University of Zurich[UZH]
Institute of Biomedical Ethics

University of Notre Dame
Psychology Department

Population and Temptation Density Determine the Effect of Social Strategies on Moral Hypocrisy in a Virtual Society

Markus Christen,

University of Zurich & University of Notre Dame

Table of Contents

University of Zurich[UZH]
Institute of Biomedical Ethics

University of Notre Dame
Psychology Department

Moral Hypocrisy: The Question(s)

A definition of moral hypocrisy (Dan Batson):

"(…) avoid the cost of being moral while maintaining the appearance of morality (...)."

Issues that frame the concept of 'moral hypocrisy'

1) **A normative condemnation:** Moral hypocrisy is the wrong way to fill the fundamental "is-ought-gap" that morality implies.
2) **A social understanding of morality:** A safeguard of society from behaviors profitable for individuals but damaging for the group.
3) **Moral reputation as a core factor in morality:** Reputation is considered to be an essential component for the development of morality in human foragers, where each individual is strongly aware that he or she must have a positive reputation in case of future need, and painfully guards it (Hrdy 2009).
4) **Temptations as gains:** Moral behavior is (partially) understood to involve disadvantages for an individual in the sense of "missed opportunities".
5) **Cover is necessary:** Moral hypocrisy requires violating moral norms such that the violation is not detected (e.g. subtle cheating; Trivers 1971) – a factor that probably increases with higher social complexity.

Research questions with respect to moral hypocrisy

1) **Philosophy:** To what extend is moral hypocrisy a necessary part of human life (to fight the “terror of morality”)? How is moral hypocrisy related to the concept of morality one holds?
2) **Personality Psychology:** How can individuals maintain a motivational state with the ultimate goal to appear moral while, if possible, avoiding the costs to self of actually being moral (Batson et al. 1997).
3) **Social Psychology:** Why do individuals’ evaluations of their own moral transgressions often differ substantially from their evaluations of the same transgressions enacted by others (Valdesolo & DeSteno 2007)?
4) **Sociology:** What is the effect of social strategies intended to avoid moral hypocrisy on the prevalence of moral hypocrisy?

University of Zurich[UZH]
Institute of Biomedical Ethics
University of Notre Dame
Psychology Department

Conceptualization of Moral Hypocrisy in the Model

Agent states and behavior types

The model implements the conceptual idea of moral hypocrisy by distinguishing two different types of agent-states:

- The **reputation** of the agent (either morally good, G, or bad, B).
  - Model parameter p(r): set-up probability for good reputation (p(r) = [0,1])
- Its **disposition to act** toward temptations (either to be tempted, T, or to resist a temptation, R).
  - Model parameter p(t): set-up probability for being tempted (p(t) = [0,1])

This offers four different behaviors to the agents:

- Appearing good and resist a temptation (GR; “good guys”, blue)
- Appearing good but being tempted (GT; “hypocrites”, yellow)
- Appearing bad and being tempted (BT; “bad guys”, red)
- Appearing bad but resisting temptations (BR; “inconsistent guys”, pink).

Moral hypocrisy as the "most rewarding behavior"

The payoff structure represents the basic idea of moral hypocrisy, i.e. an agent gains most if he takes the bait: the model assumes that moral hypocrisy is the optimal behavior for a single agent within a society.

		Disposition to act	
		Be tempted (T	Resist temptation (R)
Reputation	Good (G):	GT (yellow): one point for each temptation and for each neighbor	GR (blue): one point for each neighbor
	Bad (B)	BT (red): one point for each temptation	BR (pink): 0

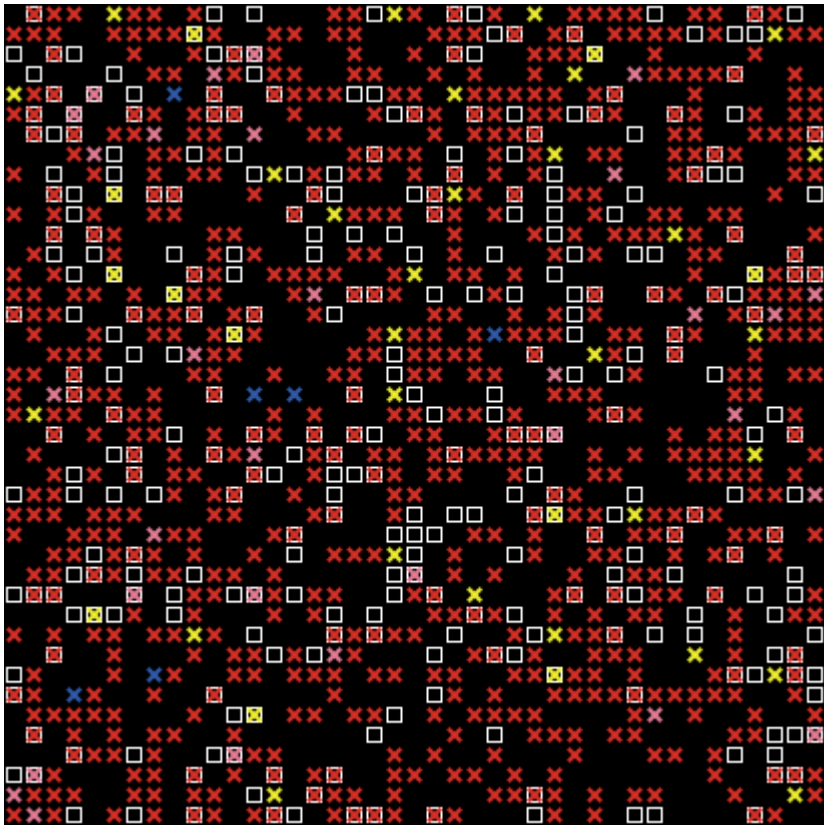
The goal of this study **is to assess the success of different strategies compared to a benchmark** (no strategy installed) in terms of changes in the population distribution of agents that follow one of the four behaviors.

