Modeling Decision-Making with Will for Cooperation in Social Dilemmas

Yizhe Huang[*1, 2], **Bin Ling**[*3], **Song-Chun Zhu**[2, 1] **& Xue Feng** [✉2]
[1]School of Intelligence Science and Technology, Peking University
[2]State Key Laboratory of General Artificial Intelligence, BIGAI
[3]Law School, Peking University

arXiv:2605.08669v2 [cs.MA] 6 Jul 2026

Abstract

Standard rational actor models often attribute cooperation failures in social dilemmas to insufficient incentives, overlooking the destabilizing effects of continuous utility maximization. To address this, we propose a framework of "will" defined as a mechanism that persistently pursues goals while ignoring local cost-benefit fluctuations. We formalize the Willed Agents as potential minimizers, distinguishing them from cumulative utility maximization. Dynamical analysis of infinite population demonstrates that willed agents shrink the feasible state space, acting as boundary constraints that accelerate convergence in canonical social dilemmas. Through multi-agent simulations in a spatiotemporal Stag Hunt Game, we show that willed agents function as "cooperation catalysts", enabling groups to surmount high-risk thresholds where purely utility maximization fails. We find that heterogeneous will strength promotes cooperation, and that agents who autonomously suspend rational re-evaluation can significantly outperform continuous optimizers. These findings suggest that successful cooperation relies on the cognitive capacity to strategically constrain calculation.

Keywords: agent-based modeling; dynamical analysis; Markov Game

Introduction

Social dilemmas arise when individually rational decision-making systematically produces collectively suboptimal outcomes (Hardin, 1968; Olson Jr, 1971; Schelling, 2006). Across classical economics and cognitive science, agents are typically modeled as efficient optimizers, such as utility maximizers in game theory and Bayesian planners in computational neuroscience, who continuously adjust behavior to maximize expected reward (C. Camerer, 2003; Simon, 1956). In social dilemmas, this commitment to scalar maximization takes the form of repeated local cost–benefit evaluation under uncertainty (C. F. Camerer et al., 2004; Van Huyck et al., 1990). A substantial body of work shows that such optimizing dynamics can amplify perceived risk and strategic caution, driving agents toward safe but collectively inferior equilibria, as in the Stag Hunt game, even when mutually beneficial coordination is feasible (Arthur, 1994; Hommes, 2013).

Most formal accounts of cooperation failures approach the problem by altering agents' incentives or the structure of their interactions. One class of approaches promotes cooperation by reshaping local payoffs or preferences, including punishment mechanisms (Fehr & Gächter, 2000; Köster et al., 2022), reputation systems (Michel-Mata et al., 2024; Nowak & Sigmund, 2005), and social preference models (Bogaert et al., 2008; Fehr & Schmidt, 1999; Hughes et al., 2018; Kong et al., 2024; Rabin, 1992). A second class stabilizes cooperation through institutional rules, governance structures, or centralized enforcement (Acemoglu, 2015; Greif, 2006; Ostrom, 2009). While effective in structured settings, institutional and governance-based approaches rely on externally imposed rules or centralized enforcement, which incur ongoing implementation and monitoring costs (North, 1990; Tirole, 2017).

Despite their differences, these approaches share a common premise: cooperation fails because agents compute or incentivize too little, and can therefore be remedied by additive computation. This perspective leaves largely unexplored the possibility that the problem lies not in insufficient optimization, but in the very reliance on continuous rational computation itself (Colman, 2003; McClennen, 1990). Accordingly, this paper asks whether a cognitive mechanism beyond the rational computation paradigm can enable groups to escape suboptimal non-cooperative equilibria without altering the underlying payoff structure.

To address this question, we introduce the concept of *will*. Rooted in political philosophy and social theory, *will* is a way of explaining how intentions, commitments, and collective expectations persist over time. At the individual level, *will* refers to an agent's capacity to form and sustain intentions beyond the moment of choice, as emphasized in philosophical accounts of intention and agency (Bratman, 1987; Frankfurt, 2018). This capacity resonates with psychological research on implementation intentions and action control, which demonstrates that individuals who form specific plans exhibit reduced deliberation and increased persistence in goal pursuit, even when immediate incentives favor alternative actions (Brandstätter et al., 2001; Gollwitzer, 1999; Heckhausen & Heckhausen, 2018). Such findings suggest that the deliberate suspension of ongoing cost–benefit evaluation is not merely a philosophical construct but an empirically documented feature of human cognition. At the social level, it describes how interdependent agents come to rely on one another's actions, allowing mutual expectations to stabilize and coordinated order to emerge (Hobbes & Popiel, 2004; Rousseau & Gourevitch, 1997). Building on this dual perspective, we conceptualize *will* as a cognitive mechanism. Once a target is selected, a willed agent persistently pursues states that bring it closer to the target over

[*]Equal contribution. ✉Corresponding author. The code is available at https://github.com/j-alpha7/willed-agent-sd.

time, regardless of cost-benefit fluctuation. Notably, *will* is not an alternative decision rule to classical models such as expected utility maximization, bounded rationality, or dual-process frameworks, but a complementary mechanism that operates on top of them.

