

Adaptive Cooperation under Flood Risk: An Agent-Based Analysis of Informational Nudging

Alessandro Sapienza^{1,*}, Rino Falcone¹

¹*Institute of Cognitive Sciences and Technologies, National Research Council (CNR-ISTC), 00196 Rome, Italy*

Abstract

In this work, we investigate how informational nudging can influence the emergence of cooperation in large-scale populations exposed to environmental risk. We propose a spatially grounded agent-based model in which individuals decide how much to invest in shared protective infrastructures. The problem is framed as a public goods game under uncertainty, where agents are characterized by bounded rationality and adapt their strategies through a contextual bandit learning mechanism. Within this framework, we introduce an informational nudging mechanism that incorporates global feedback on collective outcomes into agents' learning process, without imposing direct constraints on their decisions.

Simulation results show that cooperation exhibits strongly non-linear dynamics. The system remains robust to free-riding behavior up to a critical threshold, beyond which coordination collapses. Moreover, the impact of nudging is highly sensitive to its intensity: weak signals may interfere with learning and reduce performance, while sufficiently strong signals act as effective coordination mechanisms, fostering stable cooperative regimes even in adverse conditions.

Overall, our findings highlight both the potential and the limitations of informational nudging in complex socio-environmental systems, suggesting that its effectiveness depends on the interplay between individual learning dynamics and collective feedback mechanisms.

Keywords

Multi-agent simulation, nudging, floods, hydrogeological risk, climate change.

1. Introduction

Global challenges associated with climate change and environmental degradation are becoming increasingly pressing, requiring coordinated and collective responses from both institutions and citizens. In particular, extreme events such as floods and hydrogeological hazards represent an extremely relevant problem, causing incalculable damage every year in terms of both economic losses and human lives. These phenomena are expected to intensify in both frequency and severity, amplifying their social and economic impacts [1, 2, 3, 4]. In these contexts, the resilience of communities does not rely solely on institutional interventions, but also critically depends on the active participation and behavior of individual citizens as well as institutional stakeholders [5]. Consider, for example, that virtuous behavior by citizens can reduce damage by as much as 80% [6].

However, these situations are often characterized by inherent tensions between individual and collective interests. They can be effectively framed as social dilemmas [7, 8], in which personal incentives to act opportunistically coexist with the broader need for cooperation. A prominent example is free-riding behavior, where individuals benefit from shared resources or collective efforts without contributing themselves. This dynamic poses a significant obstacle to the sustainability of cooperative systems, especially in real-world settings where individual actions are not easily observable or traceable. Under such conditions, traditional mechanisms based on monitoring, rewards, and sanctions become difficult to implement effectively.

These limitations have led to growing interest in behavioral approaches that target the cognitive and social dimensions of decision-making, such as trust, expectations, and perception of risk [9, 10, 11, 12].

WOA 2026: 27th Workshop "From Objects to Agents", June 15–17, 2026, Salerno, Italy

*Corresponding author.

✉ alessandro.sapienza@istc.cnr.it (A. Sapienza); rino.falcone@istc.cnr.it (R. Falcone)

id 0000-0001-9360-779X (A. Sapienza); 0000-0002-0300-458X (R. Falcone)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In this perspective, nudging and boosting strategies have emerged as promising non-coercive tools to influence citizens' behavior, guiding individuals toward socially beneficial actions without altering the underlying incentive structure [13, 14]. In environmental and collective-risk domains, these approaches have evolved into so-called green nudges [15] and cooperation nudges [16], aimed at promoting sustainable and cooperative behaviors.

While the effectiveness of these interventions has been widely explored in controlled experimental settings (e.g., surveys and laboratory experiments) [17, 18], their impact in large-scale, real-world communities remains less understood. Indeed, such communities of individuals exhibit complex and non-linear dynamics, where individual behaviors, environmental factors, and social interactions intertwine, making outcomes difficult to predict.

To address this challenge, agent-based modeling offers a powerful framework for capturing such complexity [19]: citizens' behavior can be modeled using agents, thus enabling the investigation of emerging community dynamics. Thus, in this work, we develop a large-scale agent-based simulation modeling a community exposed to the risk of critical hydrogeological events. Within this scenario, citizens' behavior is modeled through a public goods game, where a shared protective resource, represented by a system of shelters, is available to mitigate the impact of critical events.

The availability and effectiveness of these shared infrastructures depend on the collective investment decisions of individuals, who must allocate limited resources under capacity constraints. Agents repeatedly decide how much to contribute, facing a trade-off between personal benefit (exploiting others' contributions for their own advantage) and collective welfare (contributing in such a way that the community as a whole is protected). Their strategies evolve over time through a reinforcement learning mechanism based on a contextual bandit formulation, enabling adaptation to both past experiences and changing environmental conditions.

Within this framework, we introduce an informational nudging mechanism designed to influence agents' decisions without imposing explicit constraints on their choices. The objective is to investigate how such signals can foster cooperation and mitigate free-riding tendencies, particularly in complex and uncertain environments.

