# Supplement

## Model description

This description follows the standardized ODD protocol for agent-based models (ABMs; Grimm et al. 2006).

*Purpose*

We developed an ABM to simulate a series of dyadic PD interactions. The aim of these models is to show that providing agents with the opportunity to mimic cooperation, silently or overtly, has implications for population distributions without drastically reducing overall cooperation.

*Parameters and schedule*

|  |  |  |
| --- | --- | --- |
| Variable | Type | Description |
| N | Global | Number of individuals in population; default = 100 |
| generations | Global | Number of generations per model run; default = 10e3 |
| µ | Global | Probability that an individual will adopt a new strategy on reproduction; default = 0.01 |
| f | Global | Sequence of rate of fluctuation of individual-level parameters on reproduction; default = -0.4 – 0.4 in increments of 0.1 |
| µs | Global | Probability that an individual will adopt or abandon a covert strategy on reproduction (model 3 only); default = 0.0001 |
| Conditional action expression (CAE) | Agent | Range of actions possible for an individual in dyadic interactions, includes cooperate and defect only across models; default = 50% each in population |
| Type | Agent | Range of possible agent-types, includes honest and mimic only across models; default = 50% each in population |
| Appearance | Agent | Range of possible agent action appearances, includes overt and silent only; default = 100% overt appearance in population |
| Potential to reproduce (PTR) | Agent | Probability that a given individual will reproduce; default = randomized normal distribution; mean = 0.5; standard deviation = 0.2 |
| Mimicry | Agent | An individual’s ability to mimic cooperative tendencies, where 0 is lowest and 1 is highest; default = randomized normal distribution; mean = 0.5; standard deviation = 0.2 |
| Sensitivity | Agent | An individual’s ability to detect mimics, where 0 is lowest and 1 is highest; default = randomized normal distribution; mean = 0.5; standard deviation = 0.2 |

**Table S1: Overview of agent- and global-level variables in model**

We represent time discreetly over generations; we do not consider space. At each generation, agents follow the commands described in the schedule.

**Figure S1: Schedule overview**

*Schedule overview*

1. Random dyads form
2. Agents play cooperate or defect
   1. In these models, an individual may switch from its default strategy depending on events described in Methods
3. Cheated strong reciprocators and mimics pay cost to punish (see Methods)
4. Adjust PTR, mimicry, and sensitivity sampling from f
5. Entire population reproduces with probability PTR
   1. Individuals inherit parent PTR, mimicry, and sensitivity after f adjustment

*Design concepts*

*Emergence —* the relationship between cost and risk of detection emerges from interactions between agents.

*Prediction* — agents have no memories and the interactions are considered as standalone events in a given cultural or genetic generation.

*Sensing* *vs signalling* — all agents can signal cooperative intent and are sensitive to mimicry

*Interaction* — dyadic interactions are formed randomly; mimics attempt to trick honest signallers into cooperating

*Stochasticity* — dyadic interactions are random; PTR is probabilistic

*Observation* — we report data per model over 10e4 runs

*Input* — we assume a range of scenarios starting with a simple strong reciprocity model, and then add in mimicry types and silent appearances to determine the effects of extra opportunities to defect against cooperators

