. Therefore, both agents earn an average payoff per round of 0. Therefore the payoff table for the infinitely repeated PD is given by

Infinitely Repeated PD Payoffs

	TFT	**ALLD**
TFT	$b - c$	0
ALLD	0	0

where $b, c > 0$.

Importantly, using total payoff (rather than average payoff per round) in a game with a finite continuation probability, such that the first round does influence payoffs and causes some negative cost for TFT and positive benefit for ALLD, does not qualitatively change our results; see Section 6 below.

Substituting in relevant payoff values yields

$$\begin{aligned}
\pi_{DD} &= p(S_r S'_r(b-c)) + (1-p)(S_1 S'_1(b-c) + S_1(1-S'_1)(-c) + (1-S_1)S'_1 b) \\
\pi_{DI} &= p(S_r S'_i(b-c)) + (1-p)(S_1 S'_i(b-c) + S_1(1-S'_i)(-c) + (1-S_1)S'_i b) \\
\pi_{ID} &= p(S_i S'_r(b-c)) + (1-p)(S_i S'_1(b-c) + S_i(1-S'_1)(-c) + (1-S_i)S'_1 b) \\
\pi_{II} &= p(S_i S'_i(b-c)) + (1-p)(S_i S'_i(b-c) + S_i(1-S'_i)(-c) + (1-S_i)S'_i b)
\end{aligned}$$

.

2 Nash equilibrium calculations

2.1 Setup

To facilitate Nash equilibria calculations, we consider a strategy space which is simplified relative to the main model in two ways: (i) agents' intuitive response S_i is limited to being either 0 (never play TFT) or 1 (always play TFT); and (ii) agents' deliberative responses are fixed to be $S_1 = 0$ and $S_r = 1$; i.e., always defecting when deliberating and facing a 1-shot game, and always playing TFT when deliberating and facing a repeated game. As in the main model, agents specify a maximum cost of deliberation T ($0 \leq T \leq d$) that they are willing to pay in order to deliberate, and this determines when they deliberate.

Thus, an agent's strategy profile is specified by two variables: 1) a binary variable S_i indicating whether or not the agent intuitively plays the cooperative strategy and 2) a continuous variable T indicating the agent's maximum cost they are willing to pay to deliberate. We denote a strategy profile for this reduced strategy space as $x = [S_i, T]$. (These simplifications of the intuitive and deliberative strategy spaces are justified by our evolutionary simulations using the full strategy space, whose results are in agreement with the results of the Nash calculation for the simplified

strategy space – see main text Figure 2.)

A strategy profile x is a Nash equilibrium if no strategy profile y is able to get a higher payoff against x than x gets against itself. That is,

$$\forall y : \pi(x, x) \geq \pi(y, x).$$

Given our restricted strategy space, the set of possible strategy profiles that an agent can adopt can be thought of as two continuous sets: 1) the set of strategy profiles that intuitively defect ($S_i = 0$) and have threshold $0 \leq T \leq d$, and 2) the set of strategy profiles that intuitively cooperate ($S_i = 1$) and have threshold $0 \leq T \leq d$.

2.2 Intuitively defecting equilibria

We first consider whether any strategy profile with $S_i = 0$ is a Nash. To do this, we calculate the expression for the payoff that an agent with $S_i = 0$ and $T = T$ gets against an agent with $S_i = 0$ and $T = T'$:

$$\pi([0, T], [0, T']) = \frac{T^2(\frac{T'}{d} - 1)}{2d} - \frac{TT'(\frac{T'}{2} - p(b - c))}{d^2}.$$

Since the concavity of this function (with respect to T) is always negative ($\frac{\partial^2}{\partial T^2}\pi([0, T], [0, T']) < 0$), there is a unique best-response $[0, T_b]$ that maximizes one's payoff when playing against $[0, T']$, which can be found by asking what value of T satisfies the equation

$$\frac{\partial}{\partial T}\pi([0, T], [0, T']) = 0.$$

Doing so yields

$$T_b = \frac{p(b - c)T'}{d}.$$

Since a strategy profile must be a best response against itself in order to be Nash, it must be the case that $T_b = T'$ in the above equation for T' to be Nash. That is, this is the unique case in which T' maximizes its payoff by playing itself. Solving for T' yields the solution of $T' = 0$ (regardless of the values of any of the parameters). Thus, $[0, 0]$, a strategy that never deliberates and always defects, is a best response to itself.

For the strategy $[0, 0]$ to be a Nash, however, it must also be the case that no intuitively cooperative strategy can beat it. This follows straightforwardly. The payoff that strategy $[0, 0]$ gets against itself is 0 (since neither player is paying a cost of cooperation to benefit the other or paying a cost to deliberate). Any intuitively cooperative strategy, on the other hand, is going to incur a cooperation cost c on the fraction of interactions that it cooperates. Moreover, since the $[0, 0]$ agent is always defecting, this intuitively cooperative strategy receives no benefit from the $[0, 0]$ agent. Thus, its payoff is always negative and it cannot invade the $[0, 0]$ strategy under any value of p. As a result, $[0, 0]$ (referred to as the "Intuitive defector (ID)" strategy profile in the main text) is always a Nash equilibrium.

2.3 Purely deliberative equilibrium

Next we investigate the other boundary case of $[0, d]$, a purely deliberative agent that never uses intuition. Note that because this agent never uses intuition, the intuitive response S_i is irrelevant, such that the strategy $[0, d]$ is functionally identical to $[1, d]$. We therefore refer to this strategy as $[-, d]$. To see whether this strategy can be Nash, we start by asking what the best response $[0, T_b]$ is when playing against $[-, d]$. Using the expression above, we find $T_b = p(b - c)$. This makes it seem that $[-, d]$ is not Nash because $T_b \neq d$ (except in the special case where $d = p(b - c)$).

