FONDAZIONE BRUNO KESSLER
MAGAZINE

What a Brain-Simulating Model Revealed About the Hidden Rules of Human Cooperation

October 30, 2025

Key findings from “AI@TN: study of Cooperative AI frameworks” - funded by the Autonomous Province of Trento and authored by Bruno Lepri and Laura Ferrarotti (from MobS @ FBK), and Riccardo Gallotti and Lucila Alvarez-Zuzek (from CHuB Lab @ FBK).

The Puzzle of Human Cooperation

Why do we choose to cooperate when pure self-interest might suggest otherwise? It is a question that lies at the heart of human society, from small teams to entire nations. We see cooperation flourish, enabling incredible achievements, but we also see it break down, sometimes with disastrous results. What is the invisible calculus that governs whether we join hands or go it alone?

To explore this puzzle, we – commented **Lucila Alvarez-Zuzek**, Researcher at CHuB Lab @ FBK Digital Society Center, enhanced a well-established cognitive model known as the **Drift Diffusion Model** (DDM). The model mimics how our brain accumulates evidence before making a choice and the cognitive mechanisms behind the process. The researchers’ key innovation was to make this model **sensitive** to **social context**, allowing it to predict how an individual’s decision-making would *evolve* based on their interactions with others. This turns the model into a powerful tool for understanding the most crucial element in cooperation: the learning process.

The results were fascinating. The model did not just confirm what we already know; it uncovered several surprising and counter-intuitive takeaways about how sensitive our cooperative instincts are to what we learn about our environment. Here are **four key insights** that challenge our common assumptions about how and why we work together.

Takeaway 1: We are More Cooperative When We Expect Less of It

This finding turns conventional wisdom on its head. The simulation placed a player in different environments, some highly cooperative and others less so. The model revealed a subtle but surprising effect: when a player was surrounded by less cooperative individuals, they showed a slight uptick in their own cooperation. Conversely, when placed in a group of mostly cooperators, the player's tendency to cooperate decreased over time.

This wasn't an act of altruistic defiance. Instead, the specific rules of this game made cooperation a surprisingly smart individual move when surrounded by defectors, and defection the most profitable move when surrounded by cooperators. The model simply uncovered the human capability of learning to play the odds, showing how quickly we master the ability of exploiting the *incentives* of a static environment. This explains the paradox of the "too-nice" team, where a lack of accountability can tempt individuals to slack off, assuming others will pick up the slack. The system's incentives—or lack thereof—overwhelm good intentions.

Takeaway 2: Shuffling the Team Hits the "Cooperation Reset" Button

The model revealed what researchers call a "restart effect." When a player was moved from one group to a new, unfamiliar one, their level of cooperation would sharply increase at the beginning of the new game. It seems a fresh start forces us to unlearn old patterns of defection or mistrust and temporarily revert to a more cooperative default.

However, this cooperative boost is not permanent. As the player gains experience with the new group and learns its dynamics, their tendency to defect gradually rises again. This is a crucial insight for real-world organizations. It explains why structures like agile development "sprints" or the start of a new school semester can temporarily boost enthusiasm and cooperation. Periodically changing team dynamics can serve as a powerful "reset," motivating people to give collaboration another chance.

Takeaway 3: The Stick is Mightier Than the Carrot

To test the power of incentives, the researchers ran simulations where the game's rules were changed to either reward cooperation (the carrot) or punish defection (the stick). Both strategies successfully increased the overall rate of cooperation, confirming that incentives matter.

However, the model found a clear difference in their effectiveness. While both approaches worked, punishing non-cooperation had a greater effect in promoting cooperative behavior. As the researchers state:

...punishing defection seems to have a slightly greater impact than rewarding cooperation.

This finding might be directly relevant to public policy debates, such as the difference between a carbon tax (a stick) and a green energy subsidy (a carrot). When crafting policies or community guidelines, this suggests that clear consequences for non-cooperative actions may be a more powerful motivator than simply rewarding good behavior.

Takeaway 4: Think Fast, Cooperate More

What happens when we don't have time to overthink our choices? The model simulated this by putting players under time pressure. The result was a significant increase in cooperative behavior, especially in the early rounds of the game.

This aligns with the "Social Heuristic Hypothesis," which posits a conflict between our intuitive, cooperative default (fast thinking) and the more calculated strategies we learn through deliberation (slow thinking). The model revealed *why* this happens: time pressure made players less cautious. Instead of carefully weighing the options, they made more impulsive, gut-level decisions. This lack of deliberation short-circuited the learned tendency to defect, allowing a more intuitive, cooperative heuristic to take over, especially early on.

The Sensitive Nature of Cooperation

These simulations reveal that our willingness to cooperate is not a fixed trait but a dynamic behavior that is highly sensitive to the context we are in. Our choices are profoundly shaped by the rules of engagement, the actions of those around us, and even the amount of time we have to think. Cooperation is not just about individual morality; it is about the systems we operate within and what they teach us.

"Our model – commented **Laura Ferrarotti,** Researcher of MobS Lab @ FBK Augmented Intelligence *– is capable of predicting patterns that are similar to those observed in real experiments of human cooperation in multiplayer social dilemmas scenarios: firstly, cooperation often declines over time as individuals realize that defection can be more advantageous, but it temporarily increases when groups are reshuffled, reflecting a reset of trust. Secondly, we also find that changing the payoff structure—by rewarding cooperation or punishing defection—significantly boosts cooperative behavior, with punishment being particularly effective. Finally, under time pressure, people (and our model) tend to respond less cautiously, which initially increases cooperation. These results suggest that cooperation is highly sensitive to context, incentives, and social dynamics, and that our model can capture these nuances in ways previous approaches could not. Beyond offering a more accurate description of decision-making, this framework provides a practical tool for exploring policies and interventions that could encourage cooperation in real-world settings, generating future applications such as AI-driven "social planners" designed to foster collective welfare."*

If our intuitive, gut reactions lean toward cooperation, what does that say about the environments we have built for ourselves—and how could we design them better?

PERMALINK

https://magazine.fbk.eu/en/news/what-a-brain-simulating-model-revealed-about-the-hidden-rules-of-human-cooperation/

TAGS

- #artificialintelligence

- #augmentedintelligence
- #chub
- #collaborative decision-making
- #DDM
- #digitalsociety
- #human behavior
- #human decision-making
- #mobs

RELATED MEDIA

- Link to the paper on the Journal of the Royal Society Interface: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2025.0168

AUTHORS

- Giancarlo Sciascia