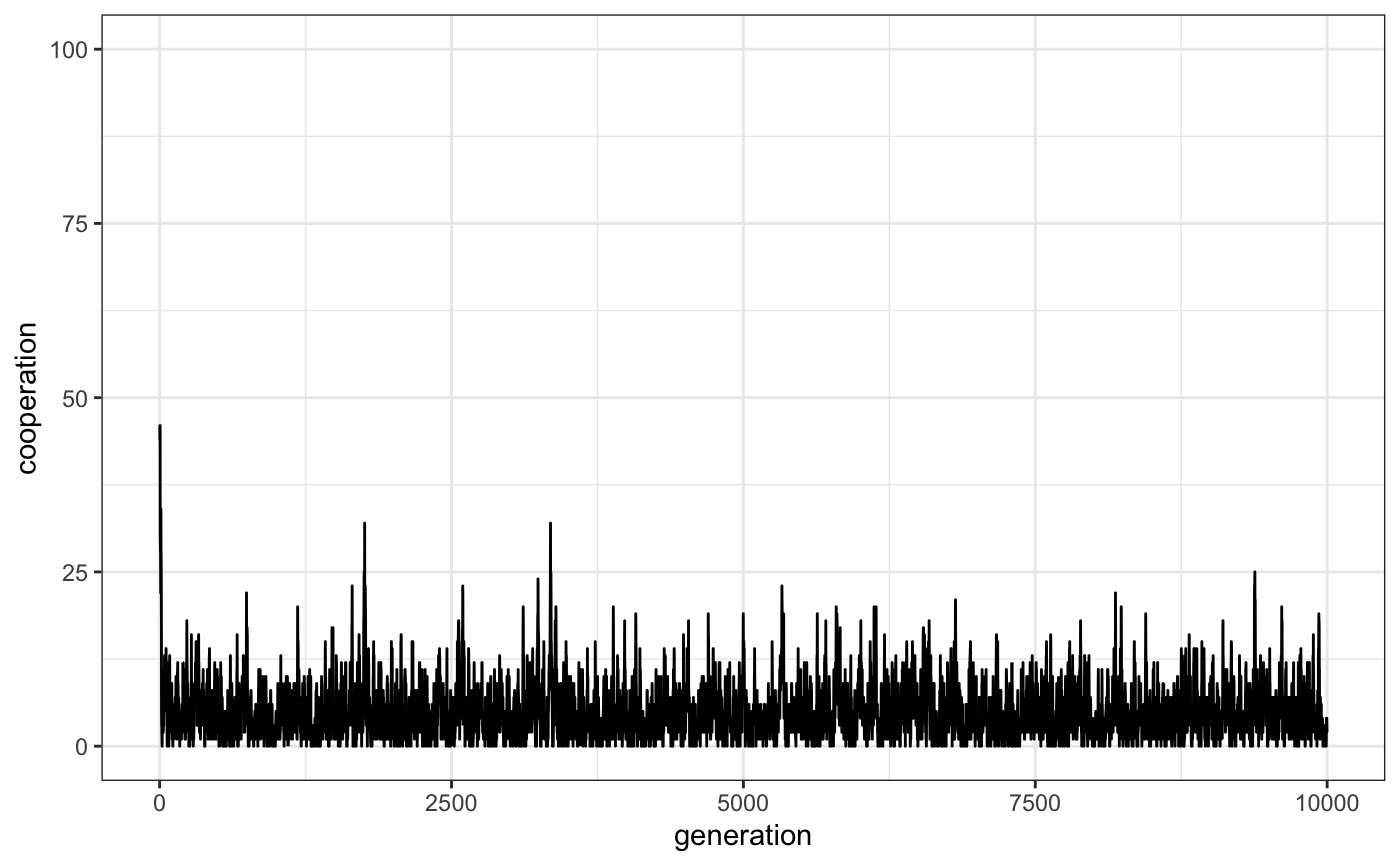
## Basic Prisoner’s Dilemma model

We initially ran a proof-of-concept basic PD model without punishment or automatic defection against defectors (strong reciprocity). The model showed, as expected, a very low rate of cooperation in the population where N = 100, generations = 10e3, µ = 0.01, ∂ = 0. The results are in figures **S1** and **S2**.

Chart

Description automatically generated

**Figure S1: the frequency of agents with cooperative strategy (green) vs agents with defection strategy (red) in basic, random dyadic pairing PD with no strong reciprocity.**

****

**Figure S2: percentage of cooperative behaviours in basic, random dyadic pairing PD with no strong reciprocity.**

## Model with probabilistic function for detection

In line with previous research (see main text), we investigated whether a probabilistic function with a mean detection rate of 65% (standard deviation = 0.1) would affect the model’s outcomes. In this case, the mimic’s *mimicry* score did not affect probability of detection, which was receiver-based only in the form of a *sensitivity* score. While cooperation was highly frequent in this case, there was regardless a consistent presence of mimics and covert defectors in the population, with no individual type or appearance in equilibrium (figures **S3**, **S4**, and **S5**).

Graphical user interface, chart, histogram

Description automatically generated with medium confidence

**Figure S3: percentage of cooperative behaviours in a covert mimicry model where receivers had a mean 65% chance of detecting mimics. Cooperation is maintained at a high degree, resembling a simple model with strong reciprocity (N = 100, generations = 10e3, µ = 0.01, µs = 0.001, ∂ = 0).**

**Bar chart

Description automatically generated**

**Figure S4: percentage of honest signallers (blue) vs mimics (orange) in a covert mimicry model where receivers had a mean 65% chance of detecting mimics. Neither type can prevent invasion over 10,000 generations.**

**Chart

Description automatically generated**

**Figure S5: percentage of overt appearance (green) vs covert appearance (black) in a covert mimicry model where receivers had a mean 65% chance of detecting mimics. Neither appearance can prevent invasion over 10,000 generations, and the covert appearance remains frequency-dependent for the majority of runs, punctuated by quickly going into fixation followed by invasion by the overt appearance.**