A Model of Moral Hypocrisy as a Model of How to Learn Agent-based Modeling

Markus Christen

Content

- Goals
- Moral Hypocrisy
- A Payoff System for Moral Hypocrisy
- Moral Hypocrisy 1.0
- Moral Hypocrisy 1.1
- Moral Hypocrisy 2.0
- Moral Hypocrisy 2.1
- Moral Hypocrisy 2.2
- Next Steps

Goals

Before I came to the workshop

"(…) we want to understand the dependence of an agent-based model of moral intelligence on moral ontologies – a semantic net that defines the connectedness and closeness of various concepts referring to moral issues (...)."

After the first few days:

1. Learn how to built a model *in concreto*
2. Find a problem relevant for morality that is simple enough to model
3. Expand the model step by step.

I report findings and insights

INSIGHT NO 1:

DON’T FEAR SIMPLICITY, DON’T BE AFRAID TO APPEAR STUPID

Moral Hypocrisy

A (canonical) definition of moral hypocrisy (Dan Batson):

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

Eliot Spitzer

Tiger Woods

Ted Haggard

Karl-Theodor zu Guttenberg

A Payoff System for Moral Hypocrisy

Appearance		**Acting towards temptations**				
Moral:	1	morally	0	=	Good Guy	1
	1	immorally	1	=	Hypocrite	2
Immoral:	0	morally	0	=	Inconsistent	0
	0	immorally	1	=	Bad Guy	1

Four strategies, whereas hypocrisy is by default the best one.

As morality is expressed in a social word, I further define:

- If agent X appears as moral, he gets +1 from each neighbor he has.
- For any “temptation” in the neighborhood of X, he gets +1, if X acts immoral
- Adapt your strategy to the strategy of your best neighbor

INSIGHT NO 2:

FIND A PROBLEM THAT CAN BE FORMULATED IN A CLEAR WAY AND THAT CAN BE TRANSLATED IN A DEFINED AGENT-BEHAVIOR

Moral Hypocrisy 1.0

Simplest model with the following parameters:

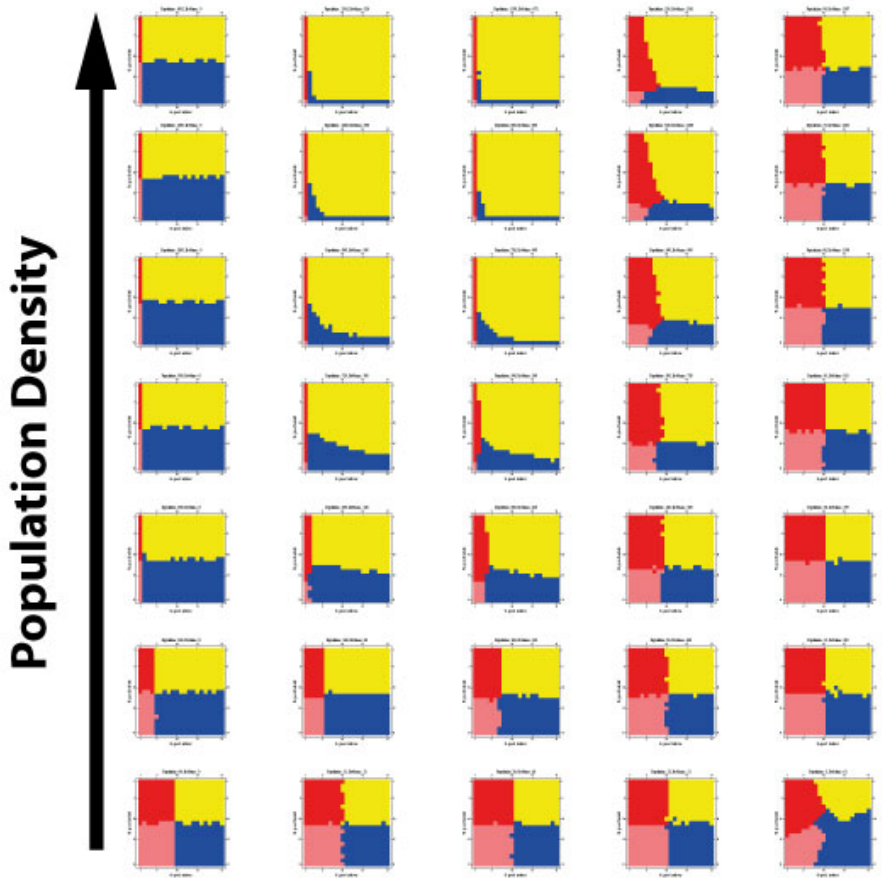
- Number of agents
- Number of temptations (occupy space)
- Probability of appearing good (setup)
- Probability of acting bad (setup)