Although decision research offers several constructs related to *will*, they remain fundamentally grounded in local cost–benefit evaluation. They promote behavioral persistence by anchoring goals, value signals, or choice sets, while leaving ongoing action continuously subject to reassessment as circumstances change. Commitment, for example, constrains future behavior by restricting the set of admissible choices in advance (Elster, 1984), yet the agent continues to evaluate which action to select among the remaining options at each decision point. Goal or intention stability maintains fixed goal representations over time (Braver, 2012; Georgeff et al., 1998), but the means of pursuing those goals remain open to ongoing cost–benefit recalculation. Motivation sustains the strength of value or reward signals (Kanfer, 1990; Ryan & Deci, 2000), yet the decision process still responds to momentary fluctuations in perceived costs and benefits. Persistence or grit operates at a descriptive level without specifying an underlying decision mechanism, characterizing sustained effort toward long-term goals despite obstacles (Duckworth & Gross, 2014). Crucially, deliberately suspending local recalculation is a documented cognitive strategy. Bounded rationality research shows humans strategically limit ongoing computation to reduce deliberative overhead and signal predictability to partners (Simon, 1956). This temporary commitment must be distinguished from zealot models that fix strategies irrespective of environmental structure (Cardillo & Masuda, 2019; Masuda, 2012), "cooperate without looking" that precludes strategic responsiveness (Hoffman et al., 2015), Bayesian Theory of Mind that relies on continuous mental-state inference (Kleiman-Weiner et al., 2025; Peysakhovich & Lerer, 2018; Shum et al., 2019), and virtual bargaining that coordinate behavior through mutual representations of joint plans (Colman, 2003). By contrast, *will* interrupts local evaluation: rather than recalculating at every step, a willed agent maintains a fixed course of action as circumstances unfold.

To the best of our knowledge, this paper provides the first mathematical formulation of *will* for multi-agent decision-making. Based on this formulation, the paper develops three lines of analysis. First, we provide a formal mathematical formulation of *will*, centered on the property that willed agents consistently pursue target states. Second, we analyze the interaction dynamics of an infinite population across three canonical social dilemma paradigms (Stag Hunt game, Snowdrift game, and Prisoner's Dilemma), examining how *will* reshapes individual behavior and collective outcomes. Third, in a more complex spatiotemporal social dilemma, the Markov Stag Hunt game, the impact of *will* on cooperation is investigated in terms of the proportion of Willed Agents and the distribution of will strength. Results show that Willed Agents function as "cooperation catalysts", enabling groups to surmount high-risk thresholds where purely utility maximization fails, and that optimal performance relies on heterogeneous will strength within the population. Moreover, a parsimonious experiment for the Willed Agents with autonomously target generation shows that autonomously suspending rational re-evaluation can significantly outperform continuous optimizers. Overall, while *will* forgoes continuous reward optimization, it functions as a critical mechanism for stabilizing cooperation in social dilemmas, where agents face conflicting incentives and strategic uncertainty.

Mathematical Framework

We model the social environment as a **Markov Game with Will**, formalizing the distinction between agents driven by external rewards and Willed Agents driven by internal distance-minimizing constraints. The system is defined as the tuple $\mathcal{M} = \langle N, \mathcal{S}, \mathcal{A}, P, \mathbf{R}, \mathbf{G}, \mathbf{D}, T \rangle$:

- $N = \{1, \ldots, n\}$: The set of n agents.
- $\mathcal{S}$: The joint state space representing the global system configuration.
- $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$: The joint action space, where $\mathcal{A}_i$ is the set of available actions for agent i.
- $P : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$: The state transition probability function, where $P(\mathbf{s}' \mid \mathbf{s}, \mathbf{a})$ is the probability of reaching $\mathbf{s}'$ given current state $\mathbf{s}$ and joint action $\mathbf{a}$.
- $\mathbf{R} = \{R_1, \ldots, R_n\}$: Reward functions $R_i : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ for each agent i.
- $\mathbf{G} = \{G_1, \ldots, G_n\}$: Target state sets $G_i \subset \mathcal{S}$ for agent i.
- $\mathbf{d} = \{d_1, \ldots, d_n\}$: Pseudometrics $d_i : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\geq 0}$ describes the distance between two states.
- $T \in \mathbb{N}$: Episode length.