It is important to note that the objective of this work is not simply to reproduce human behavior, but rather to explore how rational agents, endowed with learning and adaptation capabilities, can develop cooperative strategies that are optimal with respect to the context and the behavior of their peers. This distinction is crucial: the simulation model is not intended to replace or mimic human decision-making, but to identify the strategies that would emerge as rationally advantageous in a complex and dynamic system.

In this sense, the model should be interpreted as a form of computational exploration, aimed at understanding how cooperation may evolve in idealized artificial populations. Such an approach allows us to isolate key mechanisms—such as learning, interaction, and feedback—that are often difficult to disentangle in real-world settings. Moreover, it provides a basis for comparing these emergent dynamics with empirical observations.

More broadly, simulations can be seen as a complementary tool to empirical evidence, enabling the exploration of scenarios that are not directly observable and the identification of behavioral patterns that, while not necessarily perfectly matching human behavior, can help explain and interpret real-world trends.

2. Related Work

Hydrogeological risk has long been recognized as a major global concern due to the severe economic, environmental, and social impacts associated with extreme events such as floods and landslides. As a consequence, it has been extensively studied in the literature. However, a large portion of existing research has primarily focused on the quantification of damages, including both direct and indirect losses affecting infrastructure, economic activities, and human lives [20, 21, 22, 23]. While these approaches provide essential insights into the magnitude and distribution of risk, they offer only a

partial understanding of the problem, as they largely overlook the behavioral dimension of population response.

To address this gap, a significant body of work has investigated how individuals perceive and respond to environmental risk. The most common approach relies on survey-based methodologies [24], often combined with socio-demographic or psychological models [6]. These studies aim to identify the key factors influencing protective behaviors, such as prior experience, perceived risk, and trust in institutions. Although surveys represent a valuable tool for capturing subjective perceptions, it is well established that stated preferences do not always translate into actual behavior in real-world contexts [25]. This limitation reduces the predictive power of survey-based approaches, especially in complex and dynamic scenarios such as natural disasters. Moreover, such studies focus on individual responses and are not explicitly designed to investigate dynamics at the community level.

To overcome these limitations, simulation-based methods have been increasingly adopted. In particular, agent-based modeling (ABM) provides a suitable framework to represent heterogeneous individuals and to capture the non-linear dynamics emerging from their interactions [26, 27]. Several contributions have applied these techniques to flood-related scenarios. For instance, computational frameworks have been proposed to model individual decision-making processes during flood events [28], taking into account factors such as prior experience, warning signals, and social influence. Other works have focused on evacuation dynamics and crowd behavior, integrating environmental processes with individual movement and interaction patterns [29, 30]. Additional studies have explored the role of communication strategies and information dissemination in shaping population responses [31, 32], highlighting the importance of social interactions in determining behavioral outcomes.

Despite these advances, existing works have largely concentrated on modeling individual behavior or specific processes such as evacuation and communication, often within context-dependent case studies. As a result, comparatively less attention has been devoted to the analysis of collective behavioral dynamics, especially in large-scale populations. In particular, the interplay between individual incentives and collective outcomes, typically framed as a social dilemma, remains underexplored in the context of hydrogeological risk.

Furthermore, while recent literature has started to investigate behavioral interventions aimed at influencing individual decisions [7, 8], such as nudging and information-based strategies, their application to large-scale, real-world risk scenarios is still limited. In particular, there is a lack of studies analyzing how informational nudging mechanisms affect the emergence of cooperation in environments characterized by uncertainty, limited observability, and the presence of free-riding behavior.

This work aims to address these gaps by focusing explicitly on collective behavioral dynamics in a flood-risk context, leveraging a large-scale agent-based model to investigate how cooperation emerges and how it can be influenced through informational nudging.

3. The Simulation Model

The proposed agent-based simulation model integrates spatial interactions, bounded rationality, and adaptive learning within a public goods game framework. In this way, it captures the interplay between decentralized adaptation, spatial constraints, and collective decision-making under uncertainty.

More specifically, the model enables the analysis of how boundedly rational agents, interacting through shared resources and local information, may develop collective strategies that can be either resilient or not, depending on the incentive structure and the presence of nudging mechanisms.

The simulation is implemented in NetLogo [33], a widely used platform for agent-based modeling, and incorporates GIS spatial data to reproduce a real urban road network. This allows agents to navigate actual streets, making access to shelters dependent on their geographic location.

3.1. Environment

The simulation world is formalized as a spatial graph structure, in which nodes denote relevant urban points such as road intersections, residential areas, and shelter locations, while edges capture the

connectivity of the road network, as derived from GIS data. This representation allows the model to incorporate realistic spatial constraints that influence agents' movement and access to resources.

Public shelters are randomly distributed across the territory and constitute the key shared resource in the system. Each shelter is associated with a capacity that is not fixed, but dynamically determined by the level of collective investment contributed by agents during each simulation round. As a result, the availability of protection emerges endogenously from the aggregated decisions of the population.

Figure 1 provides a visual overview of the simulated environment.