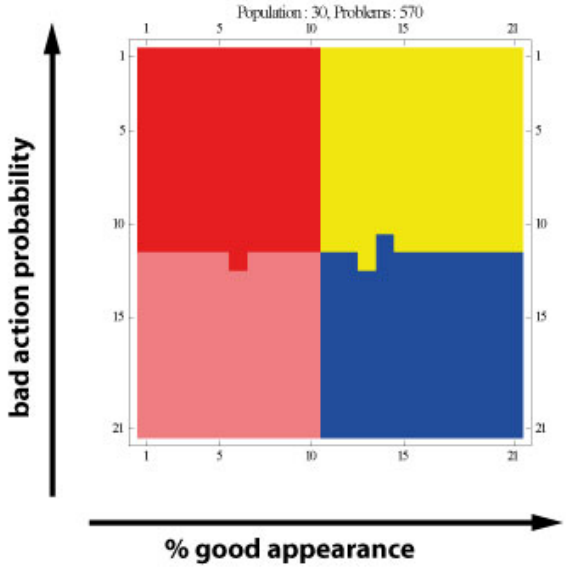
See which strategy gets the majority of agents depending on the settings of the behavior space.

Expectation:

- Hypocrits wins always unless their initial number is small and density effects (low agent density, high problem density) inhibit takeover.

Moral Hypocrisy 1.0





INSIGHT NO 3:

GETTING THE MODEL TO RUN OUTPERFORMS (IN TERMS OF MOTIVATION) THE DIAGNOSIS THAT THE MODEL IS BAD

Moral Hypocrisy 1.1

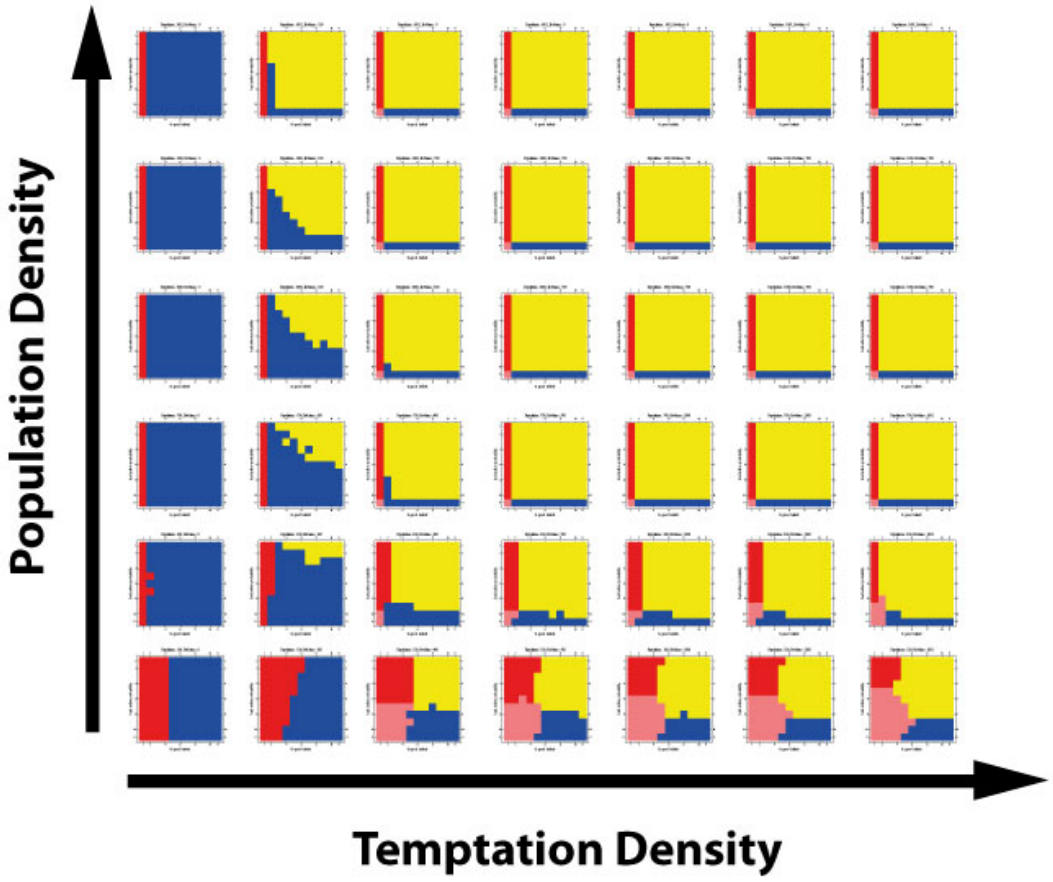
Eliminate some stupid initial settings:

