

# Meta ABS: An approach to Free Will in Agent-Based Simulation

Jonathan THALER

February 27, 2017

## Abstract

Meta Agent-Based Simulation (MetaABS) is an attempt of modelling free will on a computational basis. We focus on free will which we define as having the ability to anticipate results from one actions thus creating a feedback on the actions. Here we follow the interpretation that anticipation means 'to simulate' in the context of ABS. We follow the concept of irreducibility of computation which, due to undecidability - requires to run (=simulate) a program to actually know its output. The idea is in each step of the simulation let agents simulate the same simulation they are situated in from their point of view for a given number of steps before they take their next step. Thus the simulation is defined in terms of recursion, a novelty and something we will describe in depth in this paper. We claim that functional programming is especially well suited to implement MetaABS due to its lack of implicit side-effects, natural parallelism and declarative way of describing WHAT systems are instead of HOW they compute. Obviously this new approach has a problem: an agent would need to have complete information about the whole simulation, an assumption which is completely unrealistic. We introduce the term of 'reduced-recursive' which assumes that the next recursive level is a reduced version of the original, very local to the agent, thus solving the problem of complete information. Our method may look like another optimization method but we will show that this is not the case. Also our motivation and intention is completely different and our intended area of application are the social sciences, artificial life, philosophy and maybe religious studies.

## 1 Motivation

[1] came up with the initial simulation argument.

[3] discusses theological implications of the simulation argument and looks into reasons why we are simulated. He mentions 'evolution of complexity' and gaining of knowledge. We argue that this comes only out of free will. We are aware

that by touching on the subject of 'why' we are being simulated we touch on ideological ground but this is intentional so.

[2] defines artificial free will of a machine to be: TODO

- We ask how existence can be understood from the perspective of computer science and follow in our ideas the simulation hypothesis as of [1] which states that existence is a simulation (TODO: bostrom does not state this!) - We differ from the simulation hypothesis which does not state any deeper meaning or reason for existence by hypothesizing that existence is a simulation which allows consciousness to express free will. - We will show that granting free will outside of a simulation is too dangerous as we claim that it will always leads to ideologies which result in separation, inequality, suffering and ultimately in death, something this simulation clearly exhibits.

- We propose computational models of free will and consciousness and of the simulation itself. - Further we touch on the questions who is behind this simulation, what is outside of it and what may be beyond of it. - Our final hypothesis is that the ultimate purpose of this simulation of free will is to lead the entities which populating this existence to an understanding of the problem of free will and ideology, transcending both and giving them up for love. - We claim that love in this simulation can be understood as a total and complete understanding of what is - a complete understanding and grasping of truth: to experience objective existence. This love will enable the entities to abandon the feeling of separation, devolving them of their primal fear, resulting in true inner peace, enabling them to step up the simulation level after 'death'.

- One implication is that there is ultimately no death but only a transformation of data. - Another rather cruel implication is that the suffering serves a purpose and is not real anyway. - We hypothesize that there are further levels of simulations up to 12 layers where the final layer ultimately is what religion calls god. - we claim that free will is able to 'simulate' its actions thus anticipating them which allows to behave in various non-deterministic ways. - we give a new model of an agent-based simulation, called meta abs in which agents run the simulation locally enabling them to anticipate in a limited way what is going to happen. we show that pure functional programming is especially well suited in implementing this new technique and explore it in a model of free will - question: how could abs be capable of simulating consciousness and free will? hypothesis: we need a meta-level in our simulation which allows us to simulate the simulation - question: can we develop a model and a simulation of the simulation hypothesis?

## 2 Research Questions

### 2.1 What is Free Will in ABS?

**Hypothesis** It is the ability of an agent to simulate on a meta-level and adjust itself based on these insights

We are NOT interested in another optimization algorithm by applying this free-

will definition to optimization problems, but we want to apply it to artificial life / philosophical simulation models. Thus we must be very careful to not confuse it with e.g. game-theory where one makes assumptions about the other players - this may seem to be the same but we are not interested in this approach and we argue it is very different: we simulate instead of rational calculation. -¿ problem: how big/complex can the meta-level be? does the agent has a perfect model in its mind?

## 2.2 How can we model Free Will in ABS?

**Hypothesis** we need a meta-model: a recursive, declarative description of the simulation

## 2.3 How can we implement free will in ABS?

**Hypothesis** pure functional programming is suitable due to the recursive, declarative nature of the meta-model

## 2.4 Can we build a model of the simulation hypothesis / argument to test the difference between free will and the lack thereof?

- does free will lead to ideologies (what are ideologies?)? if yes, how many agents are sufficient?
- are there differences between having free will and using it AND of having free and not using it?
- agents try to find out if they are in a simulation or not
- agents are ordered on a scale in their beliefs about argument 1-3 of the simulation hypothesis
- agents exhibit a range of primal fear between 0 and 1 where 0 means they are enlightened and 1 means they are completely stuck in the simulation
- some agents have free will (can run a meta-simulation) and some don't but the free-will agents don't know which the others are. it is the goal of the free-will agents to find out the non-free will ones. BUT HOW? free-will agents can go on a meta-level and compare their result with the actual result, if there is a mismatch then they know the other agent has free will and otherwise not.

## References

- [1] BOSTROM, N. Are We Living in a Computer Simulation? *The Philosophical Quarterly* 53, 211 (2003), 243–255.

- [2] IRTEM, A. Simulation of Free Will. In *Current Topics in Cybernetics and Systems: Proceedings of the Fourth International Congress of Cybernetics & Systems 21–25 August, 1978 Amsterdam, The Netherlands*, J. Rose, Ed. Springer Berlin Heidelberg, Berlin, Heidelberg, 1978, pp. 148–149. DOI: 10.1007/978-3-642-93104-8\_92.
- [3] STEINHART, E. Theological Implications of the Simulation Argument. *Ars Disputandi: The Online Journal for Philosophy of Religion* 10 (2010), 23–37.