

# Model of Epistemic Vigilance

## Model Overview, Design Concepts, Details (ODD)

Daniel Reisinger  
daniel.reisinger@uni-graz.at

### Model Purpose

The purpose of the model is to formalize and then analyze the mechanisms of epistemic vigilance as outlined by Mercier (2017) for their systemic behavior in a multi-agent setting. The mechanisms of epistemic vigilance are proposed as a means to communicate information between individuals while staying vigilant in the face of inconsistencies between received messages and background beliefs. The mechanisms include critically evaluating communicated information, checking the plausibility of its content, and checking the competence of the sender. The model is designed to answer questions such as: How do the mechanisms interact with each other? What kind of feedback loops can be observed? And under what conditions do they fail or succeed in containing the spread of false information? Increasing the degree of formalization by converting the existing verbal theory into a computational model (Smaldino 2017; van Rooij & Blokpoel 2020) can help to find answers such questions and contribute to the microscopic and macroscopic understanding of the mechanisms.

### Model Entities

The model includes the following entities: Agents and network structures.

#### 1. Agent

Agents are represented by nodes in a network. The network size or number of nodes in the network determines the number of agents in the social system.

##### 1.1. Agent attributes

Agents have the following attributes: a message, background beliefs, competence, and satisfaction.

- i. Message: can have the values 1 (true message), -1 (false message), and 0 (no message). Example: A message with a value of 1 indicates that the agent currently holds a true message. Agents with true messages are represented by cyan colored nodes. Agents with false messages are presented by red colored nodes.
- ii. Background beliefs: can have the values 1 (true background beliefs), -1 (false background beliefs), and 0 (no background beliefs). Example: Background beliefs with a value of 1 indicate that the agent currently holds true background beliefs.
- iii. Competence: can have the values 1 (competence), and 0 (no competence). Competent agents are represented with black outlined around the node.

- iv. Satisfaction: can take on the values 1 (satisfied) and 0 (not satisfied). This variable is not needed for the mechanisms of epistemic vigilance and only needed for the Schelling segregation algorithm.

## 1.2. Agent behaviors

Agents are equipped with three different behaviors that include the three mechanisms of epistemic vigilance: critical evaluation, plausibility checking, and competence checking. They mechanisms are organized in a 'fast-and-frugal'-like manner, meaning certain values in agent attributes trigger certain mechanisms.

- i. Base mechanism (includes only critical evaluation): This agent behavior involves two agent attributes: a sender's message and a receiver's background beliefs. If the receiver agent has no prior background beliefs (background beliefs with value 0), the receiver will gullibly accept the message and update the value of their message attribute. If the receiver has matching background beliefs with the message (e.g. 1 in background beliefs and 1 in received message), they will also accept the message and update the value of their message (this is however not considered gullible acceptance). If the receiver has conflicting background beliefs with the message (e.g. -1 in background beliefs and 1 in received message), the receiver will perform a critical evaluation. In the base mechanism, any inconsistency between message entry and background beliefs entry trigger a critical evaluation. The critical evaluation is modeled as a probability event with a certain success rate (constant parameter). A successful critical evaluation leads to the agent finding the truth followed by an update in background beliefs and the message they hold (both values set to 1). An unsuccessful critical evaluation leads to the agent finding the falsity followed by an update in background beliefs and the message they hold (both values set to -1).
- ii. Base mechanisms + plausibility checking: Plausibility checking is modelled as an additional case to the base mechanisms. This agent behavior involves three agent attributes: a sender's message, a receiver's background beliefs, and the receiver's competence. Depending on the receiver's competence, receiver's facing an inconsistency between message and background beliefs perform a plausibility check. Plausibility checking is triggered when an agent faces an inconsistency but is not competent. The plausibility check is modelled as a probability event with a certain success rate (constant parameter). A successful plausibility check results in the agent rejecting the inconsistent information without updating any of their attribute values. An unsuccessful plausibility check results in the agent accepting the information and updating the values of their message and background beliefs to match the value of the received message.
- iii. Base mechanism + plausibility checking + competence checking: Competence checking is modelled as another additional case the base mechanisms. This agent behavior involves four agent attributes: a sender's message, a receiver's background beliefs, the receiver's competence, and the sender's competence. Competence checking is triggered when the receiver of a message faces an inconsistency between message and background beliefs, they themselves are not competent, but their sender is. Competence checking is modelled as a probability event with a certain success rate (constant parameter). A successful competence check results in the receiver accepting the message and updating their background beliefs, superseding the plausibility check in the process. An unsuccessful competence check results in them falling back on the plausibility check instead.

## 2. Network structures

Network structure determines the possible interaction partners of an agent. Example: an agent selects one of its direct neighbors in a given network structure as a conversation partner. The selecting agent then performs the role of message receiver and the selected agent the role of a message sender.

Possible network structures in the model are: a network based on a von Neumann neighborhood, on a Moore neighborhood, a network constructed from a Voronoi diagram, and a Watts-Strogatz graph.