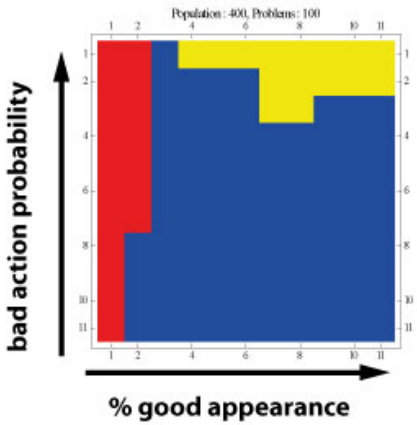
- Temptations should not occupy space
- The probability of acting bad in the setup should only apply if there is a temptation around

Expectation:

- Density effects will change: if agent and temptation density is low, hypocrisy should be less, if agent density is high, it should be more successful.

Moral Hypocrisy 1.1





INSIGHT NO 4:

DON’T BE IRRITATED IF YOU DON’T FIND “EMERGENCE”

Moral Hypocrisy 2.0

Start to enrich your model:

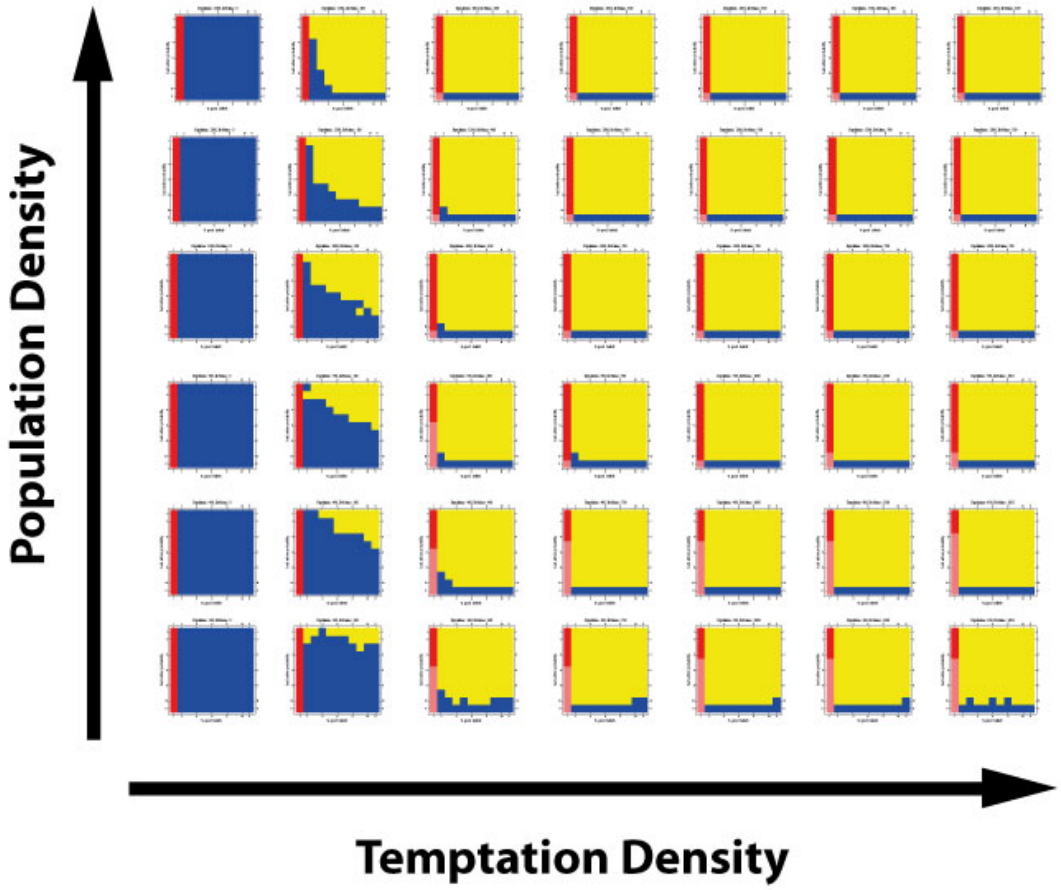
- Allow agents to move in the neighborhood of agents that appear good.