Social Strategies to overcome moral hypocrisy

#	Description of Strategy
1	Avoid agents that are tempted: Every agent that has either a yellow or red neighbor moves to the closest free cell on the lattice without such neighbors (if possible).
2	Seek agents with good reputation: Every agent that does not yet have either a blue or yellow neighbor moves to the closest free cell on the lattice with at least one such neighbor (if possible).
3	Disclose hypocrite (local version): Whenever the majority of agents in a two-degree Moore neighborhood of a yellow agent is non-yellow, the yellow agent changes his behavior to BT (red).
4	Disclose hypocrite (global version): Whenever the majority of agents of a specified yellow agent is non-yellow, the yellow agent changes his behavior to BT (red).
5	First strategy 2, then strategy 1
6	First strategy 1, then strategy 3
7	First strategy 1, then strategy 4
8	First strategy 2, then strategy 3
9	First strategy 2, then strategy 4
10	First strategy 3, then strategy 2, then strategy 1
11	First strategy 4, then strategy 2, then strategy 1

How the model works

a) It randomly distributes agents and temptations and assigns former an initial behaviors according to p(r) and p(t).
b) It selects an agent X
c) It calculates the payoff of X
d) It does a) and b) for each agent chosen in a random order
e) It changes the behavior of each agent to the behavior of its best performing neighbor
f) It applies the strategy (1-11), checks whether stop condition g) applies and if not goes back to b).
g) It stops when the model reaches a quasi-stable state (no significant population size changes)



Model Parameters

- Set-up-probability of having a good reputation p(r) - Set-up-probability of being tempted if a temptation is present p(t) - Strategies (1 to 11) - Spatial distribution of temptations and agents - # of agents/temptations for fixed p(r) and p(t)	**Full sampling**
- Population Density - Temptation Density - Sequencing of single steps within (complex) strategies - Dynamic change of strategies (exploratory)	**Pre-Test**
- Neighborhood for strategy comparison - Hypocrite disclosure majorities - Payoffs - Agent-temptation interaction	**No changes**

University of Zurich[UZH]
Institute of Biomedical Ethics
University of Notre Dame
Psychology Department

Pre-Tests and Determination of Main Scenarios

Paradigmatic scenarios – Description

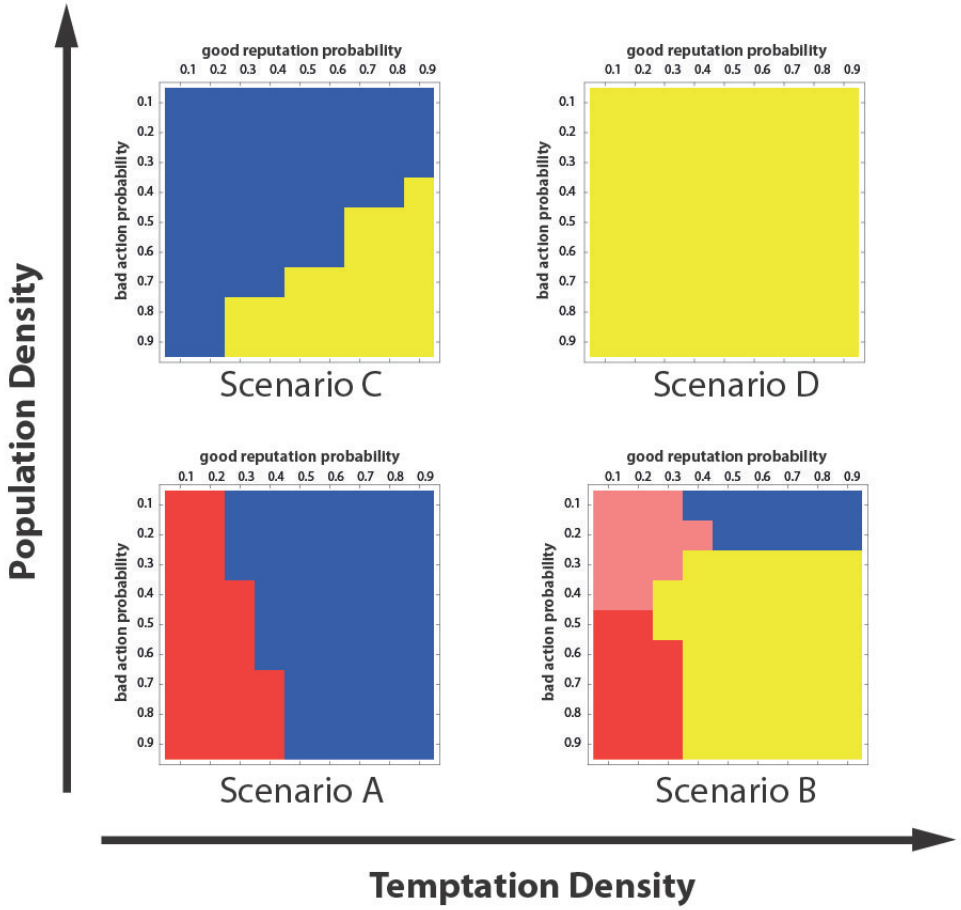
Scenario A – Pre-Modern: Low population (10%) and low temptation density (5%). Pre-modern societies consist of small groups with high social control minimizing the number of available temptations.