Will as Potential Minimization

To formalize "will," we define the Willed Agent by its direct minimization of distance to a goal, rather than by accumulated payoff. A Willed Agent i is governed by a potential field D_i:

$$D_i(\mathbf{s}) = \inf_{\mathbf{g} \in G_i} d_i(\mathbf{s}, \mathbf{g}), \tag{1}$$

where $d(\cdot, \cdot)$ is a pseudometric on the state manifold. Unlike rational agents that integrate scalar signals over time to **maximize cumulative extrinsic utility** $\sum_t R_i(s_t, \mathbf{a}_t)$, the Willed Agent operates via descent on the potential field to **minimize distance error** $D_i(\mathbf{s}_T)$, a bounded objective defined strictly by state space topology. Its decision rule greedily minimizes expected distance to G_i:

$$\pi_{will}(\mathbf{s}) = \arg\min_{a \in \mathcal{A}_i} \mathbb{E}_{\mathbf{s}' \sim P(\cdot \mid \mathbf{s}, a_i, \mathbf{a}_{-i})} [D_i(\mathbf{s}')]. \tag{2}$$

The Potential Attractor

A **Potential Attractor** emerges when a Willed Agent becomes physically "pinned" by environmental boundaries or interaction dynamics.

Definition: A state $\mathbf{s} \in \mathcal{S}$ is a Potential Attractor for agent i if:

$$\forall a_i \in \mathcal{A}_i, \quad \mathbb{E}_{\mathbf{s}' \sim P(\cdot|\mathbf{s}, a_i, \mathbf{a}_{-i})}[D_i(\mathbf{s}')] \geq D_i(\mathbf{s}). \tag{3}$$

Locally trapped in this potential well, an agent might fail to reach its exact goal ($D_i(\mathbf{s}) > 0$) if further progress requires coordinated joint actions. Inside a Potential Attractor, the Willed Agent acts as a **fixed boundary condition** in the phase space, resisting deviations from G_i.

Dynamical Analysis of Population with Will

Having formalized the Willed Agent as a potential minimizer, we analyze how these individual constraints scale to shape collective behavior. For tractability, we examine an infinite population ($N \to \infty$) making independent choices between two targets: G_1 (Cooperation) and G_2 (Defection).

Population Structure and Payoffs

Let the system state $x \in [0, 1]$ denote the total proportion of the population pursuing G_1. The population comprises two decision-making modes:

- **Willed Mode:** Agents constrained by Potential Attractors (Eq. 3). We define n_1 and n_2 as the fixed proportions rigidly committed to G_1 (Willed Cooperators) and G_2 (Willed Defectors), with $0 \leq n_1 + n_2 \leq 1$.
- **Rational Mode:** The remaining agents, $m = 1 - n_1 - n_2$, adapt strategies to maximize cumulative extrinsic reward.

Rewards $\mathbf{R}$ translate into frequency-dependent payoffs. Let $f_1(x)$ and $f_2(1-x)$ be the rewards for choosing G_1 and G_2, given the population shares x in G_1. The rational segment's evolutionary dynamics are driven by the payoff differential:

$$\Delta f(x) = f_1(x) - f_2(1-x). \tag{4}$$

The functional forms of f_1 and f_2 correspond to game paradigms.

Analysis in Three Game Paradigms

By remaining locked in Potential Attractors, Willed Agents restrict the accessible phase space, constraining the system state x to a **Feasible Region** $\Omega = [n_1, 1 - n_2]$.

Rational agents drive system dynamics by switching to higher-reward groups. We model this evolution via a Langevin equation subject to social noise σ:

$$dx_t = \alpha \cdot m \cdot \Delta f(x_t)dt + \sigma dW_t, \tag{5}$$

where α is the agent movement rate and W_t is a Wiener process representing choice volatility. Natural equilibria are defined as $\mathcal{E} = \{x^* \mid \Delta f(x^*) = 0\}$, with stable equilibria $\mathcal{E}_s = \mathcal{E} \cap \{x^* \mid (\Delta f)'(x^*) < 0\}$. However, overall stability is also bounded by Ω:

Theorem 1 *A state $x^* \in \Omega$ is a stable equilibrium if it satisfies one of the following:*

1. ***Interior Stability:*** $x^* \in (n_1, 1 - n_2)$ *and* $x^* \in \mathcal{E}_s$.
2. ***Lower Boundary Stability:*** $x^* = n_1$ *and* $\Delta f(n_1) < 0$.