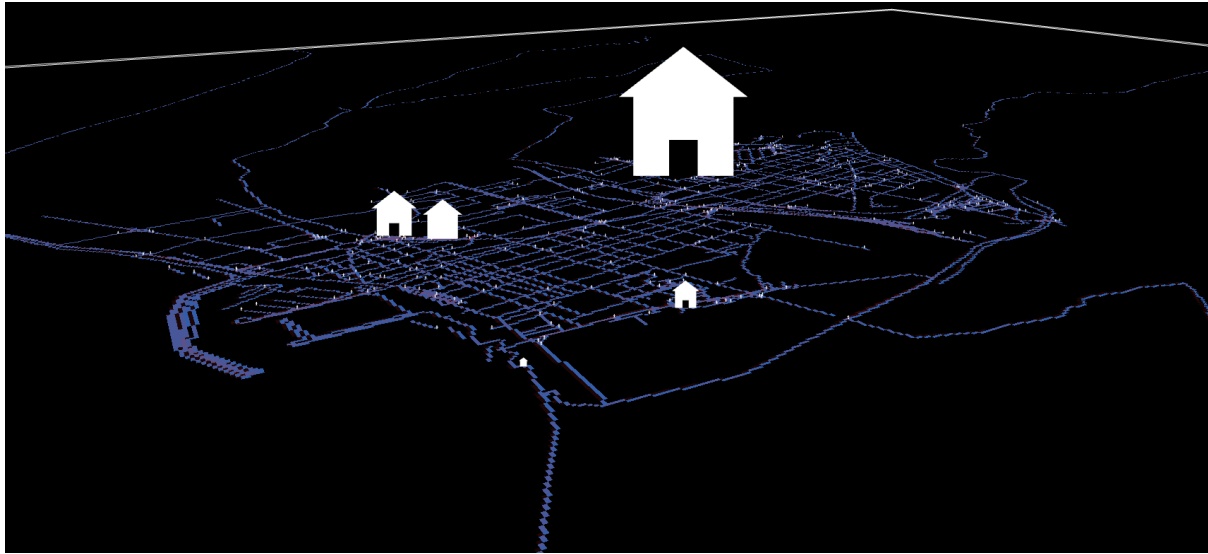


Figure 1: Visual representation of the simulated environment.

Time in the model advances in discrete steps, referred to as ticks, each representing a single day. During each tick, agents engage in mobility processes across the spatial network, followed by the possible occurrence of a hazardous event at a randomly determined moment. When such an event takes place, agents must attempt to reach a shelter within the constraints imposed by distance and capacity.

Within this simulated environment, we find citizens, modeled as autonomous agents. Each agent represents a citizen characterized by: a home location, the origin of citizens' movement and the place to which they return every night; a monetary capital that can be allocated to protective investments and is dynamically updated based on the outcome of the event, increasing in case of successful protection and decreasing otherwise; a sensitivity level to prosocial behavior.

The goal of the agents is to maximize expected utility. This therefore implies, on the one hand, the intention to avoid individual damage from critical events, and on the other hand, the intention to minimize the amount of resources invested in their own protection.

From a normative perspective, the optimal collective outcome would require agents to coordinate their contributions so as to ensure the minimum sufficient investment needed to provide protection for the entire population. However, the presence of free-riding behavior may lead some agents to rely on the contributions of others without investing themselves, while additional variability arises from the stochastic nature of the environment and the decentralized interaction of individual decision-making processes. These factors jointly make the emergence of stable cooperation a non-trivial outcome.

At the operational level, at the beginning of each round, every agent decides how much personal capital to invest in protective infrastructures. Each citizen is associated with a reference shelter, the geographically nearest one. Each agent can therefore decide to invest an integer amount of capital between 0 and 3, where 0 implies no willingness to cooperate in protection, 1 represents the minimum contribution to the local shelter, and 3 implies a clear overinvestment compared to what is strictly necessary.

It is worth noting that, if all agents invested exactly one unit, all shelters would be constructed. However, given the underlying mobility dynamics, this would still lead to uneven usage patterns, with

Table 1

Payoff structure as a function of investment level and outcome.

Is the agent safe?	Gain	Fair Gain	Gain0	Gain1	Gain2	Gain3
no	-2	-3.5	-2	-3	-4	-5
Yes	4	2.5	4	3	2	1
Yes, but there was an overinvestment	4.25	2.75	4.25	3.25	2.25	1.25

some shelters becoming overcrowded while others remain underutilized, ultimately preventing the entire population from being protected. Ideally, a slightly higher level of uniform investment would be sufficient to ensure full coverage. However, the presence of free-riders requires some agents to invest more to compensate for the lack of cooperation.

The model offers agents the possibility to cooperate, including a *shared contribution* mechanism. In other words, for the two highest levels (2 and 3), the agent can decide not to assign the amount exclusively to the local shelter, but to redistribute it proportionally across all shelters, according to the number of citizens associated with each one. This distinction is fundamental. Purely local investments represent self-protective cooperation within a neighborhood, whereas shared investments represent system-level cooperation aimed at improving resilience of the whole community.

The total investment accumulated by each shelter determines its operational state. Let n be the number of citizens associated with a specific shelter. Then three cases are possible:

- **Unavailable shelter:** if investment is lower than n , the shelter is not constructed and nobody can use it.
- **Efficient cooperation:** if investment lies between n and $2n$, the shelter becomes usable and can host a number of citizens proportional to the collected investment.
- **Overinvestment:** if investment exceeds $2n$, all potential users can be protected, but the excess resources are wasted.