Scenario B – Modern Agricultural: Low population (10%) and high temptation density (50%). Modern agriculture consists of large farms (low population density) that have access to all means of modern societies in terms of mobility, communication etc. that increase the "temptation space".

Scenario C – Brave New World: High population (66%) and low temptation density (5%). A city state (Singapore?) with a tight control regime with respect to temptations.

Scenario D – Sin City: High population (66%) and high temptation density (50%). It implements the idea of a densely populated city full of temptations (Charlotte?).

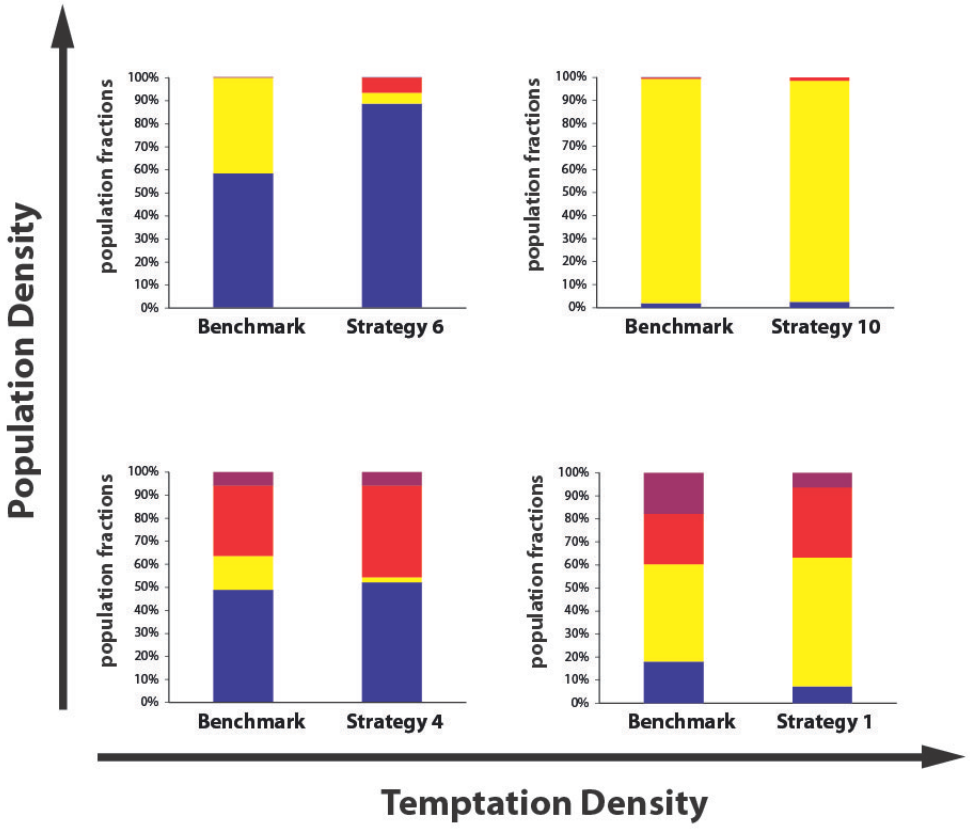
Paradigmatic scenarios – Majorities (Benchmark)



University of Zurich[UZH]
Institute of Biomedical Ethics
University of Notre Dame
Psychology Department