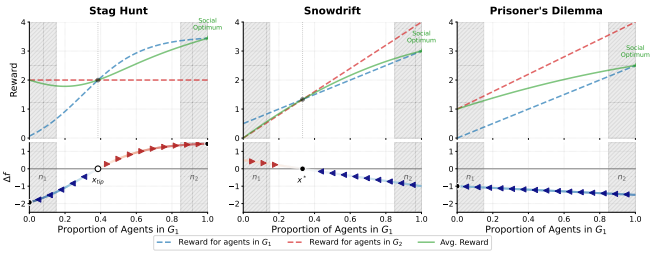

Figure 1: Willed Constraints in Canonical Games. **Top:** Payoff functions (f_1, f_2) and average reward for Stag Hunt Game, Snowdrift Game, and Prisoner's Dilemma. **Bottom:** The evolutionary gradient $\Delta f(x)$ and equilibrium points. Hatched regions indicate the excluded state space due to Willed Agents (n_1, n_2), defining the feasible region Ω.

3. ***Upper Boundary Stability:*** $x^* = 1 - n_2$ *and* $\Delta f(1 - n_2) > 0$.

We apply this framework to observe how n_1 and n_2 reshape three canonical games (Fig. 1).

The Stag Hunt Game In the Stag Hunt Game, increasing coordination yields higher rewards, creating stable equilibria at $x = 0$ and $x = 1$ separated by an unstable tipping point x_{tip}. The interaction between Ω and x_{tip} yields three regimes:

1. Degeneration into Coordination ($n_1 > x_{tip}$): If Willed Cooperators exceed the tipping point, x_{tip} is excluded from Ω. Because $\Delta f(x) > 0$ across Ω, the system inevitably converges to $x = 1 - n_2$, eliminating the coordination trap.

2. Degeneration into Dilemma ($1 - n_2 < x_{tip}$): If Willed Defectors push the Ω ceiling below x_{tip}, the cooperative basin becomes inaccessible. The game functionally degenerates into a Prisoner's Dilemma, collapsing to $x = n_1$.

3. Stochastic Acceleration ($n_1 \leq x_{tip} \leq 1 - n_2$): If x_{tip} remains within Ω, Willed Cooperators cannot deterministically force coordination. However, raising the starting floor to $x = n_1$ shortens the distance against the gradient, increasing the probability that social noise bridges the gap to x_{tip}. By Kramers' Rate Law (Kramers, 1940), the expected escape time τ drops exponentially as n_1 grows (Fig. 2):

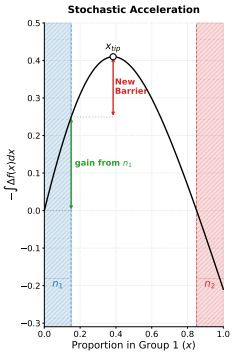

Figure 2: n_1 reduces the potential barrier, exponentially accelerating coordination.

$$\tau \propto \exp\left(-\frac{2}{\sigma^2}\int_{n_1}^{x_{tip}} \Delta f(x)dx\right) \tag{6}$$

The Snowdrift Game In the Snowdrift Game, $\Delta f(x)$ transitions from positive to negative, establishing a stable interior equilibrium x^*. Ω acts as a window strictly limiting outcomes:

1. Over-Cooperation ($n_1 > x^*$): If Willed Cooperators

exceed the optimal ratio, Rational Agents fully defect, pushing the system to $x = n_1$.

2. Under-Cooperation ($1 - n_2 < x^*$): If Willed Defectors restrict the ceiling below x^*, Rational Agents fully cooperate, and the system converges to $x = 1 - n_2$.

3. Neutral Substitution ($n_1 \leq x^* \leq 1 - n_2$): If x^* lies within Ω, final outcomes remain unchanged. Rational agents simply reduce their efforts in response to Willed Cooperators' fixed contributions, maintaining overall cooperation at x^*.

The Prisoner's Dilemma Because defection is strictly dominant ($\Delta f(x) < 0$) in the Prisoner's Dilemma , the gradient continuously drives the system downward. The upper boundary $1 - n_2$ is irrelevant, and the system collapses to the lower boundary $x^* = n_1$. Here, Rational Agents fully defect, while Willed Agents provide a stability floor and prevent total utilitarian collapse, albeit via their own exploitation.