This mechanism creates a social dilemma: underinvestment produces collective failure, whereas excessive uncoordinated generosity generates inefficiency.

At each simulation round, a hazardous event occurs at a random time. When the event occurs, agents react by attempting to reach the closest accessible shelter. Due to their spatial distribution and ongoing movement within the environment, the nearest shelter at that moment may differ from their reference one.

As previously stated, access to shelters is subject to capacity constraints, which are determined by the cumulative level of collective investment. At the same time, spatial proximity plays a crucial role: agents located closer to a shelter have a higher probability of reaching it before capacity is exhausted. In practice, admission follows an implicit first-come, first-served mechanism.

The payoff obtained by each agent depends on two main components:

1. the outcome of the event, i.e., whether the agent manages to reach a shelter and thus survives
2. the cost associated with the investment made during the previous decision phase

Agents who successfully access a shelter receive a positive payoff, whereas those who fail to do so incur a negative outcome. In both cases, the cost of the investment is deducted from the resulting payoff, capturing the trade-off between individual expenditure and collective benefit.

Table 1 summarizes the payoff structure. In the table, *gain* denotes the gross payoff received by an agent regardless of the amount invested, while *fair gain* represents the normalized payoff each agent would obtain when the total investment is evenly distributed across the population. We also report the actual payoff for each choice, net of the associated cost. It is easy to observe that choosing to invest 0 potentially yields the highest payoff. The more an agent invests, the lower the payoff becomes.

When shelters are overfunded, the resulting inefficiency is reflected in a reduced payoff for agents, with respect to the actual investment, capturing the effect of resource waste.

Overall, the reward represents the trade-off between individual cost and collective protection, and is used as feedback for subsequent learning.

3.2. Learning and Nudging

Agents' learning is modeled through a contextual multi-armed bandit approach[34, 35]. The context is defined by the state of the agent's reference shelter in the previous round: underinvestment; efficient cooperation; overinvestment.

At each decision step, the agent observes the current context and selects an action based on the corresponding row of the utility matrix. The action with the highest expected utility is chosen, with a fixed probability of random exploration to allow for strategy diversification.

After the event, the agent updates the expected utility associated with the selected context–action pair. The update is based on the difference between the obtained reward and the previously expected utility:

$$U_i(c, a, t + 1) = U_i(c, a, t) + \eta (R_i(t) - U_i(c, a, t)) \quad (1)$$

where η is the learning rate and $R_i(t)$ is the realized reward.

This core version of the model is further extended modeling the influence of the global outcome of the population. The focus here is on modeling agents' sensitivity to others' losses (such as observing others fail to reach protection or incur losses, even when the agent itself is safe) which can be interpreted as a form of *prosocial behavior* [36, 37] toward individuals at risk of exclusion. In our specific case, agents become more sensitive to collective risk and are more likely to adopt cooperative strategies, including shared investments across shelters.

When many agents fail to survive, the reinforcement signal is adjusted to penalize strategies that contribute to poor collective performance. This signal simulates a non-coercive intervention by the authority, informing agents about the negative collective outcome (e.g., many citizens failed to survive). Importantly, nudging does not directly constrain agents' choices. Instead, it modifies the learning dynamics by introducing a social component into the reinforcement signal. In practice, when nudging is active, agents partially take into account the overall success or failure of the population, not just their individual outcome.

The modeling of prosocial behavior is inspired by the framework proposed by [38], which introduces a parameter capturing proactive altruism, balancing individual payoff against collective welfare. In this setting, the parameter γ represents the extent to which an agent weights its own utility relative to the overall well-being of the population.

Operationally, the prosocial effect is incorporated into the agents' learning process as a discounting factor: when the community experiences significant losses, the individual utility of the selected action is reduced accordingly. This mechanism allows agents to internalize collective outcomes within their decision-making dynamics.

Let $U_i(c, a, t)$ be the expected utility of action a in context c for agent i . First, a standard update based on individual reward is computed:

$$U'_i(c, a, t + 1) = U_i(c, a, t) + \eta (R_i(t) - U_i(c, a, t)) \quad (2)$$

Then, the final utility is obtained by combining the individual update with a global signal:

$$U_i(c, a, t + 1) = (1 - \gamma) U'_i(c, a, t + 1) + \gamma G(t) \quad (3)$$

where:

- $\gamma \in [0, 1]$ represents the sensitivity of the agent to social feedback;
- $G(t) \in \{0, 1\}$ is a global performance indicator.

In particular, $G(t) = 0$ when the fraction of agents that failed to reach protection exceeds a predefined threshold, and $G(t) = 1$ otherwise.

As a consequence, when collective performance is poor, the expected utility of the selected action is systematically reduced. This mechanism discourages strategies that, while possibly beneficial at the individual level, lead to unfavorable outcomes at the population level.

Conversely, when most agents successfully reach protection, the global signal reinforces the current strategy. This simple feedback mechanism allows the model to capture how informational nudging can align individual learning with collective objectives.

Over time, this process leads agents to adapt their investment strategies to recurring patterns in the environment, balancing self-interest with emerging collective behavior.