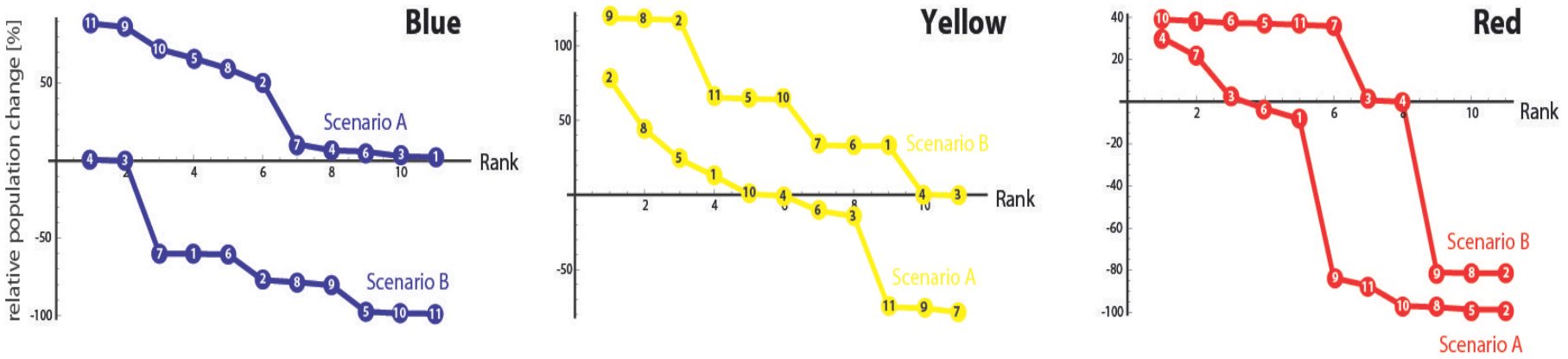
Results 1: separate scenarios

Result 1-A: *Scenario parameters are the main determinants of population distributions (no strategy induced majority change)*

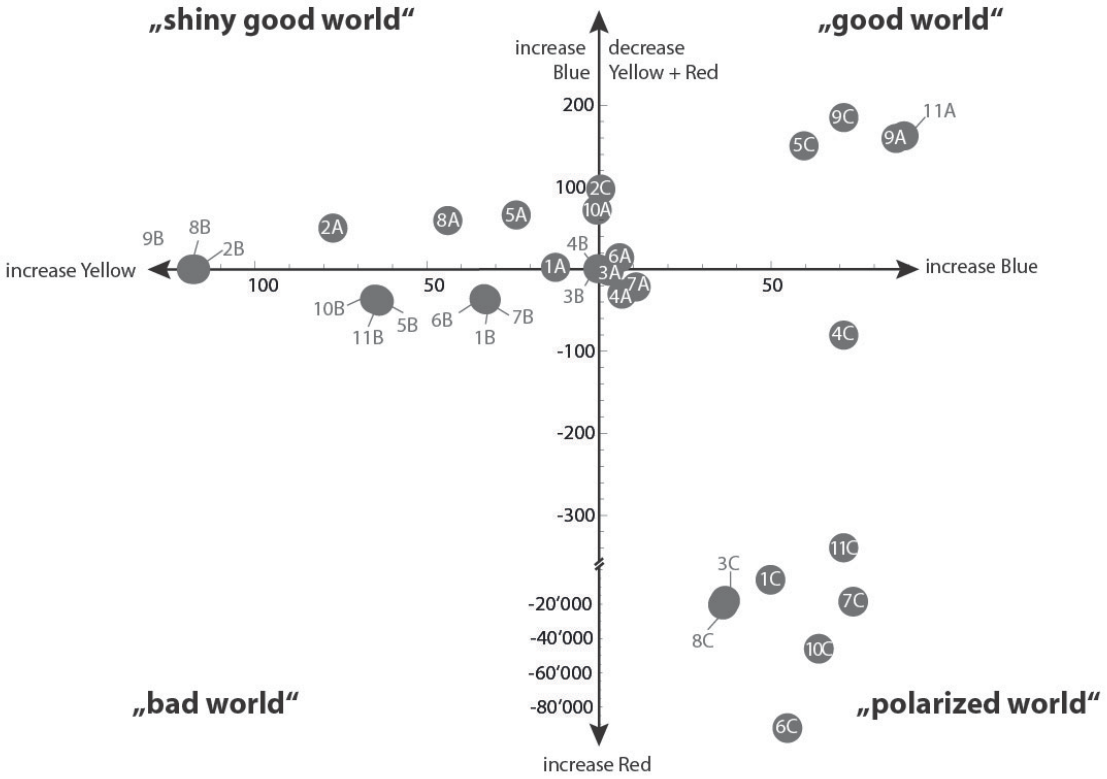


Result 1-B: *Strategy rankings reveal conflicting effects of interventions*

We ranked the strategies according to their ability to increase the population of a specific behavior type relative to the benchmark population size. We display the two sequences for each population with the highest dissimilarity measured by the Kendall Rank correlation (a measure for the similarity of rankings).