Experiments

Experimental Setup

The infinite population analysis assumes that the movements of a single agent are instant and have a negligible impact on the group. However, real-world coordination occurs in finite populations with temporal constraints and spatial dynamics. In such settings, behavior is a dynamic process: individual actions are continuously observed by peers, actively shaping their belief estimations and judgments of optimal behavior over time.

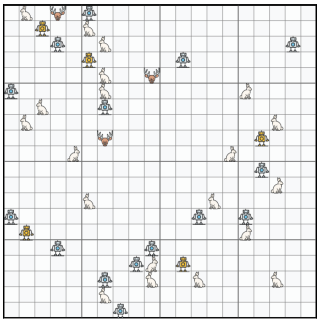

Figure 3: Overview of Markov Stag-Hunt Game.

To investigate the impact of Will in this context, we utilize a grid-world instantiation of the **Markov Stag Hunt Game** (Fig. 3). We simulate an $H \times W$ grid populated by N agents, N_h hares, and N_s stags. Each episode lasts T steps. At each step, agents act synchronously by selecting from six actions: {idle, left, right, up, down, hunt}. Hunting a hare is a unilateral action yielding an individual reward R_h. Hunting a stag is a cooperative action that succeeds only if at least θ agents are co-located with the target, yielding a shared total reward R_s. Each agent can hunt only one prey per episode. The difficulty and incentive for cooperation are parameterized by the coordination threshold θ and the effective individual share $\bar{R}_s = R_s/\theta$. Unless otherwise specified, our experiments use $H = W = 20$, $N = 20$, $N_s = 3$, $N_h = 20$, $R_h = 1$, and $\bar{R}_s = 5$.

Agent Implementation We instantiate the theoretical types defined in the previous section as distinct decision-making modes:

- **Willed Mode:** The agent operates as a potential minimizer (Eq. 1). The potential field is defined by the Manhattan distance d_M to the target prey: $D_i(\mathbf{s}) = \inf_{g \in G_i} d_M(p_i(\mathbf{s}), p_g)$, where $p_i(\mathbf{s})$ and p_g are the positions of agent i and prey g, respectively.
- **Rational Mode:** The agent maximizes expected extrinsic reward via model-based planning. First, it employs a Bayesian Theory of Mind (Baker et al., 2011) to infer peer goals from observed trajectories. Assuming a uniform prior $b^0_{ij}(g_j) = 1/(N_s + N_h)$, the belief b^t_{ij} over agent j's target g_j updates via $b^{t+1}_{ij}(g_j) \propto b^t_{ij}(g_j) Pr(a^t_j \mid \mathbf{s}^t, g_j)$, utilizing a Boltzmann likelihood $Pr(a_j \mid \mathbf{s}^t, g) \propto \exp(-\beta d_M(p_j(\mathbf{s}^t, a_j), p_g))$, where β is the rationality coefficient. Second, it performs K Monte Carlo simulations. In each simulation k, peer targets are sampled from the posterior $\mathbf{g}_{-i} \sim \prod_{j \neq i} b^t_{ij}(g_j)$, assuming peers follow the aforementioned Boltzmann policy. Finally, the agent estimates the value of hunting each prey g as $V^k_i(g) = \sum_{\tau=0}^{T-t} \gamma^\tau R^{k,\tau}_i$, where $R^{k,\tau}_i$ is the reward gained at future step τ. It selects the target that maximizes the average estimated reward $\sum_{k=1}^{K} V^k_i(g)/K$ (Huang et al., 2024). (Parameters used: $\beta = 10, K = 15, \gamma = 0.98$).

In complex sequential interactions, rational calculation is often computationally expensive and unreliable early in an episode due to high uncertainty and insufficient observations. Thus, agents can initially rely on the Willed Mode before switching to the Rational Mode. We formalize this transition using the **will strength** parameter $\alpha \in [-1, 1]$, which controls the temporal duration of the agent's commitment. For the first $|\alpha|T$ steps, the agent follows the Willed Mode; the sign determines the target type ($\alpha > 0$ for the Will to Stag, and $\alpha < 0$ for the Will to Hare). For the remaining steps, the agent switches to the Rational Mode. We manipulate the population **composition** (proportion of Willed Agents) and **will strength** (α) to examine how these constraints shape collective coordination.

Effect of Population Composition on Cooperation

We examine how the density of Willed Agents influences collective outcomes by fixing agents to specific decision-making modes: the Will to Stag ($\alpha = 1$), the Rational ($\alpha = 0$), and the Will to Hare ($\alpha = -1$). We analyze the normalized group payoff across varying cooperation thresholds θ while keeping the incentive for cooperation constant at $\bar{R}_s = 5$.