However, because d is the maximum cost of deliberation, T is bounded such that $0 \leq T \leq d$. Therefore, when $d < p(b - c)$, this best response $T_b = p(b - c)$ lies outside the range of possible T values. Recall that because $\frac{\partial^2}{\partial T^2}\pi([0, T], [0, T']) < 0$ is satisfied for all T, T',the payoff $\pi([0, T], [-, d])$ decreases monotonically as T moves further from the best-response value $p(b - c)$. Thus, when $d < p(b-c)$ (such that the best response is greater than the maximum value of T), the value of T within the allowed interval which best responds to $[-, d]$ is in fact $[-, d]$ itself (i.e. the maximum allowed value of T).

We find a similar result when asking which intuitively cooperative strategy best-responds to $[-, d]$. Solving $\frac{\partial}{\partial T}\pi([1, T], [-, d]) = 0$ gives a best response of $[1, c(1-p)]$. Thus, by the logic from the preceding paragraph, $[-, d]$ cannot be beaten by any intuitively cooperative strategies if $d < c(1 - p)$. As a result, we see that the purely deliberative strategy $[-, d]$ can be Nash when the maximum cost of deliberation is sufficiently small, such that both $d \leq p(b - c)$ and $d \leq c(1 - p)$ are satisfied.

This result is natural - if deliberating were free, it would obviously be better to deliberate in our model than to use intuition. Thus it is no surprise that there is a minimum d above which it is no longer worth paying to deliberate on all occasions. Given the wide-spread use of intuition by humans, we believe it is a safe assumption that the $d > p(b - c), c(1 - p)$ condition is satisfied.

2.4 Intuitively cooperating equilibria

We next consider whether any intuitively cooperative strategy profile is a Nash. Following the procedure used above, we calculate the expression for the payoff that an intuitively cooperative agent with strategy profile $[1, T]$ gets against an intuitively cooperative agent with strategy profile $[1, T']$:

$$\begin{aligned}\pi([1, T], [1, T']) = &\frac{((1-p)(-b) - (b-c)p + \frac{T}{2})(\frac{T'}{d} - 1)T}{d} \\ &+ (p(b-c) - (1-p)(c-b))(1 - \frac{T}{d})(1 - \frac{T'}{d}) \\ &- \frac{(\frac{T}{2} - p(b-c))TT'}{d^2} \\ &- \frac{((1-p)c - p(b-c))(1 - \frac{T}{d})T'}{d}.\end{aligned}$$

We then find the best-response T_b by solving for when the partial derivative of this expression with respect to T is 0, yielding

$$T_b = (1-p)c.$$

Thus, an intuitively cooperative agent's best response against another intuitively cooperative agent is to deliberate only in cases where the cost of deliberation is not greater than $(1-p)c$. Note that this is the product of the probability of a 1-shot game occurring $(1-p)$ and the cost of cooperating c, which is precisely the expected benefit of deliberation for an intuitive cooperator (since what deliberation does here is allow the agent to override her cooperative intuition when she finds herself in a 1-shot game).

In order to test whether the strategy profile $[1, c(1-p)]$ is Nash, we must also consider whether any intuitively defective strategy profile $[0, T']$ can beat it. To do this, we find the intuitively defective strategy profile that is a best response against the optimal intuitively cooperative strategy profile $[1, (1-p)c]$ by solving $\frac{\partial}{\partial T}\pi([0,T],[1,(1-p)c]) = 0$ for T.

This yields the strategy profile $[0, p(b-c)]$ as the intuitively defecting strategy that performs best against the intuitively cooperative strategy $[1, (1-p)c]$. We then find the conditions under which the optimal intuitively cooperative strategy profile does better against itself than the best response intuitive defecting strategy profile does,

$$\pi([1,(1-p)c],[1,(1-p)c]) \geq \pi([0,p(b-c)],[1,(1-p)c]),$$

in order to find out when $[1, (1-p)c]$ is Nash. We find that this inequality is satisfied when $p \geq \frac{c}{b}$.

It is also necessary to consider whether $[1, c(1-p)]$ can be beaten in the boundary case where $d > (1-p)c$, but $d \leq p(b-c)$, such that the best response against $[1, (1-p)c]$ is actually $[-, d]$ (as $T = p(b-c)$ is outside the allowed range). Doing so, we find that it is always the case that $\pi([1,(1-p)c],[1,(1-p)c]) \geq \pi([-,d],[1,(1-p)c])$ when $p \geq \frac{c}{b}$. Thus the purely deliberative agent $\pi([-, d]$ cannot invade the intuitively cooperative strategy under these conditions.

We therefore conclude that the intuitively cooperative strategy profile $[1, (1-p)c]$ is a Nash equilibrium when $p \geq \frac{c}{b}$.

2.5 Summary of Nash results

In sum, we find two main equilibria:

1. The Intuitive Defector (ID) strategy profile that intuitively defects ($S_i = 0$) and never deliberates ($T = 0$) is always Nash (the deliberative strategy variables S_1 and S_r are irrelevant, as this strategy never deliberates).

2. The Dual-process Cooperator (DC) strategy profile that intuitively plays TFT ($S_i = 1$), deliberates when the cost of deliberation is no greater than $T = (1-p)c$, and deliberatively plays TFT in repeated games ($S_r = 1$) and deliberatively defects in 1-shot games ($S_1 = 0$), is Nash when repeated games are sufficiently common, $p \geq c/b$.

In addition, a purely deliberative strategy that never uses intuition ($T = d$, thus the value of S_i is irrelevant), and deliberatively plays TFT in repeated games ($S_r = 1$) and deliberatively defects in 1-shot games ($S_1 = 0$), is Nash when the maximum cost of deliberation is sufficiently small, $d \leq c(1-p)$ and $d \leq p(b-c)$, such that it is always worth paying to deliberate. As this behavior is psychologically unrealistic, we focus our evolutionary analyses on parameter regions where d is large enough to make this strategy not an equilibrium.

2.6 Why is there no equilibrium with $S_i = 0$ and $T > 0$?

A notable feature of our Nash results is the absence of a strategy that intuitively defects but uses deliberation to play TFT when faced with a repeated game. Why can't such a strategy be Nash? The answer is as follows. Unlike in 1-shot games, where it is always beneficial for an agent to defect no matter what the other agent does (because she always avoids paying the cost of cooperation c), the benefit of playing TFT in repeated games depends on coordinating with the other agent. Hence, when two intuitively defecting agents interact and play a repeated game, an agent that pays a cost to deliberate and thereby switch to TFT only benefits from doing so when her partner also deliberates (and thus also plays TFT). As a result, the returns from deliberative cooperation in repeated games for these agents depend not only on the benefit of mutual cooperation $b - c$ and the probability of repeated games p, but also on the probability that the other player deliberates.