Result 1-C: *Strategy effects can be attributed to four "moral worlds"*



How scenarios frame the effect of interventions

- The Modern Agricultural scenario creates a context that promotes bad world strategies – i.e. red and yellow can often increase their weight.
- The Brave New World scenario creates a context that promotes polarizing world strategies increasing both the blue and the red population.
- The Pre-Modern scenario creates a context that promotes shiny good world strategies.

Effect of strategies:

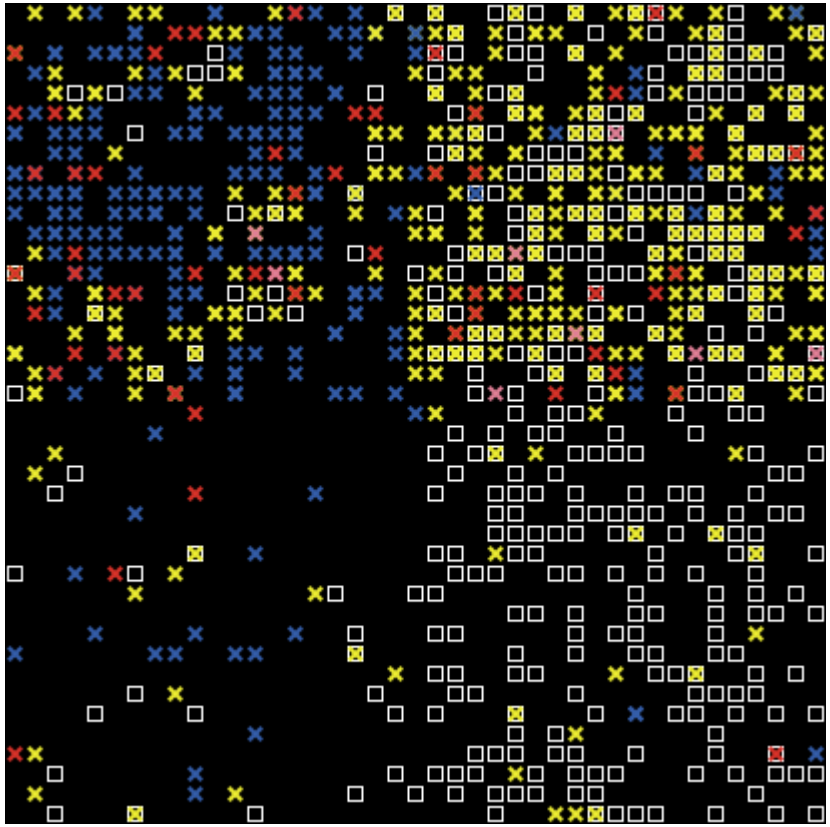
- Disclosing strategies in their pure form (3 and 4) tend to be polarizing, i.e. form strong minorities of red agents.
- Strategy 2 tends to be “shiny”, which is plausible as the yellow population profits from a strategy that benefits good reputation.

University of Zurich[UZH]
Institute of Biomedical Ethics
University of Notre Dame
Psychology Department

Results 2: combined scenarios

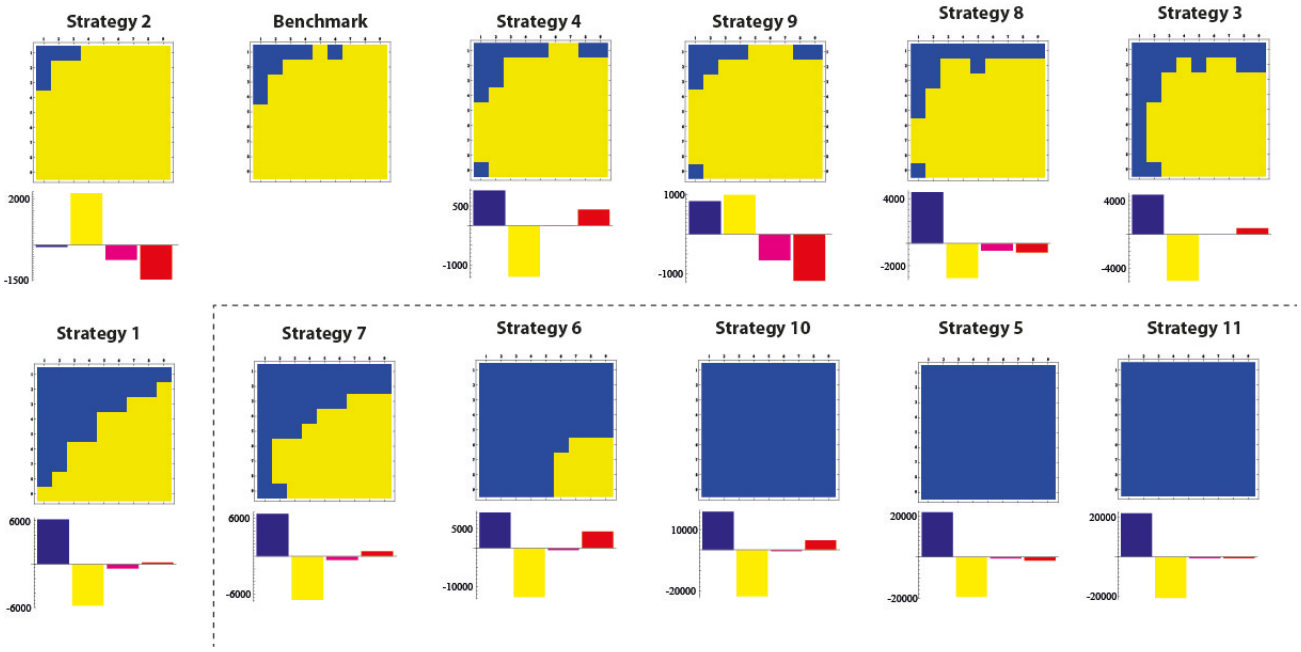
Including all four basic scenarios in one model

