Abstract Submitted
for the CUWIP21 Meeting of
The American Physical Society

Modeling discrimination in societies of neural networks MARIANA MERCUCCI, OTAVIO CITTON, FELIPPE ALVES, NESTOR CATICHA, Institute of Physics - University of Sao Paulo, CNAIPS TEAM — This ongoing project deals with the quantitative study of polarization and formation of groups holding opposite opinions on a set of issues in the context of agent based models where the agents are neural network classifiers. The agents exchange binary opinions on a set of multidimensional issues. In this case, polarization is driven by adaptive affective distrust and the inclusion of irrelevant features to the problem being discussed. We consider the case where some agents extend the correctly parsed assertion with a set of numbers that are irrelevant to the classification problem, but depend only on the emitter agent. This irrelevant addition acts like disrupting noise, driving agents to effectively learn from a group of similar agents and to unlearn from other agents. We use the Entropic Dynamics learning algorithm for neural networks (EDNNA) which has been extended to model societies where the agents are perceptrons. At a given time of a discrete dynamics a pair of agents is chosen at random, one acts as the emitter and the other as the receiver of a pair (input vector - label) information. The dynamics is performed by the update of the weights and the distrust of the receiver towards the emitter. We present some preliminary results from simulations.

Mariana Mercucci
Institute of Physics - University of Sao Paulo

Date submitted: 04 Jan 2021 Electronic form version 1.4