Specifically, when two intuitive defectors $[0, T]$ and $[0, T']$ interact, the expected gain from deliberating for the first agent is $\frac{p(b-c)T'}{d}$, the product of the probability of there being a repeated game p, the benefit of mutual cooperation $b - c$, and the probability that the partner also deliberates $\frac{T'}{d}$. As a result, she should be willing to pay a maximum cost of deliberation $T^* = \frac{p(b-c)T'}{d}$ to get this benefit; and indeed, as we saw above, the best response to $[0, T']$ is $[0, \frac{p(b-c)T'}{d}]$. Thus, assuming that one's partner has $T' > 0$, there is always an incentive to deviate by deliberating less ($T^* < T'$). In other words, because of the coordination problem presented by cooperation in repeated games, any nonzero amount of deliberation T' among intuitive defectors is unstable and will be out-performed by intuitive defectors who engage in less deliberation. Therefore, the only equilibrium level of deliberation for a population of intuitive defectors is none at all ($T = 0$). (Or, as discussed above, if the maximum cost of deliberation d is sufficiently low, $d < p(b-c)$, then agents with $T' < d$ will instead be beaten by more deliberative agents with $T^* > T'$, resulting in the equilibrium where agents always deliberate and never use intuition.)

2.7 Nash calculations with assortment

We now consider a version of the game with assortment $a > 0$. In the context of population dynamics, assortment represents non-random mixing, such that with probability $(1 - a)$ a given agent plays with another agent selected at random from the population, whereas with probability a that agents plays with another agent having the same strategy as herself. To incorporate assortment in our Nash calculations, we therefore modify the Nash condition to be

$$\forall y : \pi(x, x) \geq ((1 - a)\pi(y, x) + a\pi(y, y).$$

We then solve for strategies that are best responses to themselves, in the manner described above. (Note that when $a = 0$, this is exactly equivalent to the above calculations.) Doing so finds that

the ID strategy remains the same when assortment is added ($T = 0$), but that the DC strategy now deliberates with $T = (1-p)(c-ba)$. Note that a consequence of this is that when $a = c/b$, the DC strategy reaches the boundary case of $[1, 0]$, such that $a \geq c/b$ implies no deliberation by DC, just intuitive cooperation (such that DC stops being an actual dual-process strategy).

3 Risk dominance calculations

3.1 Without assortment

Given that we have identified the game's two Nash equilibria, we are now interested in identifying when one equilibrium or the other will be favored by natural selection. For parameters where ID is the only Nash, it is clearly predicted that evolution will lead to ID. When DC becomes Nash, however, ID also remains Nash. Thus knowing when DC becomes Nash is not enough to know when selection will favor DC.

Risk-dominance, which is a stricter criterion than Nash, has been shown to answer this question: in symmetric $2x2$ games such as the one we study, when two symmetric equilibria exist, evolution will favor the risk-dominant equilibrium [1].

One Nash risk-dominates another Nash when the first Nash earns a higher expected payoff than the second Nash when there is a 50% chance of playing against either of the two strategies. Or, in population dynamic terms, the risk-dominant strategy profile is the one that fares better in a population where both are equally common.

We now ask when DC $[1, c(1-p)]$ risk-dominates ID $[0, 0]$ as a function of p. First, we consider the expected payoffs of these two strategy profiles against themselves and each other:

$$
\begin{aligned}
\pi(ID, ID) &= 0 \\
\pi(DC, ID) &= -c(1-p)(1 - \frac{(1-p)c}{d}) - \frac{(1-p)^2 c^2}{2d} \\
\pi(ID, DC) &= b(1-p)(1 - \frac{(1-p)c}{d}) \\
\pi(DC, DC) &= ((1-p)(b-c) + p(b-c))(\frac{(1-p)c}{d} - 1)^2 \\
&\quad - \frac{(1-p)^2 c^2 (.5(1-p)c - p(b-c))}{d^2} \\
&\quad + \frac{(1-p)c(-b(1-p) - p(b-c) + .5(1-p)c)(\frac{(1-p)c}{d} - 1)}{d} \\
&\quad - \frac{(1-p)c(-c(1-p) + p(b-c))(\frac{(1-p)c}{d} - 1)}{d}.
\end{aligned}
$$

DC risk-dominates ID when

$$\frac{1}{2}\pi(DC, ID) + \frac{1}{2}\pi(DC, DC) > \frac{1}{2}\pi(ID, ID) + \frac{1}{2}\pi(ID, DC).$$

Solving for p in the above equation yields the following condition:

$$p > \frac{\frac{1}{2}(2c^2 - bd - cd + \sqrt{d}\sqrt{-4bc^2 + 4c^3 + b^2d + 2bcd + c^2d})}{c^2}.$$

As we will see below, this value of p successfully captures the transition point we observe in evolutionary dynamics from a population of all ID players to a population of all DC players. For example, for $b = 4$, $c = 1$, and $d = 1$, DC begins to risk dominate ID when $p > .30$ (see Figure 2 of main text). Some other values, which we explore in steady state analyses below, include the following:

b	c	d	p at which DC risk-dominates ID
2	1	1	.62
8	1	1	.14
4	2	1	.50
4	.5	1	.19
4	1	.75	.25
4	1	2	.36

3.2 With assortment

When including assortment $a > 0$, the risk dominance condition for DC becomes

$$a\pi(DC, DC) + (1-a)(\frac{1}{2}\pi(DC, ID) + \frac{1}{2}\pi(DC, DC)) > a\pi(ID, ID) + (1-a)(\frac{1}{2}\pi(ID, ID) + \frac{1}{2}\pi(ID, DC))$$

with the DC agent's deliberation threshold now being $T = (1-p)(c - ba)$ (as shown above in the Nash calculations with assortment).