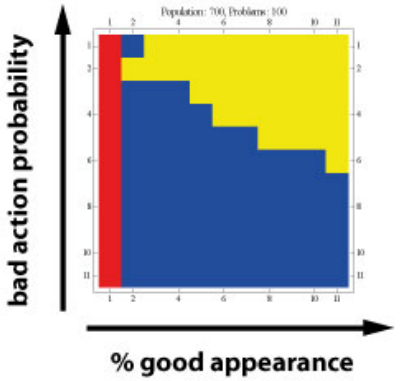
Rational: good appearing agents are a good predictor for good acting agents, as we usually don't know whether there are any temptations around.

Expectation:

- Good guys should be more successful, as long as the number of temptations is not too high.

Moral Hypocrisy 2.0





INSIGHT NO 5:

EXPLORING BEHAVIOR SPACE MAY REVEAL BUGS YOU NEVER THOUGHT THAT THEY COULD EXIST

Moral Hypocrisy 2.1

Start to enrich your model (2):

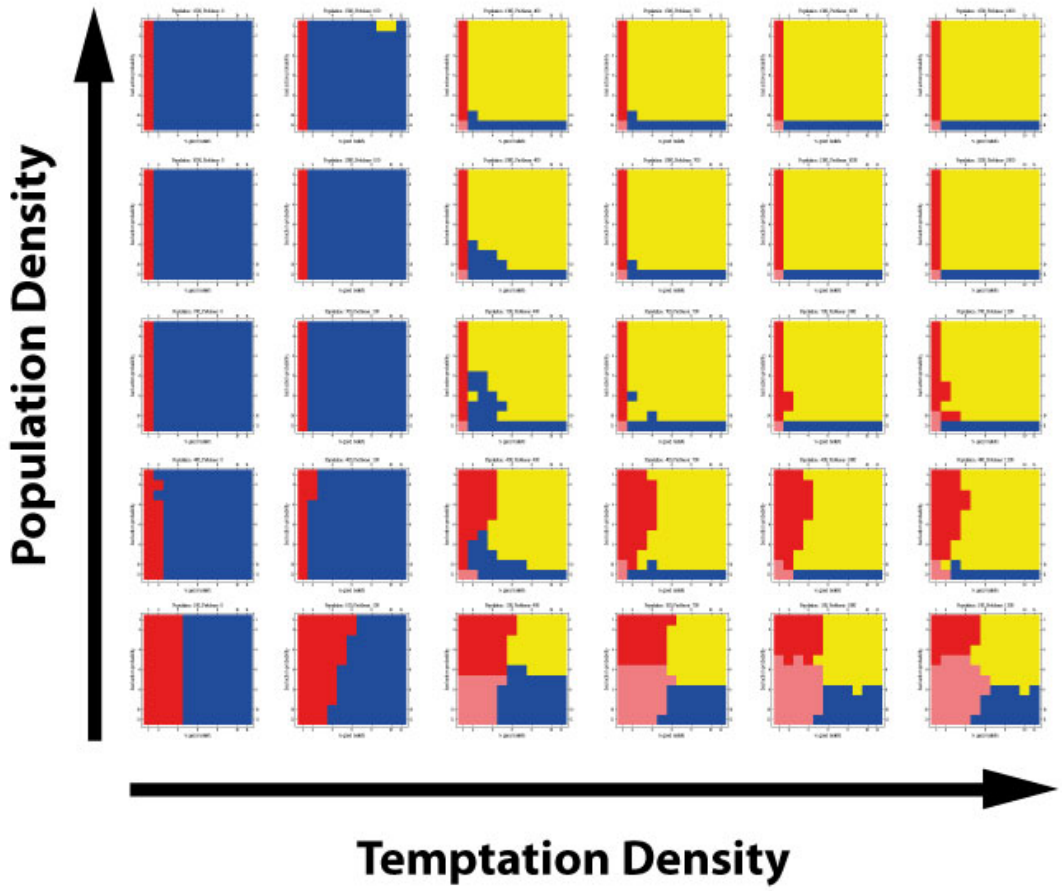
- Allow agents to avoid the neighborhood of agents that act immorally towards temptations.

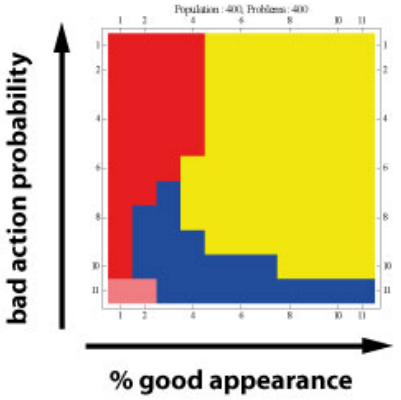
Rational: not the (potential) bad belief of a neighbor, but his bad actions actually can cause avoidance.