The model now has spatial structure with respect to the probability of distributing agents and temptations such that the four main scenarios are present simultaneously.



Result 2-A: *Scenario-diversity allows to defeat moral hypocrisy for some strategies (majority change)*

Result 2-B: *There is a pronounced but not consistent local-global effect in hypocrite disclosure*



University of Zurich[UZH]
Institute of Biomedical Ethics
University of Notre Dame
Psychology Department

Results 3: Population uniformity change

A measure for population uniformity

Calculate for each setting of reputation and temptation probability a population uniformity measure $U(x_1, x_2, x_3, x_4)$ such that (x_i: relative population size):

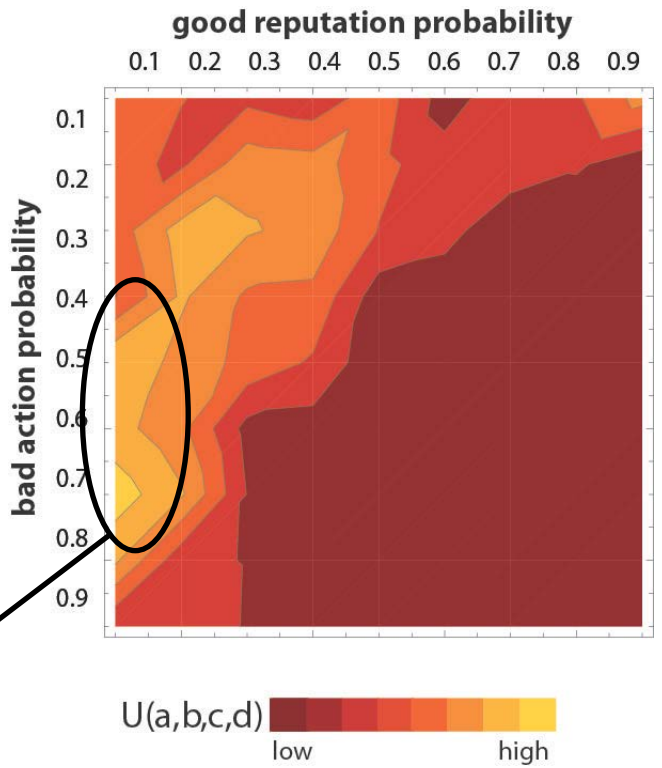
1) $U(0.25, 0.25, 0.25, 0.25) = 0$
2) $U(1, 0, 0, 0) = U(0, 1, 0, 0) = etc. = 1$
3) $U(0.5, 0.5, 0, 0) = U(0, 0, 0.5, 0.5) = etc.\ (Symmetry)$

This leads to:

$$U(x_1, x_2, x_3, x_4) = \frac{1}{6} \sum_{i,j=1}^{4} (x_i - xj)^2$$

Calculate $\Delta U = 2[if\ majority\ change] * \|U[benchmark] - U[strategy]\|$ for each strategy and all p(r), p(t) and calculate the mean over all strategies in order to identify regions in the p(r)-p(t)-space that are sensible for large population changes due to strategy interventions.

Result 3 (mean ΔU for all strategies): *The effect of strategies is not uniform in p(r)-p(t)-space, but depends in particular on a low p(r) – but has also a local maxima for high p(r) and low p(t) (for some strategies).*



Hold p(r) and p(t) fixed and vary #agents and #temptations.