Thus, the minimum a value at which DC comes to risk dominate ID is given by

$$a > \frac{1}{2(b^2 - 2pb^2 + p^2b^2)} (2bc - 4pbc + 2p^2bc - 2bd + pbd + pcd + \sqrt{(-2bc + 4pbc - 2p^2bc + 2bd + pbd + pcd)^2 - 4(b^2 - 2pb^2 + p^2b^2)(c^2 - 2pc^2 + p^2c^2 + pbd - 2cd + pcd)}).$$

4 Evolutionary dynamics

4.1 Basic setup

We now turn from Nash calculations to evolutionary dynamics. We study the transmission of strategies through an evolutionary process, which can be interpreted either as genetic evolution or as social learning. In both cases, strategies that earn higher payoffs are more likely to spread in the

population, while lower payoff strategies tend to die out. Novel strategies are introduced by mutation in the case of genetic evolution or innovation and experimentation in the case of social learning.

We study a population of N agents evolving via a frequency dependent Moran process with an exponential payoff function [2]. In each generation, one agent is randomly selected to change strategy. With probability u, a mutation occurs and the agent picks a new strategy at random. With probability $(1-u)$, the agent adopts the strategy of another agent j, who is selected from the population with probability proportional to $e^{w\varphi_j}$, where w is the intensity of selection and φ_j is the expected payoff of agent j when interacting with agents that have the same strategy with probability a, and interacting with agents picked at random from the population with probability $(1-a)$.

For ease of calculation, our main analyses focus on the limit of low mutation. Later, we also explore higher mutation rates using agent-based simulations, and demonstrate the robustness of our low mutation limit calculations.

4.2 Limit of low mutation calculation method

In the low mutation limit, a mutant either goes to fixation or dies out before another mutant appears. Thus, the population makes transitions between homogeneous states, where all agents use the same strategy. Here the success of a given strategy depends on its ability to invade other strategies, and to resist invasion by other strategies. We use an exact numerical calculation to determine the average frequency of each strategy in the stationary distribution [3, 4, 5].

Let s_i be the frequency of strategy i, with a total of M strategies. We can then assemble a transition matrix between homogeneous states of the system. The transition probability from state i to state j is the product of the probability of a mutant of type j arising ($\frac{1}{M-1}$) and the fixation probability of a single mutant j in a population of i players, $\rho_{i,j}$. The probability of staying in state i is thus $1-\frac{1}{M-1}\sum_k \rho_{k,i}$, where $\rho_{i,i}=0$. This transition matrix can then be used to calculate the steady state frequency distribution s^* of strategies:

$$\begin{pmatrix} s_1^* \\ s_2^* \\ \vdots \\ s_M^* \end{pmatrix} = \begin{pmatrix} 1-\Sigma_j \frac{\rho_{j,1}}{M-1} & \frac{\rho_{1,2}}{M-1} & \cdots & \frac{\rho_{1,M}}{M-1} \\ \frac{\rho_{2,1}}{M-1} & 1-\Sigma_j \frac{\rho_{j,2}}{M-1} & \cdots & \frac{\rho_{2,M}}{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\rho_{M,1}}{M-1} & \frac{\rho_{M,2}}{M-1} & \cdots & 1-\Sigma_j \frac{\rho_{j,M}}{M-1} \end{pmatrix} \begin{pmatrix} s_1^* \\ s_2^* \\ \vdots \\ s_M^* \end{pmatrix}$$

The eigenvector corresponding to the largest eigenvalue (1) of this matrix gives the steady state distribution of the stochastic process.

Note that this method requires discretizing the strategy space, such that there is some finite number of strategies M that agents can select. We consider a strategy space in which: (i) agents' cooperation strategies S_i, S_1, and S_r are limited to being either 0 (never play the cooperative strategy) or 1 (always play the cooperative strategy); and (ii) agents' maximum cost of deliberation T ($0 \leq T \leq d$) that they are willing to pay in order to deliberate is rounded to the nearest $\frac{d}{10}$ (so T is selected from the set $\{0, d/10, 2d/10, \ldots, d\}$). Thus, the strategy space consists of a total of $2*2*2*11 = 88$

strategies.

Using the Moran process, the fixation probability $\rho_{B,A}$ (the probability that a single A mutant introduced into a population of B-players will take over) is calculated according to an exponential fitness function. In a population of i A-players and $N-i$ B-players, the fitness of an A-player f_i and B-player g_i are defined as

$$\begin{aligned} f_i &= e^{w(a\pi(A,A)+(1-a)(\frac{i-1}{N-1}\pi(A,A)+\frac{N-i}{N-1}\pi(A,B)))} \\ g_i &= e^{w(a\pi(B,B)+(1-a)(\frac{i}{N-1}\pi(B,A)+\frac{N-i-1}{N-1}\pi(B,B)))} \end{aligned}$$

where $\pi(A,A)$ is the expected payoff of an A-player against an A-player, $\pi(A,B)$ is the expected payoff of an A-player against a B-player, etc.

The fixation probability of a single A-player in a population of B-players can then be calculated as follows:

$$\rho_{B,A} = \frac{1}{1+\sum_{k=1}^{N-1}\prod_{i=1}^{k}\frac{g_i}{f_i}}$$

The calculations presented in the main text numerically evaluate this expression for each strategy pair and then solve for the steady state distribution according to the procedure described above. As shown in Figure S1, these evolutionary calculations are in quantitative agreement with the risk-dominance calculations across p and a values shown in the main text Figure 3.

5 Robustness of evolutionary results

5.1 Robustness to parameter variation

Figure S2 shows results of evolutionary steady state calculations for various parameter sets. In each case, we see a qualitatively equivalent pattern to what is observed in the main text Fig 2a: the steady state transitions from intuitive defection $S_i = 0$ with little deliberation $T \approx 0$ when p is small, to intuitive cooperation $S_i = 1$ with substantial deliberation $T >> 0$ implementing cooperation in repeated games $S_r = 1$ and defection in 1-shot games $S_1 = 0$ when p is sufficiently large. Then, as p increases further, the steady state value of T decreases. (Note that for some parameter values (e.g. Figure S2 panels b, c, and e), when DC becomes risk-dominant, the equilibrium level of T is close to 1 such that agents almost never use intuition, and therefore initially there is little selection pressure on S_i, leading to $S_i \approx 0.5$.)