4. Results

The purpose of the simulation is to investigate the emergence of collective cooperation under varying conditions, with a particular focus on the impact of free-riding behavior and informational nudging mechanisms.

Experiments were carried out within a synthetic urban environment that reproduces a real-world road network. The main simulation parameters are summarized below:

- Number of agents: 14,000
- Number of shelters: 5
- Nudge threshold: 5% of the population
- Learning rate: 0.1
- Exploration probability: 20%
- Number of simulated events: 2,000

These parameters were not intended to provide a calibrated representation of a specific real flood event. Rather, they were selected to define a controlled exploratory scenario in which the effects of two theoretically relevant dimensions could be isolated: the proportion of persistent free-riders and the intensity of informational nudging. The remaining parameters were kept fixed in order to reduce the dimensionality of the experimental space and make the interpretation of the observed dynamics more transparent.

Two key dimensions were systematically explored:

- *Percentage of fixed free-riders*: 0%, 10%, 20%, 30%. These agents permanently adopt a non-cooperative strategy, exploiting the contributions of others without participating themselves. The rest of the population must adapt to their presence, attempting to compensate for the lack of cooperation and maintain a viable collective outcome. Failure to do so results in the inability to construct shelters, exposing all agents to the consequences of adverse events.
- *Nudging intensity*: 0%, 10%, 20%, 30%. This parameter determines the degree to which agents are influenced by global informational signals. Higher values correspond to greater sensitivity to collective feedback.

The percentage of fixed free-riders and the nudging intensity were therefore selected as the main experimental variables because they directly correspond to the central research question of the paper: whether and under which conditions informational feedback can mitigate the destabilizing effect of non-cooperative behavior. Other parameters, such as the number and spatial distribution of shelters, the learning rate, the exploration probability, and the threshold used to activate the global signal, are expected to influence the quantitative values of the outcomes. However, in this work they are treated as baseline assumptions, while their systematic variation is left to future sensitivity analyses.

Each experimental configuration was repeated over 20 independent simulation runs in order to reduce stochastic variability.

Cooperation is evaluated through the average shelter construction failure rate, defined as the proportion of scenarios in which agents fail to coordinate effectively and remain unprotected. Lower values indicate higher levels of cooperation and increased collective resilience.

Table 2 summarizes the results obtained under the different experimental conditions.

Table 2

Shelter construction failure rate as a function of the percentage of fixed free-riders and nudging intensity.

Fixed free-riders (%)	No nudge	10% nudge	20% nudge	30% nudge
0%	0.76	18.24	6.13	0.69
10%	3.50	27.13	12.54	2.14
20%	10.42	68.89	20.08	0.99
30%	60.93	65.09	27.34	5.29

Interestingly, although trust is not explicitly encoded in the model, the simulations reveal the emergence of trust-like dynamics as an endogenous property of agent interactions. Rather than relying on predefined cognitive structures, agents progressively adapt their strategies based on past outcomes and observed behaviors within the population. This adaptive process leads to the spontaneous formation of trust and distrust patterns [39], which in turn influence the system’s capacity to sustain cooperation. In this sense, trust is not treated as a static attribute but as a dynamic cognitive process shaped by experience and social feedback [40]. Agents implicitly learn when cooperation is advantageous and when disengagement becomes the rational choice.

This perspective allows us to distinguish between different types of non-cooperative behavior. On the one hand, opportunistic agents (free-riders) consistently avoid contributing while benefiting from others’ efforts [41, 42]. On the other hand, the simulations reveal the presence of agents who initially cooperate but gradually withdraw when they perceive that their contribution is insufficient to compensate for widespread non-cooperation.

In the latter case, non-cooperation is not driven purely by selfish incentives, but rather by a form of collective disillusionment, namely a loss of confidence in the effectiveness of joint action. This phenomenon can be interpreted as an indicator of social fragmentation: when cooperation is no longer perceived as mutually beneficial or supported by shared expectations, agents tend to disengage, leading to reduced collective efficiency [43].

These dynamics are clearly reflected in the quantitative results shown in Table 2. In the baseline scenario, where no nudging mechanism is present, the system achieves high levels of cooperation in the absence of free-riders. Even with up to 20% of fixed free-riders, the population is still able to compensate for non-cooperative behavior, and shelters are successfully constructed in over 90% of cases.

However, when the proportion of free-riders reaches 30%, the system experiences a sharp transition. In this regime, shelters fail to be constructed in approximately 60% of the simulations, indicating a breakdown of collective coordination. This result suggests the existence of a critical threshold beyond which, without any form of intervention, cooperative behavior is no longer sustainable.

The introduction of informational nudging produces non-linear and, in some cases, counterintuitive effects. At low intensity levels (e.g., 10%), nudging may even degrade system performance compared to the baseline. In this regime, the signal is insufficient to stabilize cooperation, yet strong enough to interfere with the agents’ learning process. As a consequence, it introduces ambiguity in the perceived utility, leading to suboptimal decisions.

From a cognitive standpoint, weak nudging can distort the perceived trade-off between costs and benefits, potentially encouraging short-term opportunistic behavior or an overestimation of cooperation costs. This effect is consistent with findings in reinforcement learning, where weak or inconsistent signals may hinder convergence and reduce learning stability [44].