The Catalytic Role of Will Our infinite population analysis suggested that the Will to Stag benefits the group. To verify if this holds in a spatiotemporal environment, we isolate the interaction between Agents with Will to Stag and Rational Agents (Fig. 4a).

Results indicate that Willed Agents function as *cooperation catalysts*, though their efficacy is strictly context-dependent based on the difficulty θ. When cooperation demands are minimal ($\theta = 2$), increasing the number of Willed Agents (n_1) actually reduces payoff, as persistence imposes unnecessary

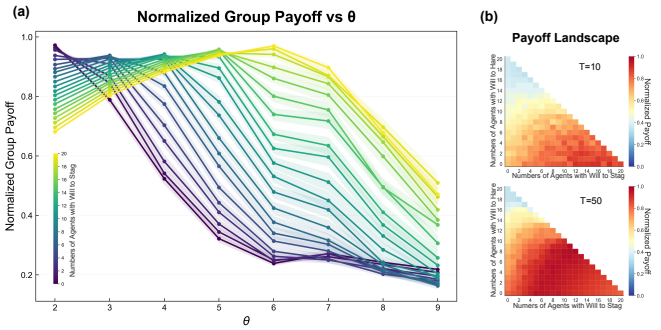

Figure 4: (a) Normalized group payoff vs. cooperation threshold θ for varying numbers of agents with Will to Stag. (b) Normalized group payoff landscape for different agent compositions at $\theta = 3$.

rigidity on a task that Rational Agents can easily solve. As difficulty increases, the presence of Willed Agents becomes beneficial. When $3 \leq \theta \leq 4$, this benefit is non-monotonic, where a small number improves performance but too many reduce it. When $5 \leq \theta \leq 6$, group payoff increases monotonically with n_1, and the gains from will are most pronounced. However, under extreme cooperation demands ($\theta \geq 7$), small additions of willed agents are insufficient; a critical mass is required to trigger any improvement in outcomes.

Collectively, these results show that **while will can enhance cooperation, it exhibits both critical mass effects and non-monotonicities depending on cooperation difficulty**.

Full Composition Analysis under Temporal Constraints We further analyze the full ternary composition space under varying time horizons ($T = 10$ and $T = 50$) with $\theta = 3$. As shown in Fig. 4b, shorter episodes ($T = 10$) require a substantially higher proportion of agents with Will to Stag to achieve high performance compared to longer episodes.

This result aligns with the analysis of the Stag Hunt Game in infinite population. Rational Agents require time to adjust beliefs. Given sufficient steps, Rational Agents can observe trajectories, infer peer intent, and align with the Willed Agents. However, under tight temporal constraints, the system relies on a critical density of the Will to Stag to "fast-track" the group toward the Stag equilibrium before the episode terminates.

Overall, **Willed Agents provide the initial momentum to overcome cooperation friction, which is crucial when the cooperation threshold is high or time is scarce**.

Effect of Will Strength on Cooperation

While the previous section established that the presence of Willed Agents catalyzes coordination, those agents operated with fixed, infinite persistence. We now investigate the impact of **will strength** ($\alpha \in [-1, 1]$), which governs the duration of an agent's commitment before reverting to rational mode. This allows us to examine how the intensity of individual "will" translates into collective outcomes.

Homogeneous Populations We first analyze populations where all agents share a uniform will strength α under tight temporal constraints ($T = 10$). As shown in Fig. 5, the relationship between will strength and group payoff is non-linear and strictly modulated by the coordination difficulty θ.

For the Will to Stag ($\alpha > 0$), we observe a distinct **inverted-U relationship** between will strength and performance at intermediate difficulties ($3 \leq \theta \leq 7$). A moderate level of will ($\alpha \approx 0.5$) effectively anchors agents to the cooperative equilibrium, overcoming initial coordination friction without inducing pathological stubbornness. However, as α approaches 1.0, the benefits diminish; excessive will induces functional rigidity, preventing agents from making necessary local adjustments, especially when the target prey is likely to be captured by other agents.

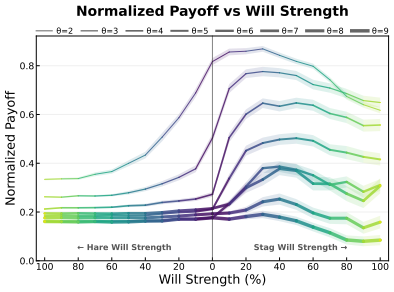

Figure 5: Normalized group payoff as a function of will strength α across coordination thresholds θ. $\bar{R}_s = 5, T = 10$. Shaded regions indicate 95% confidence intervals over 300 independent episodes.