Quantitatively, the transition from intuitive defection and non-deliberation to intuitive cooperation and deliberation occurs at precisely the value of p where DC begins to risk-dominate ID; and after this point the average value of T matches that of DC, with $T = (1-p)c$. Thus, these evolutionary calculations show the power of the Nash calculations for characterizing the behavior of our system.

5.2 Robustness to higher mutation rates and probabilistic strategies

We now compare the results of the steady state calculations presented in the main text with agent-based simulations. These simulations use exactly the same evolutionary process as the calculations described above, but relax two simplifying assumptions made in the calculations: the simulations (i) allow agents' probabilities of playing cooperative strategies S_i, S_1 and S_r to take on any value on the interval $[0, 1]$, instead of only allowing 0 or 1 as in the calculations; allow agents' deliberation threshold T to take on any value on the interval $[0, d]$, instead of only allowing discrete values in steps of $d/10$, and (iii) relax the calculation's assumption of vanishingly small mutation and instead use a relatively high mutation rate of $u = 0.05$. For each set of parameters, we conduct 10 simulation runs, each of which lasts 10^7 generations. We then show the value of each of the 4 strategy variables S_i, S_1, S_r, and T, averaged over all generations of all 10 simulation runs (Figure S3 symbols). For comparison, we also show the low mutation limit calculation results (Figure S3 lines). Critically, Figure S3 shows that these agent-based simulations produce very similar results to the calculations. This demonstrates the validity of the calculation, despite its simplifying assumptions.

6 Generalized coordination game analysis

6.1 Setup

The key idea underlying our model is that cooperation sometimes involves a social dilemma (e.g. the 1-shot PD), but other times involves coordination. In our main model, we focus on the infinitely repeated PD as our example of coordination. Doing so, we find that there are two main strategies that can be Nash in this setup: (i) a strategy that intuitively cooperates and sometimes deliberates when the cost of deliberation is less than $T = c(1 - p)$,and (ii) a strategy that always intuitively defects and never deliberates. (We also find that when the maximum possible cost of deliberation d is especially low ($d \leq p(b - c), c(1 - p)$), agents who always deliberate and never use intuition ($T = d$) can also be Nash.)

Here, we demonstrate that these basic results extend to cooperative interactions that involve coordination more generally, rather than being specific to infinitely repeated PDs. To do so, we consider a game where with probability $1 - p$ agents play the 1-shot PD defined above, and with probability p they play a coordination game with the following payoff structure:

	Cooperate	**Defect**
Cooperate	$A + B$	$A - C$
Defect	$A + B - D$	A

where $A, C \geq 0$ and $B, D > 0$.

This payoff structure has the following features. First, it captures the essence of coordination problems, which is that you cannot improve your payoff by playing something different from the other person (the penality of not coordinating when the partner defects is captured by $C \geq 0$, and when the other person cooperates by $D \geq 0$). As we are interested in *cooperative* coordination problems, we introduce two additional features: that the cooperative equlibrium is more efficient (higher payoff) than the non-cooperative equilbrium, captured by $B > 0$; and that it requires coordination to achieve the full benefits of this cooperation, such that defecting when the partner cooperates leads to a strictly lower payoff than cooperating when the partner cooperates, $D > 0$ (rather the more

general coordination requirement of just $D \geq 0$). Note that this payoff structure reduces to the infinitely repeated PD using $A = 0$, $B = b - c$, $C = 0$, $D = b - c$.

Using this much more general specification of cooperative coordination problems, we perform a Nash analysis and ask whether (i) we continue to observe the dual process, intuitively cooperative strategy profile that we found using the repeated PD, and whether (ii) an intuitively defecting Nash that sometimes deliberates, which was not observed using the repeated PD, can occur here. We use the same approach described above, in which we focus our Nash analysis on strategies with $S_r = 1$, $S_1 = 0$, and S_i either 0 or 1.

6.2 Intuitively cooperating equilibria

As we did for the repeated PD model, we calculate the best response deliberation threshold with T_b for an intuitively cooperating agent playing against an intuitively cooperating agent with deliberation threshold T'. We find, as before, that the best response is $T_b = c(1-p)$, regardless of the value of T' (or any of the coordination game parameters). To determine when this strategy $[1, c(1-p)]$ is Nash, we next consider under what conditions an intuitively defecting agent could beat it. To do so, we find the best response intuitively defecting strategy against $[1, c(1-p)]$, which we find to be $[0, pD]$ (note that this matches the result from repeated PD model, where the best response was $[0, p(b-c)]$). We find that $\pi([1, c(1-p)], [1, c(1-p)]) \geq \pi([p, pD], [1, c(1-p)])$, such that $[1, c(1-p)]$ is Nash, when $p \geq \frac{c}{c+D}$ and $d > c(1-p)$. (Note, again, that this matches the results from the repeated PD model in which the Dual-process Cooperator was an equilibrium when $p \geq c/b$ and $d > c(1-p)$.)

6.3 Purely deliberative equilibrium

Next, we consider the boundary case that always deliberates, $[-, d]$. Analogous to the results for the repeated PD version, we find that $[-, d]$ is Nash when $d \leq pD$ and $d \leq c(1-p)$ (as the best response strategy when these conditions are met has $T > d$).

6.4 Intuitively defecting equilibria

Finally, we consider the intuitively defecting case. Unlike in the repeated PD model, we now find that there are two possible intuitively defecting equilibria.

We begin by considering the boundary case $[0, 0]$. We find that the best responding intuitive defector against $[0, 0]$ is $[0, -Cp]$. Because T cannot be negative, this means that among the allowed values of T, $[0, 0]$ is the best response to itself (following the logic explained above in the repeated PD Nash calculations for the purely deliberative equilibrium). Moreover, we find that no intuitively cooperative strategy can ever do better against $[0, 0]$ than $[0, 0]$ does against itself. Thus, as in the repeated PD model, $[0, 0]$, a strategy that never deliberates and always intuitively defects, is always Nash.