University of Zurich[UZH]
Institute of Biomedical Ethics
University of Notre Dame
Psychology Department

Results 4: Effect of varying population and temptation numbers

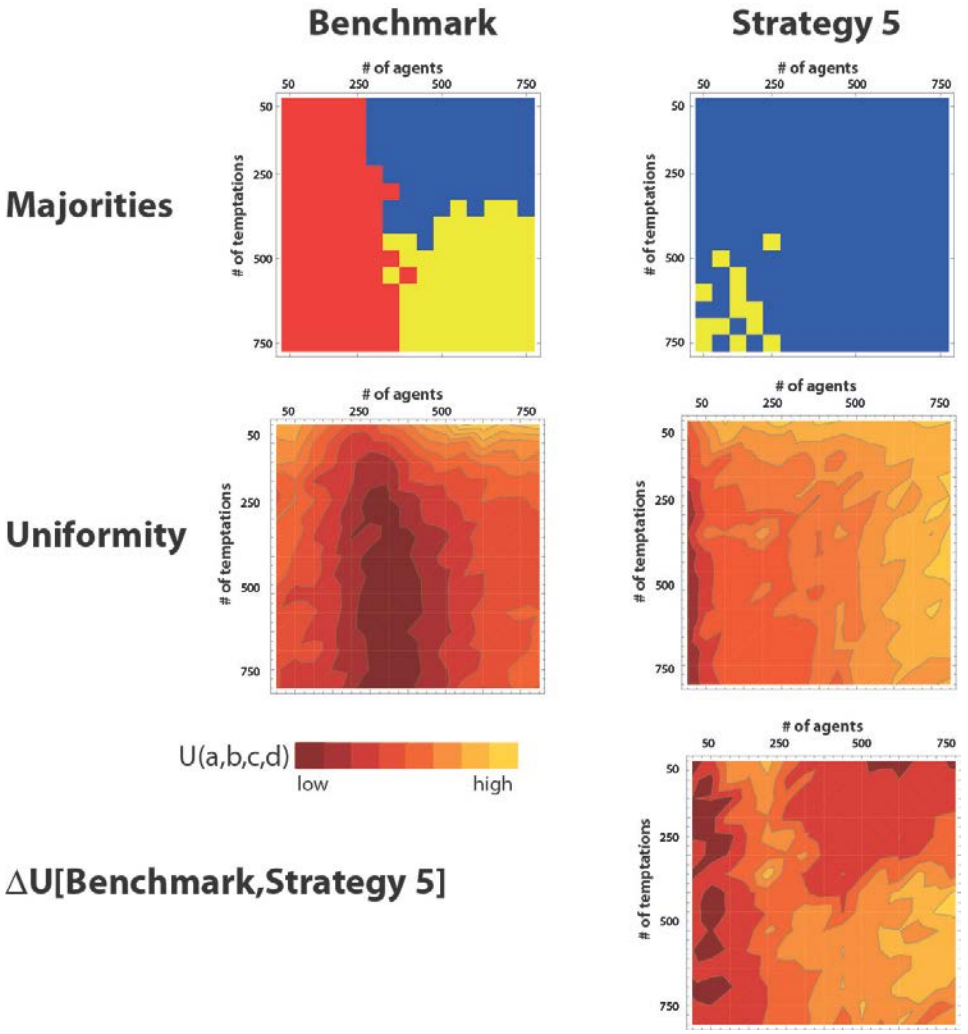
Example: Strategy 5 versus Benchmark

Reasoning:

- We use the model setting that includes all four basic scenarios.
- We use strategy 5 where a global majority change can be observed.
- We use the setting for p(r) and p(t) that has the highest ΔU value for strategy 5 in the extended model (i.e. this set-up condition is most sensible for large changes in population numbers.

Question: How does the population uniformity depend on #agents and #temptations?

Result 4: ***The success of strategy 5 relies in the fact that it can defeat hypocrisy even when many agents and temptations are present (i.e. the strategy compensates for the additional gain the presence of temptations could allow)***



University of Zurich[UZH]
Institute of Biomedical Ethics
University of Notre Dame
Psychology Department

Exploratory result: Dynamic strategy change

Analyze dynamic strategy change

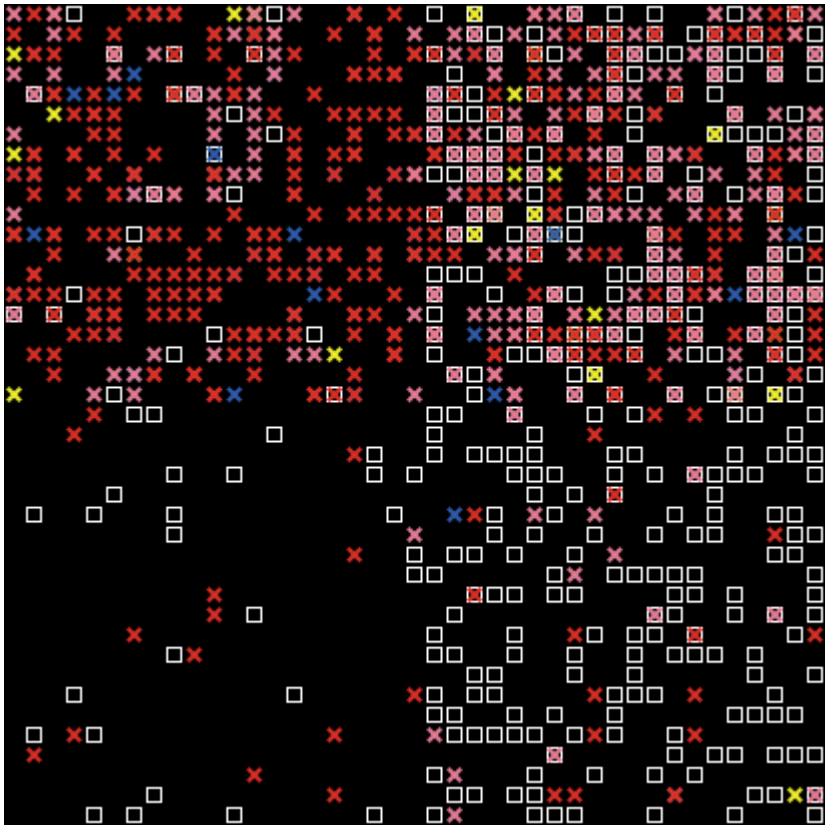
Simulate a “policy change” within a single run, i.e. change basic strategies (1,2, 3, 4) and their combinations.