Moreover, low-intensity nudging may generate heterogeneous expectations among agents, leading to temporary misalignment and reduced coordination. Some agents adapt to the signal, while others do not, resulting in transient instability at the system level.

As the intensity of nudging increases, the system exhibits a qualitatively different behavior. At higher levels (e.g., 30%), the informational signal becomes strong enough to align expectations across agents and promote convergence toward cooperative strategies.

In this regime, nudging acts as an emergent coordination mechanism, capable of mitigating the negative effects of free-riders even at higher proportions. The increased sensitivity of agents to collective signals transforms nudging into a stabilizing input within the system. Notably, collective performance remains positive even when a substantial proportion of agents adopt a fixed free-riding strategy, indicating a strong degree of robustness of the system to non-cooperative behavior.

Additionally, successful cooperative outcomes reinforce agents' contribution strategies, generating a positive feedback loop. This cascading effect further strengthens cooperation and enhances the overall resilience of the system.

5. Conclusions

In this work, we investigated the emergence and stability of cooperative behavior in large-scale populations exposed to environmental risk, with particular emphasis on the role of informational nudging in social dilemma contexts.

To this end, we proposed a large-scale agent-based model that combines spatially grounded dynamics, collective-risk scenarios, and adaptive decision-making based on a contextual bandit learning framework. Within this setting, we introduced an informational nudging mechanism that provides agents with global feedback about collective outcomes, influencing their learning process without imposing explicit constraints on their choices.

The results highlight the intrinsically non-linear nature of collective cooperation [45, 46]. The system demonstrates a significant degree of robustness to free-riding behavior up to a critical threshold, beyond which cooperation rapidly deteriorates and the population fails to coordinate effectively. This behavior suggests the presence of tipping points in the dynamics of cooperation, where small changes in the proportion of non-cooperative agents can lead to large-scale systemic effects.

At the same time, the impact of informational nudging is shown to be highly sensitive to its intensity. Weak nudging signals may inadvertently disrupt coordination by introducing noise into the learning process and generating misaligned expectations among agents [47]. Conversely, when the signal reaches a sufficient strength, it acts as an effective coordination device, aligning individual behaviors and fostering the emergence of stable cooperative regimes even under unfavorable conditions.

These findings emphasize that informational nudging is not universally beneficial per se, but its effectiveness depends on both its design and the characteristics of the population [48, 49]. In particular, the results suggest that the presence of a minimum level of prosocial responsiveness among agents is a necessary condition for nudging to produce positive outcomes. In this sense, nudging mechanisms should be carefully calibrated and contextualized, especially in real-world applications involving heterogeneous populations.

More broadly, this work contributes to the understanding of how local interactions, learning dynamics, and global informational signals jointly shape collective outcomes in complex socio-environmental systems. By explicitly modeling the interplay between individual incentives and collective risk, the proposed framework provides a useful tool for exploring policy interventions aimed at enhancing community resilience.

Certainly, this work is not without limitations. The main objective of the study was to investigate whether and under which conditions informational nudging can affect cooperative behavior in the presence of free-riding dynamics. For this reason, we deliberately focused on a specific and controlled experimental setting, keeping several model parameters fixed in order to isolate the interaction between free-riding and nudging intensity. The results show that informational feedback can have a positive effect on collective outcomes, but also that this effect is highly non-linear and depends on the strength of the signal. At the same time, the present study does not provide a full sensitivity analysis of the entire parameter space. Parameters such as the number and spatial distribution of shelters, the activation

threshold of the global signal, the learning and exploration rates, and alternative payoff structures may influence the quantitative outcomes and should therefore be systematically investigated in future work.

Future research will extend the model along several directions. First, we plan to investigate the role of spatial and infrastructural factors, such as the number, distribution, and capacity of shelters. While increasing the availability of protective infrastructure is expected to improve accessibility and reduce exposure to risk, it may also introduce additional coordination challenges, potentially affecting the emergence of cooperative behavior.

Second, we aim to explore alternative investment and governance regimes, including purely private contribution schemes, centralized public investment, and hybrid approaches. This will allow us to analyze how different institutional configurations influence collective dynamics and resilience outcomes.

Finally, future work will consider the inclusion of richer behavioral dimensions, such as heterogeneous risk perception[50, 51], trust in information sources[52, 53], and dynamic social networks[54, 55]. Incorporating these aspects would enable a more realistic representation of human decision-making and provide further insights into the design of effective intervention strategies in real-world contexts.

Acknowledgments

This work has been supported by the project ICRAS - Interventi Comportamentali per la Resilienza a rischi Ambientali e Sanitari (MIUR-PNRR, PE00000005, CUP B83C22004820002) and by the project FAIR- Future Artificial Intelligence Research (MIUR- PNRR, PE00000013).