Unlike in the repeated PD model, however, our general calculation of the best response deliberation threshold for an intuitive defector playing against $[0, T']$ gives

$$T_b = \frac{p(C+D)}{d}T' - Cp,$$

such that the intuitive defector strategy that best responds to itself is given by $[0, \frac{cDp}{(C+D)p-d}]$. We find that if the maximum cost of deliberation is sufficiently small, such that $d < pD$ and either $d \leq c(1-p)$ or $0 < C \leq \frac{c(1-p)(Dp-d)}{p(d-c(1-p))}$ are satisfied, then this strategy (which has $S_i = 0$ and $T > 0$) is Nash.

Thus, unlike in the repeated PD model, it *is* therefore possible to have a Nash equilibirium that intuitively defects but sometimes uses deliberation to cooperate in the coordination game. Critically, however, this strategy can never be risk-dominant, and therefore is never favored by selection! Whenever $[0, \frac{cDp}{(C+D)p-d}]$ is Nash, there are always two other strategies which are Nash, and both of these other strategies always risk-dominate $[0, \frac{cDp}{(C+D)p-d}]$: $[0, 0]$ is always Nash; when $d < pD$ and $d \leq c(1-p)$, the purely deliberative strategy $[-, d]$ is also Nash; when $d < pD$ and $d > c(1-p)$ but $0 < C \leq \frac{c(1-p)(Dp-d)}{p(d-c(1-p))}$, then $[1, c(1-p)]$ is also Nash. Therefore, as in the repeated PD model, the more general coordination model finds that an intuitively defecting strategy that uses deliberation to cooperate when it is beneficial to do so can never be favored by selection.

6.5 Summary

In sum, the more general social dilemma versus coordination model we have analyzed provides two main conclusions.

1. If the maximum cost of deliberation d is sufficiently large, we observe precisely the same two equilibria observed in the simpler model: (i) an equilibrium that intuitively cooperates and sometimes deliberates $[1, c(1-p)]$, and (ii) an equilibrium that always intuitively defects and never deliberates $[0, 0]$.

2. If the maximum cost of deliberation d is smaller, more complicated equilibria can emerge, such as an equilibrium that intuitively defects and does sometimes deliberate. Crucially, however, this equilibrium is always risk-dominated by another equilibrium, and therefore will never be selected for.

Thus, the conclusions from the repeated PD model hold across all models where agents sometimes play 1-shot PD social dilemmas and other times play cooperative coordination games: selection can favor dual process cooperators, but not dual process defectors.

6.6 Application to repeated PD with finite continuation probability

Our main analyses used the average payoff per round from an infinitely repeated PD between TFT and ALLD for the game with reciprocity. Here, we use the generalized coordination game calculations above to show that these results extend to the more realistic case of total payoff in a repeated PD between TFT and ALLD where after every round, another round occurs with probability δ (such that on average there are $1/(1-\delta)$ rounds per game), yielding the payoff matrix

PD with continuation probability δ

	TFT	**ALLD**
TFT	$\frac{b-c}{1-\delta}$	$-c$
ALLD	b	0

where $b, c > 0$, $0 < \delta < 1$.

Thus, in terms of the generalized coordination game, this gives $A = 0$, $B = \frac{b-c}{1-\delta}$, $C = c$, $D = \frac{b-c}{1-\delta} - b$. Plugging in these values, we find that the DC strategy continues to be specified by $[1, c(1-p)]$, just as it was for the infinitely repeated PD, and that the condition for DC to be an equlibrium becomes $p \geq \frac{c}{\frac{b-c}{1-\delta} - b + c}$ and $d > c(1-p)$.

6.7 Application to general PD with reciprocal consequences

Finally, we use our generalized results to show that the conclusions of the main model, which used an infinitely repeated PD, extend to the general PD with reciprocity framework outlined in the main text. Here, with probability p, the PD payoff structure is modified such that when one player defects and the other cooperates, the defector's payoff is reduced by α and the cooperaters payoff is increased by β, yielding the payoff matrix

PD with Reciprocal Consequences

	C	**D**
C	$b - c$	$-c + \beta$
D	$b - \gamma$	0

where $b, c, \gamma, \beta > 0$.

In our main model, we focused on the case where $\gamma = b$ and $\beta = c$, yielding a payoff structure that is equivalent to average payoff per round of TFT and ALLD playing an infinitely repeated PD. Plugging this more general form into our results for the cooperative coordination game (using $A = 0, B = b - c, C = c - \beta, D = \gamma - c$), we find that the DC strategy continues to be specified by $[1, c(1-p)]$, just as it was for the infinitely repeated PD, and that the condition for DC to be an equlibrium becomes $p \geq \frac{c}{\gamma}$ and $d > c(1-p)$.

References

[1] Michihiro Kandori, George J Mailath, and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica: Journal of the Econometric Society*, pages 29–56, 1993.

[2] Tibor Antal, Martin A Nowak, and Arne Traulsen. Strategy abundance in 2× 2 games for arbitrary mutation rates. *Journal of Theoretical Biology*, 257(2):340–344, 2009.

[3] Drew Fudenberg and Lorens A Imhof. Imitation processes with small mutations. *Journal of Economic Theory*, 131:251–262, 2006.

[4] Lorens A Imhof, Drew Fudenberg, and Martin A Nowak. Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31):10797–10800, 2005.

[5] David G Rand and Martin A Nowak. The evolution of antisocial punishment in optional public goods games. *Nature Communications*, 2:434, 2011.