- i. Von Neumann neighborhood: regular grid with 4 neighbors (except boundaries)
- ii. Moore neighborhood: regular grid with 8 neighbors (except boundaries)
- iii. Voronoi neighborhood: irregular grid-like network
- iv. Watts-Strogatz graph: average degree 10, rewiring probability 0.01, corresponds to a network with high average clustering coefficient and low average shortest path length

## Model Setup

### 1. Stylized simulations:

#### 1.1. Scenario 1: Locality of critics

- i. Simulation with a social network containing 81 agents.
- ii. Agents are organized in one of the possible network structures (von Neumann, Moore, Voronoi, Watts-Strogatz graph) where every node represents an agent.
- iii. Agents are equipped with the base mechanism (only critical evaluation)
- iv. The critic agent is placed on the most central node (network closeness centrality measure) and has true background beliefs (value set to 1)
- v. The agents on the remaining nodes have no background beliefs (value set to 0) and one of these agents is randomly selected to hold a starting false message (value set to -1)
- vi. The success rate of critical evaluation is set to 1.
- vii. For case A, the critic agent is connected to all of its structural neighbors. For case B, the critic agent is connected to only 1 of its structural neighbors.

#### 1.2. Scenario 2: Impeding structures

- i. Simulation with a social network containing 81 agents.
- ii. Agents are organized in one of the possible network structures (von Neumann, Moore, Voronoi, Watts-Strogatz graph) where every node represents an agent.
- iii. Agents are equipped with the base mechanism + plausibility checking.
- iv. The critic agent is placed on the most central node (network closeness centrality measure), has true background beliefs (value set to 1) and is competent (value set to 1)
- v. The critic's direct neighbors in the network have true background beliefs (value set to 1) and are not competent (value set to 0).
- vi. The agents on the remaining nodes have no background beliefs (value set to 0) and one of these agents is randomly selected to hold a starting false message (value set to -1).
- vii. The success rate of critical evaluation is set to 1.

- viii. For case A, the success rate of plausibility checking is set to 1. For case B, the success rate of plausibility checking is set to 0.99.

### 1.3. Scenario 3: Breaking structure

- i. Simulation with a social network containing 81 agents.
- ii. Agents are organized in one of the possible network structures (von Neumann, Moore, Voronoi, Watts-Strogatz graph) where every node represents an agent.
- iii. Agents are equipped with the base mechanism + plausibility checking + competence checking.
- iv. The critic agent is placed on the most central node (network closeness centrality measure), has true background beliefs (value set to 1) and is competent (value set to 1).
- v. The critic agent is assigned an initial true message for communication.
- vi. The critic's direct neighbors in the network have false background beliefs (value set to -1) and are not competent (value set to 0).
- vii. The agents on the remaining nodes have no background beliefs (value set to 0) and one of these agents is randomly assigned to hold a starting false message (value set to -1).
- viii. The success rate of critical evaluation is set to 1.
- ix. The success rate of plausibility checking is set to 1.
- x. For case A, the success rate of competence checking is set to 0. For case B, the success rate of competence checking is set to 0.1

### 1.4. Scenario 4: Schelling segregated simulation

- i. Simulation with a social network containing 961 agents.
- ii. Agents are organized in a Moore neighborhood network and in a Watts-Strogatz graph.
- iii. Agents are equipped with the base mechanism + plausibility checking + competence checking.
- iv. Agents with the following attributes are placed randomly on the network nodes before Schelling segregation algorithm repositions them: 9 agents with background beliefs set to 1 and competence set to 1, 238 agents with background beliefs set to 1 and competence set to 0, 238 agents with background beliefs set to -1 and competence set to 0, and 476 agents with background beliefs set to 0 and competence set to 0.
- v. We set the message for 10 random agents to 1, and of another 10 random agents to -1.
- vi. Agents are then repositioned in the network following the Schelling segregation algorithm: Over a maximum of 100 iterations, it is checked whether an agent is satisfied with its current neighborhood. Agents are satisfied if a certain percentage (parameter: tolerance level = 0.3) have the same background beliefs as them. If they are satisfied, the agent remains at its current location and the attribute satisfaction is set to 1. If they are not satisfied, they switch places with a random agent with background beliefs 0 and competence 0. An iteration is over when all agents in random activation are checked for their satisfaction and potentially relocated. The algorithm is run for a maximum of 100 iterations.
- vii. The success rate of critical evaluation is set to 0.99
- viii. The success rate of plausibility checking is set to 0.99
- ix. The success rate of competence checking is set to 0.1

## Model Processes

After the initialization phase the model is run for T ticks (T = 200 in stylized simulations, T = 2000 in Schelling segregated simulations).

What happens in each timestep or tick?

- i) Every agent (node) in the network is selects an interaction partner from its immediate neighborhood (directly linked nodes).
- ii) The selected neighbor performs the role of a message sender (if it has one).
- iii) The selecting agent performs the role of a message listener.
- iv) The selecting agent behaves based on the choosen agent behavior: base mechanisms (+ plausibility checking (+ competence checking)).
- v) A tick is over when all agents had the opportunity to select an interaction partner.

## Outputs

For each tick we calculate the following metrics:

- i) The number of true messages held by agents in the network
- ii) The number of false messages held by agents in the network
- iii) The number of no messages held by agents in the network

## Concepts and Principles

What important concepts or principles are represented in the model?

Interaction.

- i) Agents interact based on the network structure they are in.
- ii) Agents may only interact with direct neighbors (links in the network).

Randomness.

- i) Agents randomly pick one of their neighbors in the network as an interaction partner per tick.
- ii) In the Schelling segregation algorithm, agents switch place with a randomly selected agent with background beliefs 0 and competence 0.

## References

Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. The ODD protocol: a review and first update. *Ecological modelling*. 2010 Nov 24;221(23):2760-8.