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

- [1] R. Bailie, Climate-related natural disasters: reflections on an agenda for rural health research, *International journal of environmental research and public health* 20 (2023) 5553.
- [2] M. Gao, Z. Wang, H. Yang, Review of urban flood resilience: insights from scientometric and systematic analysis, *International Journal of Environmental Research and Public Health* 19 (2022) 8837.
- [3] K. Zeuli, A. Nijhuis, R. Macfarlane, T. Ridsdale, The impact of climate change on the food system in toronto, *International journal of environmental research and public health* 15 (2018) 2344.
- [4] N. A. of Sciences, Medicine, D. on Earth, L. Studies, B. on Atmospheric Sciences, C. on Extreme Weather Events, C. C. Attribution, Attribution of extreme weather events in the context of climate change, National Academies Press, 2016.
- [5] L. A. Gammoh, I. G. Dawson, K. Katsikopoulos, How flood preparedness among jordanian citizens is influenced by self-efficacy, sense of community, experience, communication, trust and training, *International Journal of Disaster Risk Reduction* 87 (2023) 103585.
- [6] T. Grothmann, F. Reusswig, People at risk of flooding: Why some residents take precautionary action while others do not, *Natural hazards* 38 (2006) 101–120.
- [7] A. A. Bondarev, B. Hintermann, F. Krysiak, R. Winkler, The intricacy of adapting to climate change: Flood protection as a local public goods game (2017).
- [8] M. Kumar, V. Dutt, Understanding cooperative behavior against climate change through a public-goods game, *Climate Change* 1 (2015) 68–71.
- [9] C. Castelfranchi, R. Falcone, Trust theory: A socio-cognitive and computational model, John Wiley & Sons, 2010.
- [10] A. Sapienza, R. Falcone, Flood risk and preventive choices: a framework for studying human behaviors, *Behavioral Sciences* 14 (2024) 74.

- [11] R. Falcone, A. Sapienza, An agent-based model to assess citizens' acceptance of covid-19 restrictions, *Journal of Simulation* 17 (2023) 105–119.
- [12] A. Sapienza, F. Cantucci, C. Castelfranchi, R. Falcone, Trust evolution in agent and multi-agent systems: a computational modeling perspective, in: *The Agents Journey: Twenty-Five Years of Multi-agent Systems*, Springer, 2026, pp. 79–102.
- [13] R. H. Thaler, C. R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*, Penguin, 2009.
- [14] R. Hertwig, T. Grüne-Yanoff, Nudging and boosting: Steering or empowering good decisions, *Perspectives on Psychological Science* 12 (2017) 973–986.
- [15] C. Schubert, Green nudges: Do they work? are they ethical?, *Ecological economics* 132 (2017) 329–342.
- [16] V. Capraro, G. Jagfeld, R. Klein, M. Mul, I. v. de Pol, Increasing altruistic and cooperative behaviour with simple moral nudges, *Scientific reports* 9 (2019) 11880.
- [17] R. Böhm, Ö. Güreker, T. Lauer, Nudging climate change mitigation: a laboratory experiment with inter-generational public goods, *Games* 11 (2020) 42.
- [18] C. T. DesRoches, P. Bartha, K. Mintz-Woo, A. Rodriguez, D. Steel, Climate nudging, catastrophes, and cost benefit analysis, *European Journal for Philosophy of Science* 16 (2026) 39.
- [19] M. Wooldridge, *An introduction to multiagent systems*, John Wiley & sons, 2009.
- [20] F. Luino, M. Chiarle, G. Nigrelli, A. Agangi, M. Biddoccu, C. Cirio, W. Giulietto, A model for estimating flood damage in Italy: preliminary results, *WIT Trans. Ecol. Environ* 98 (2006) 65–74.
- [21] C. Scawthorn, N. Blais, H. Seligson, E. Tate, E. Mifflin, W. Thomas, J. Murphy, C. Jones, Hazus-mh flood loss estimation methodology. i: overview and flood hazard characterization, *Natural hazards review* 7 (2006) 60–71.
- [22] L. Tierolf, H. de Moel, J. van Vliet, Modeling urban development and its exposure to river flood risk in southeast Asia, *Computers, Environment and Urban Systems* 87 (2021) 101620.
- [23] N. Gentle, S. Kierce, A. Nitz, Economic costs of natural disasters in Australia, *The Australian Journal of Emergency Management* 16 (2001) 38–43.
- [24] N. C. Bronfman, P. C. Cisternas, E. López-Vázquez, L. A. Cifuentes, Trust and risk perception of natural hazards: implications for risk preparedness in Chile, *Natural hazards* 81 (2016) 307–327.
- [25] J. Whittaker, J. Handmer, Community bushfire safety: a review of post-black Saturday research, *Australian Journal of Emergency Management*, The 25 (2010) 7–13.
- [26] H. Van Dyke Parunak, R. Savit, R. L. Riolo, Agent-based modeling vs. equation-based modeling: A case study and users' guide, in: *International workshop on multi-agent systems and agent-based simulation*, Springer, 1998, pp. 10–25.
- [27] H. Rahmandad, J. Sterman, Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models, *Management science* 54 (2008) 998–1014.
- [28] S. P. Simonovic, S. Ahmad, Computer-based model for flood evacuation emergency planning, *Natural Hazards* 34 (2005) 25–51.
- [29] R. J. Dawson, R. Peppe, M. Wang, An agent-based model for risk-based flood incident management, *Natural hazards* 59 (2011) 167–189.
- [30] Y. Li, B. Hu, D. Zhang, J. Gong, Y. Song, J. Sun, Flood evacuation simulations using cellular automata and multiagent systems—a human-environment relationship perspective, *International Journal of Geographical Information Science* 33 (2019) 2241–2258.
- [31] R. Falcone, A. Sapienza, How can subjective impulsivity play a role among information sources in weather scenarios?, in: *WOA*, 2017, pp. 19–24.
- [32] R. Falcone, A. Sapienza, C. Castelfranchi, Which information sources are more trustworthy in a scenario of hydrogeological risks: a computational platform, in: *Advances in Practical Applications of Scalable Multi-agent Systems. The PAAMS Collection: 14th International Conference, PAAMS 2016, Sevilla, Spain, June 1-3, 2016, Proceedings*, Springer, 2016, pp. 84–96.
- [33] U. Wilensky, I. Evanston, *Netlogo: Center for connected learning and computer-based modeling*, Northwestern University, Evanston, IL 4952 (1999).
- [34] D. Bouneffouf, I. Rish, C. Aggarwal, Survey on applications of multi-armed and contextual bandits,