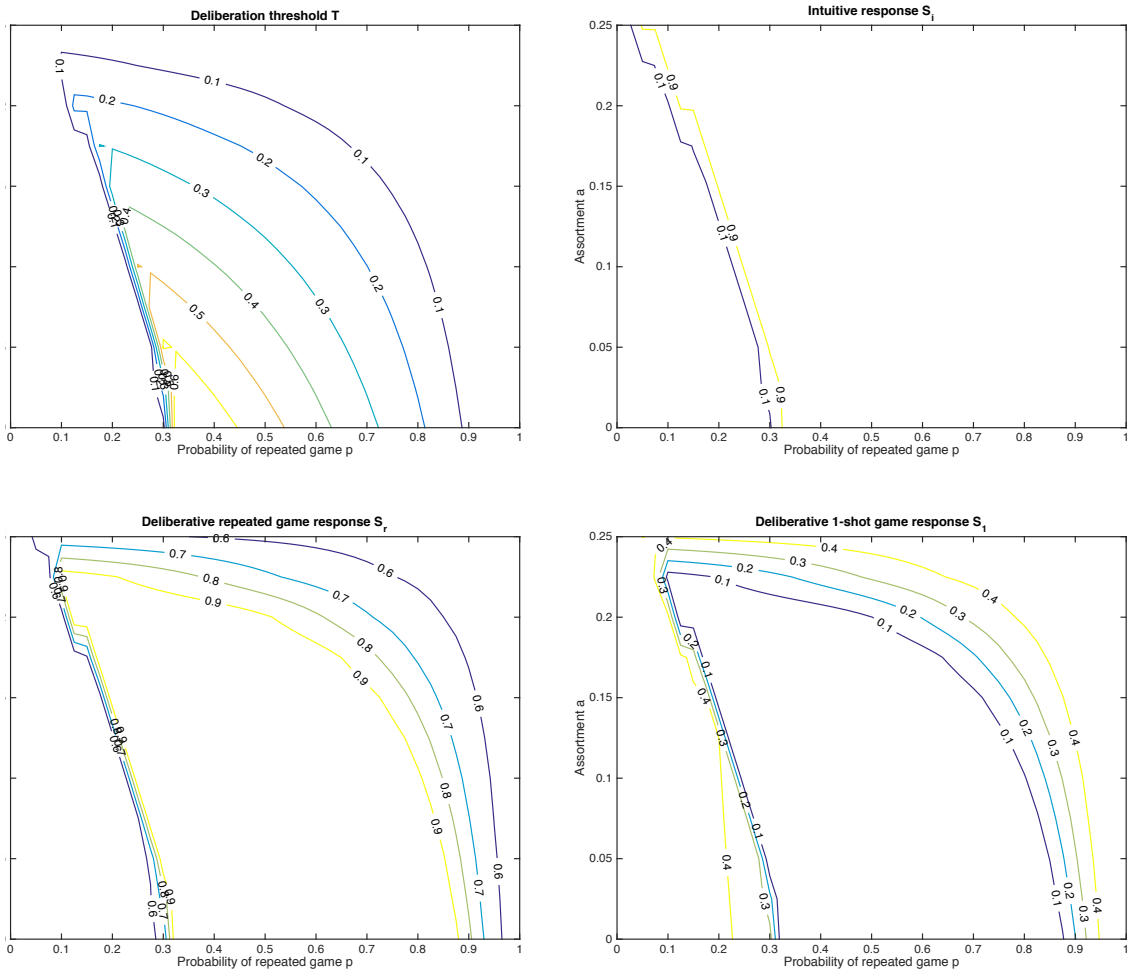

Figure S1: *Evolutionary calculations of the steady state distribution using* $N = 50$, $b = 4$, $c = 1$, $d = 1$, $w = 6$, *for various values of* p *and* a. *Shown are the average values of* T *(a),* S_i *(b),* S_r *(c), and* S_1 *(d). We see quantitative agreement with the risk-dominance calculations shown in the main text Figure 3:* S_i *is near 0 when ID is risk-dominant and near 1 when DC is risk-dominant;* T *is near 0 when ID is risk-dominant and equal to* $(c - ba)(1 - p)$ *when DC is risk-dominant; and* S_r *is near 1 while* S_1 *is near 0, except when* T *is close to zero such that there is little selection pressure on deliberative responses, leading neutral drift to pull* S_r *and* S_1 *toward 0.5.*