At the extremes of coordination difficulty, the benefits of will vanish. When coordination is trivial ($\theta = 2$), rigidity becomes a liability, and rational agents can coordinate spontaneously and adjust your goals promptly based on the observation. and rational population ($\alpha = 0$) outperform population with $\alpha > 0.4$. Conversely, when coordination is severe ($\theta \geq 8$), the high requirement for simultaneous presence makes stag hunting statistically improbable. In these cases, the inflexible pursuit of stags yields lower returns than the rational baseline of safe hare-hunting. Finally, the Will to Hare ($\alpha < 0$) consistently underperforms or matches the baseline, confirming that persistence is evolutionarily advantageous only when directed toward high-risk, high-reward equilibria that rational agents fail to reach.

Heterogeneous Populations Biological and social systems are rarely homogeneous. To determine the evolutionarily stable distribution of will, we utilize a genetic algorithm to optimize the population's composition of α ($N = 10, N_s = 2, N_h = 10, \bar{R}_s = 10, T = 10$).

The results (Fig. 6a) reveal that the ideal group composition varies systematically with coordination difficulty. The optimal average will strength peaks at $\theta = 4$ with average $\alpha \approx 0.6$ before declining. When coordination is feasible, the optimal population exhibits high heterogeneity: high-α agents anchor the cooperative goal, creating a focal point, while lower-α agents retain the flexibility to adapt to stochastic dynamics. As θ increases beyond critical thresholds, the population converges toward homogeneity and rationality, abandoning the

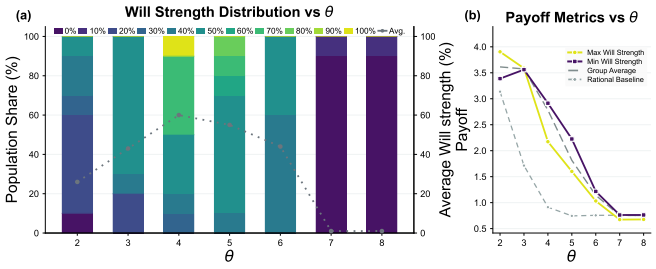

Figure 6: (a) Optimal will strength distribution across θ. Stacked bars indicate the population share at each strength level. The dashed line tracks the mean. (b) Individual payoff for maximal vs. minimal will strength, with dashed lines indicating optimized group and rational baseline payoffs.

high-risk stag strategy.

Fig. 6b highlights the cost of this specialization. At most coordination thresholds, agents with high will strength earn lower individual payoffs than their lower-will counterparts. High-will agents lower the energetic barrier for coordination but pay a fitness cost for their rigidity. This suggests that the Will to Stag functions as a form of altruistic commitment, enabling the group to achieve superior outcomes compared to a purely rational population, often at the expense of the willed individual's local utility.

Endogenous Will Selection

In previous experiments, target states were exogenously assigned. A fully autonomous agent, however, must internally generate the "will" to pursue a specific goal. We implement hybrid agents that utilize Monte Carlo Planning (Rational Mode) to identify a high-value target, then switch to Potential Minimization (Willed Mode) to greedily pursue it without re-planning. Let $k \in [1/T, 1]$ be the *Rational Ratio*, representing the proportion of timesteps allocated to re-evaluation. We compare two temporal integration strategies:

1. **Intermittent:** The agent engages in rational planning periodically (every $1/k$ steps) to update its target. In the interim steps, it acts as a Willed Agent committed to the most recently selected target. This mimics an agent that frequently "checks in" on the validity of its goal.
2. **Phased:** The agent confines rational planning to the initial phase of the episode (the first kT steps). Once this phase concludes, the agent "locks in" the final target and operates in Willed Mode for the remainder of the episode.

We evaluate these strategies under coordination difficulty $\theta = 4$ (Tab. 1). The results highlight a trade-off between flexibility and commitment. At equivalent rational ratios, the Phased strategy excels when cooperative rewards are low ($\bar{R}_s = 10$), as early commitment prevents abandoning marginal opportunities. Conversely, the Intermittent strategy performs better at high cooperative rewards, where false-positive commitments are costlier.

Table 1: Average normalized group payoff under endogenous goal selection ($\theta = 4$). "Instant" refers to an agent that plans only at $t = 0$ and maintains that commitment for the entire episode. Data averaged over 300 independent episodes.