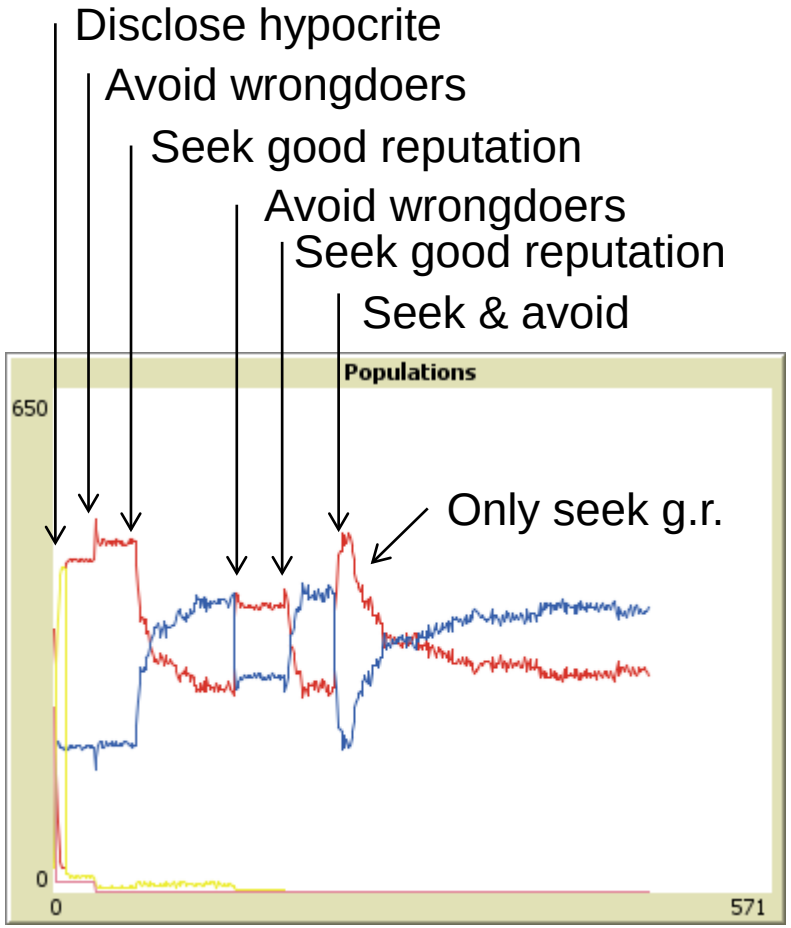
Do this for:

- Complex model (i.e. all four scenarios)
- Low p(r), medium p(t)

No systematic testing.

Dynamics – preliminary example





University of Zurich[UZH]
Institute of Biomedical Ethics
University of Notre Dame
Psychology Department

Conclusion

Main Points

For a simple scenario-setting, there is no optimal strategy to act against moral hypocrisy independent of population and temptation density – parameters that determine paradigmatic social scenarios. In fact, the most successful strategies in one scenario can have disastrous consequences in other scenarios.

For a more complex scenario-setting, there are successful strategies against moral hypocrisy. The effect on population uniformity depends strongly on setup-conditions and on the number of agents and temptations.

Dynamics (in the sense of strategy change) seems to play a very important role.

Reminder of shortcomings

The current model does not take into account the psychological complexity of moral hypocrisy with respect, e.g., to the type of temptation.

The model does not account for real world handling of moral hypocrisy, e.g. forgiveness.

The current analysis does not involve all possible social strategies against moral hypocrisy.

Other model parameters may become object of further investigations (e.g. changes in the payoff-structure, non-Moorean interactions between agents, agent-temptation interactions).

Of particular interest: dynamic strategy-changes.

University of Zurich[UZH]
Institute of Biomedical Ethics

University of Notre Dame
Psychology Department

Thank you!

And thanks to the members of the Institute "Computer Simulation in the Humanities"

And Daniel Singer for advice