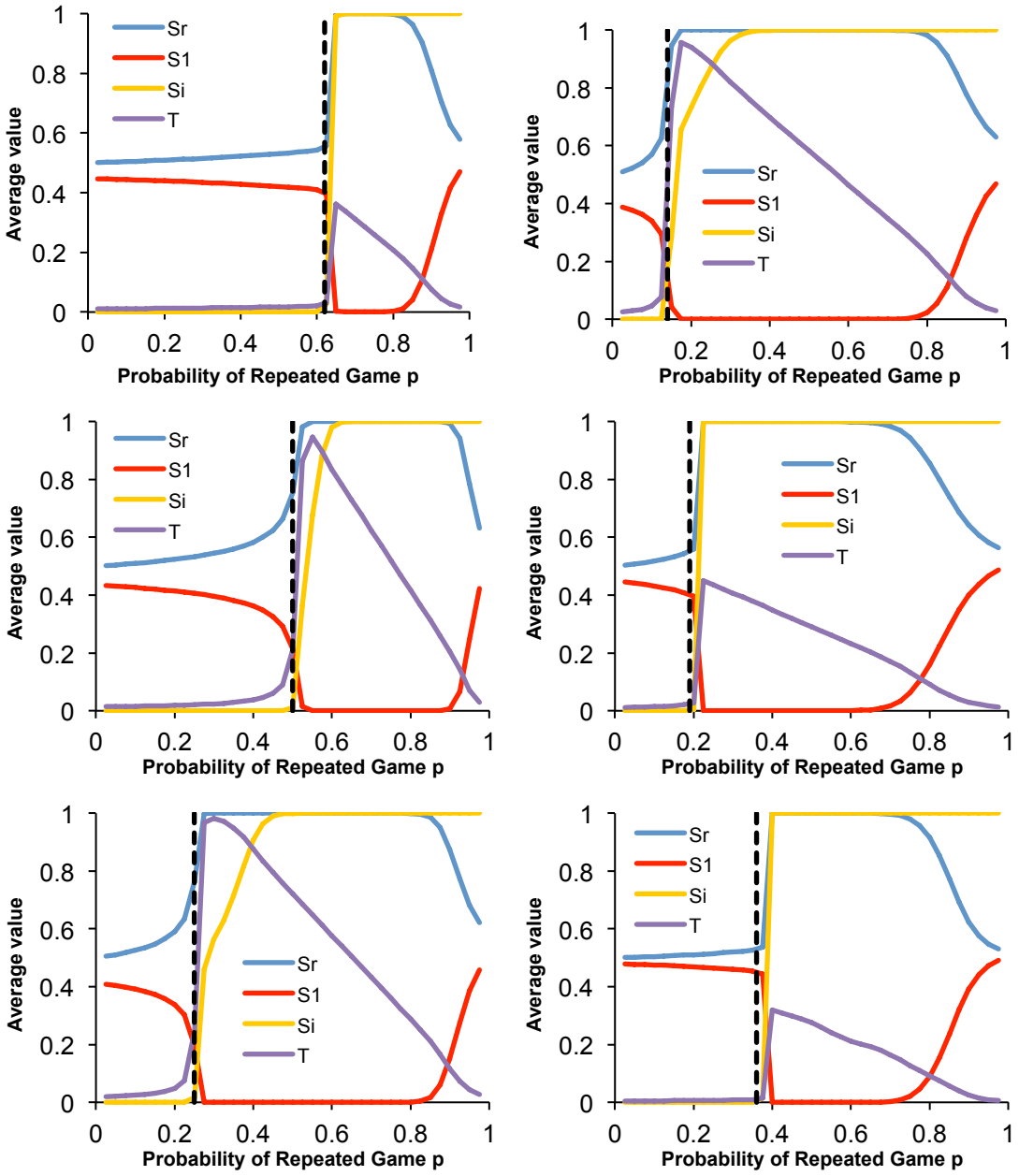

Figure S2: *Evolutionary calculations of the steady state distribution using* $N = 50$, $a = 0$ *and (a)* $b = 2$, $c = 1$, $d = 1$, $w = 6$; *(b)* $b = 8$, $c = 1$, $d = 1$, $w = 3$; *(c)* $b = 4$, $c = 2$, $d = 1$, $w = 5$; *(d)* $b = 4$, $c = .5$, $d = 1$, $w = 6$; *(e)* $b = 4$, $c = 1$, $d = .75$, $w = 5$; *(f)* $b = 4$, $c = 1$, $d = 2$, $w = 5$. *The point at which DC transitions to risk-dominating ID is presented as a dotted black line for comparison. (Note that because of our use of exponential fitness, for certain parameter sets a smaller selection strength* w *was needed to prevent the post-exponentiation fitnesses from exceeding MATLAB's computational limits.)*

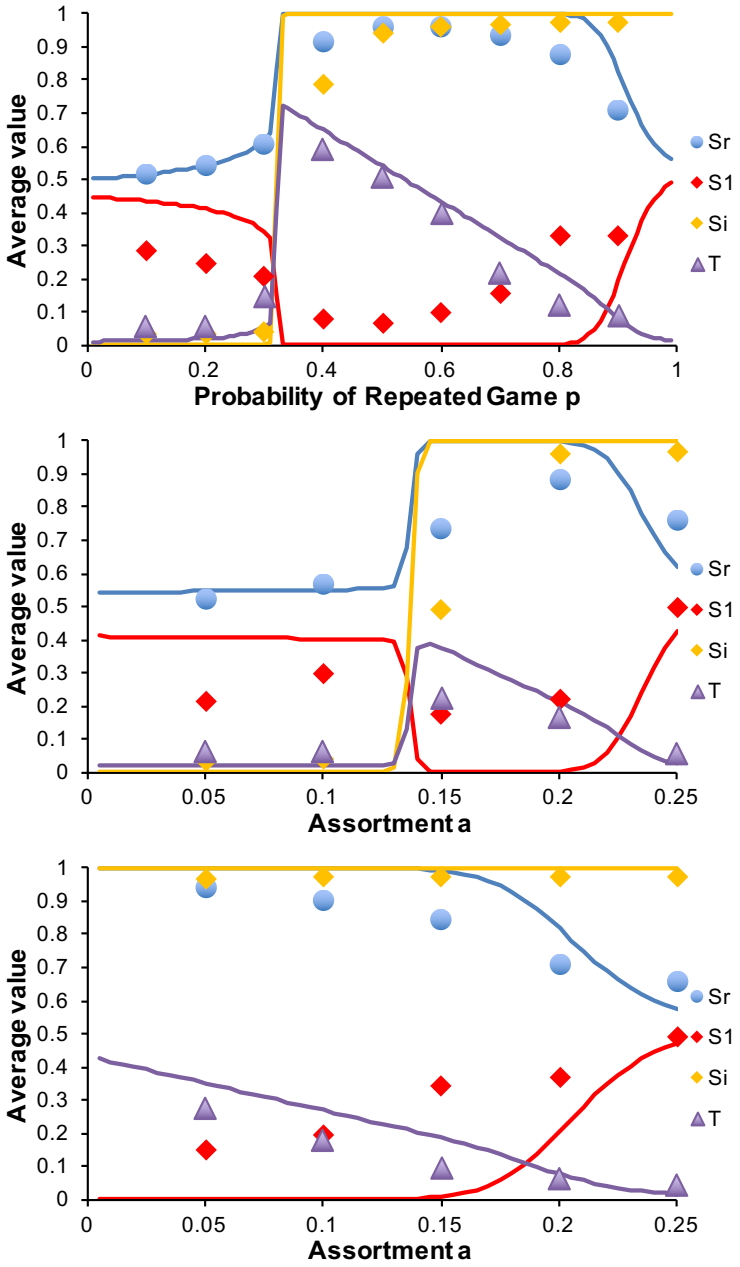

Figure S3: *Results of agent-based simulations (symbols) and steady state calculations (lines) showing the average value of each strategy variable, using* $N = 50$, $b = 4$, $c = 1$, $d = 1$, $w = 6$. *(a) Fixing* $a = 0$*; (b) fixing* $p = 0.2$*; (c) fixing* $p = 0.6$.