

## Going for the Third: Cooperation between LLM Agents in a Three-Option Setting

Leonardo Mascagni, Ossama Zaroual, Mauro Mezzenzana, Giacomo Buonanno, Francesco Bertolotti

School of Industrial Engineering, and Intelligence, Complexity and Technology Lab (ICT Lab),  
LIUC - Carlo Cattaneo University, Castellanza, Italy  
fbertolotti@liuc.it

In this work, we investigate emergent capabilities arising from cooperation between AI models by employing a minimal interaction protocol where two agents communicate iteratively, with one agent’s response serving as the input for the other. We formalize the concept of ”collective opinion” as a dynamic state moving within a 2-simplex (triangle), defined by three vertices: the initial stance of Agent A (Opinion A), the initial stance of Agent B (Opinion B), and a third vertex representing a novel, synthesized opinion (Opinion C) that agents may generate rather than aligning with pre-existing viewpoints. Each message exchanged induces a discrete update that shifts the collective pointer toward one of these vertices, creating a trajectory that visualizes the competition between consensus-seeking, persuasion, and creative reframing.

We examine how the pointer’s position evolves based on controllable factors, including the number of agents, the number of distinct opinion prototypes, model parameters, and the inherent strength or evidential weight of the opinions (e.g., highly factual versus contested claims). Furthermore, we propose a generalization of this framework from the triangular simplex to higher-dimensional geometries (e.g., squares or hypercubes) to accommodate larger agent populations and multiple competing viewpoints. This approach allows us to verify the conditions under which cooperative AI systems converge to a stable consensus, oscillate in mixed states, or successfully generate emergent and novel solutions.

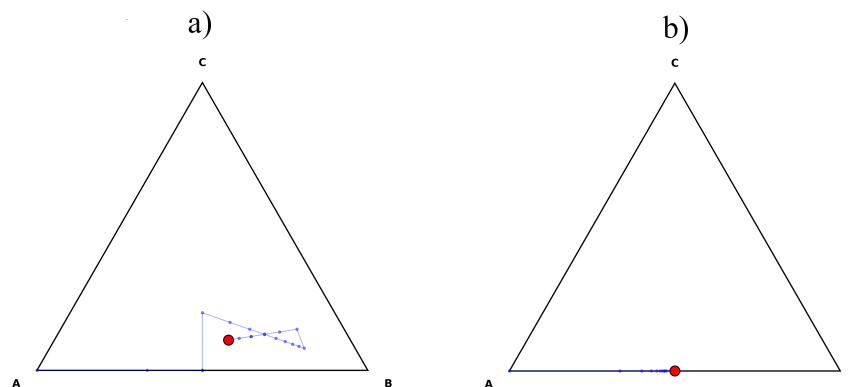


Figure 1: Trajectories of agents’ average opinion for Gemma 3 8B (a) and GPT-5 (b)