- in: 2020 IEEE congress on evolutionary computation (CEC), IEEE, 2020, pp. 1–8.
- [35] T. Lu, D. Pál, M. Pál, Contextual multi-armed bandits, in: Proceedings of the Thirteenth international conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 485–492.
 - [36] J. Andreoni, Impure altruism and donations to public goods: A theory of warm-glow giving, *The economic journal* 100 (1990) 464–477.
 - [37] C. D. Batson, D. A. Lishner, E. L. Stocks, D. A. Schroeder, W. G. Graziano, The empathy-altruism hypothesis, *Positive psychology* (2022) 12–32.
 - [38] G. Charness, M. Rabin, Social preferences: Some simple tests and a new model (2000).
 - [39] R. Falcone, A. Sapienza, F. Cantucci, C. Castelfranchi, To be trustworthy and to trust: The new frontier of intelligent systems, *Handbook of Human-Machine Systems* (2023) 213–223.
 - [40] R. Falcone, C. Castelfranchi, Social trust: A cognitive approach, in: *Trust and deception in virtual societies*, Springer, 2001, pp. 55–90.
 - [41] G. Cullity, The Free Rider Problem, in: E. N. Zalta, U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*, Fall 2025 ed., Metaphysics Research Lab, Stanford University, 2025.
 - [42] G. Marwell, R. E. Ames, Economists free ride, does anyone else?: Experiments on the provision of public goods, iv, *Journal of public economics* 15 (1981) 295–310.
 - [43] E. Fehr, S. Gächter, Cooperation and punishment in public goods experiments, *American Economic Review* 90 (2000) 980–994.
 - [44] R. S. Sutton, A. G. Barto, et al., *Reinforcement learning: An introduction*, volume 1, MIT press Cambridge, 1998.
 - [45] L. Avilés, Cooperation and non-linear dynamics: an ecological perspective on the evolution of sociality, *Evolutionary Ecology Research* 1 (1999) 459–477.
 - [46] B. Pavard, J. Dugdale, The contribution of complexity theory to the study of socio-technical cooperative systems, in: *Unifying themes in complex systems: new research volume IIIB proceedings from the third international conference on complex systems*, Springer, 2006, pp. 39–48.
 - [47] C. R. Sunstein, Nudges that fail, *Behavioural public policy* 1 (2017) 4–25.
 - [48] L. Congiu, I. Moscati, A review of nudges: Definitions, justifications, effectiveness, *Journal of Economic Surveys* 36 (2022) 188–213.
 - [49] D. Hummel, A. Maedche, How effective is nudging? a quantitative review on the effect sizes and limits of empirical nudging studies, *Journal of behavioral and experimental economics* 80 (2019) 47–58.
 - [50] K. Papagiannaki, M. Diakakis, V. Kotroni, K. Lagouvardos, E. Andreadakis, Hydrogeological and climatological risks perception in a multi-hazard environment: The case of greece, *Water* 11 (2019) 1770.
 - [51] S. Fuchs, K. Karagiorgos, K. Kitikidou, F. Maris, S. Paparrizos, T. Thaler, Flood risk perception and adaptation capacity: a contribution to the socio-hydrology debate, *Hydrology and Earth System Sciences* 21 (2017) 3183–3198.
 - [52] R. Falcone, A. Sapienza, C. Castelfranchi, Trusting information sources through their categories, in: *International Conference on Practical Applications of Agents and Multi-Agent Systems*, Springer, 2015, pp. 80–92.
 - [53] B. Ryan, Establishing information seeking pathways in slow and flash floods, *International journal of disaster risk reduction* 31 (2018) 9–19.
 - [54] T. Haer, W. W. Botzen, J. C. Aerts, The effectiveness of flood risk communication strategies and the influence of social networks—insights from an agent-based model, *Environmental Science & Policy* 60 (2016) 44–52.
 - [55] R. Falcone, A. Sapienza, C. Castelfranchi, Recommendation of categories in an agents world: The role of (not) local communicative environments, in: *2015 13th Annual Conference on Privacy, Security and Trust (PST)*, IEEE, 2015, pp. 7–13.