Expectation:

- Hypocrites should have harder times.

Moral Hypocrisy 2.1





Moral Hypocrisy 2.2

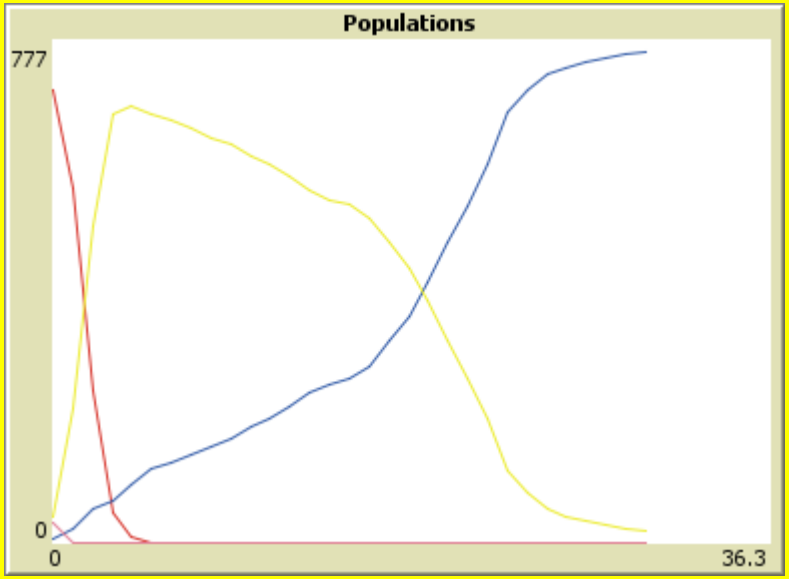
Now combine

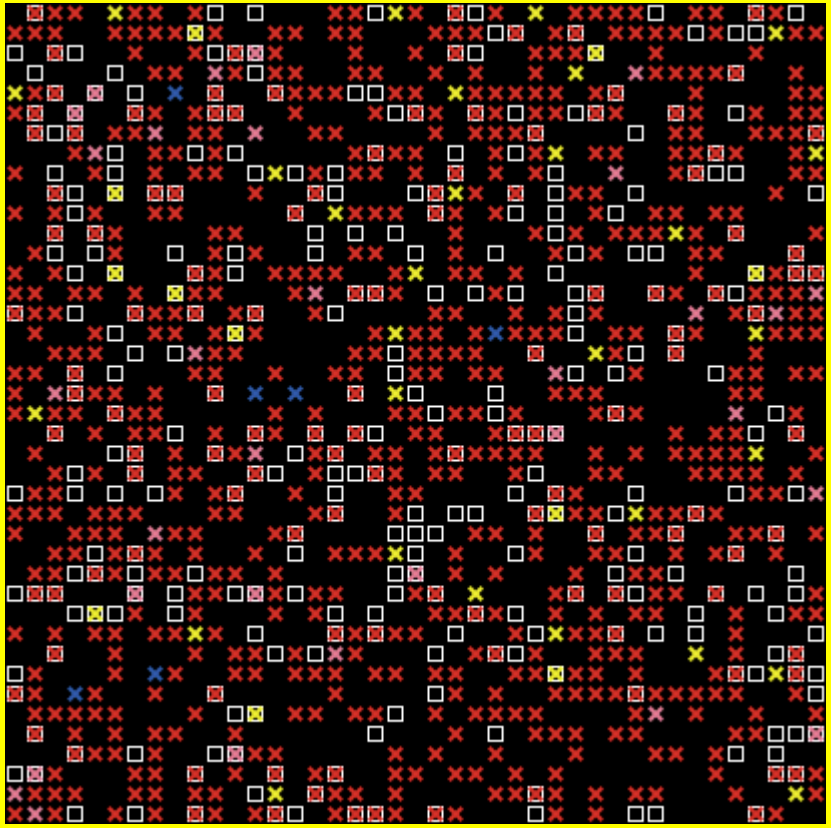
- Seeking the neighborhood of agents that appear moral
- Avoiding the neighborhood of agents that act immoral

Expectation:

- ???.

Moral Hypocrisy 2.2





INSIGHT NO 6:

ENJOY YOUR MODEL

Next Steps

1: Stay with Hypocrisy 2.2

- Explore agent distributions
- Explore dynamics of population growth/ decline

2: Implement more "external" factors:

- Change payoff-structure
- Change neighborhood
- Give "temptation-space" a structure
- Implement "asymmetric information flow" (e.g. effect of mass media: bad actions are appreciated in a wider environment compared to good appearance).

3: Start to include internal structure into the agents along the model of moral intelligence