Strategy	Rational Ratio	$\bar{R}_s = 10$	$\bar{R}_s = 50$
Pure Rational	100%	**0.607**$\pm$ 0.024	0.716$\pm$ 0.027
Intermittent	50%	0.577$\pm$ 0.023	0.700$\pm$ 0.027
Phased	50%	0.587$\pm$ 0.024	0.693$\pm$ 0.025
Intermittent	20%	0.521$\pm$ 0.022	0.687$\pm$ 0.025
Phased	20%	0.578$\pm$ 0.026	0.678$\pm$ 0.026
Intermittent	10%	0.505$\pm$ 0.019	0.724$\pm$ 0.025
Phased	10%	0.562$\pm$ 0.023	0.684$\pm$ 0.025
Instant	2%	0.510$\pm$ 0.019	**0.782**$\pm$ 0.024

Crucially, under high incentives ($\bar{R}_s = 50$), the Instant strategy (a single initial plan followed by unwavering execution) significantly outperforms the Pure Rational baseline. This challenges the assumption that more information processing yields better outcomes, as continuous re-evaluation makes rational agents hypersensitive to stochastic noise, often leading to the premature abandonment of joint goals.

While suspending rational planning undeniably serves as a computational cost-saving shortcut, its primary evolutionary and social function is far richer: **it actively signals predictability to peers.** By intentionally constraining its own cognitive flexibility and "binding its will" to a target, the agent transforms into a reliable focal point, effectively solving the coordination problem. This empirically validates the concept of *Resolute Choice* (McClennen, 1990): under high-stakes uncertainty, the optimal strategy is often to decide wisely once, and then refuse to reconsider.

Conclusion and Discussion

We have proposed a decision-making framework where social agents are driven by potential minimization rather than pure utility maximization. By constraining the feasible solution space, willed agents act as cooperation catalysts, enabling populations to traverse high coordination thresholds that purely rational groups fail to overcome.

Although "willed" behavior can positively impact society, its local fitness costs raise questions about its evolutionary survival. This persistence may be explained by cooperative mechanisms in evolutionary game theory (Nowak, 2006; Santos et al., 2006) and economic risk models, where some agents pursue high-risk, high-reward goals (Kahneman & Tversky, 2013; Schildberg-Hörisch, 2018).

Future research should explicitly investigate the evolutionary dynamics of willed agents across diverse spatiotemporal conditions. Additionally, integrating endogenous Will Selection with multi-agent reinforcement learning could explore how agents learn a meta-policy, dynamically regulating *when* to bind their will to balance the stability required for coordination against the flexibility needed for adaptation.

Acknowledgement

This work is supported by the National Science and Technology Major Project (No. 2022ZD0114900). This work is also supported by Beijing Natural Science Foundation (No. 4264117).

References

Acemoglu, D. (2015). Why nations fail? *The Pakistan Development Review*, *54*(4), 301–312.

Arthur, W. B. (1994). Inductive reasoning and bounded rationality. *The American economic review*, *84*(2), 406–411.

Baker, C., Saxe, R., & Tenenbaum, J. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. *Proceedings of the annual meeting of the cognitive science society*, *33*(33).

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British journal of social psychology*, *47*(3), 453–480.

Brandstätter, V., Lengfelder, A., & Gollwitzer, P. M. (2001). Implementation intentions and efficient action initiation. *Journal of personality and social psychology*, *81*(5), 946.

Bratman, M. (1987). Intention, plans, and practical reason.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in cognitive sciences*, *16*(2), 106–113.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton university press.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *The quarterly journal of economics*, *119*(3), 861–898.

Cardillo, A., & Masuda, N. (2019). Critical mass effect in evolutionary games triggered by zealots. *arXiv preprint arXiv:1912.00400*.

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and brain sciences*, *26*(2), 139–153.

Duckworth, A., & Gross, J. J. (2014). Self-control and grit: Related but separable determinants of success. *Current directions in psychological science*, *23*(5), 319–325.

Elster, J. (Ed.). (1984). *Ulysses and the sirens: Studies in rationality and irrationality*. Editions de la Maison des sciences de l'homme.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), 817–868.

Frankfurt, H. (2018). Freedom of the will and the concept of a person. In *Agency and responsiblity* (pp. 77–91). Routledge.

Georgeff, M., Pell, B., Pollack, M., Tambe, M., & Wooldridge, M. (1998). The belief-desire-intention model of agency. *International workshop on agent theories, architectures, and languages*, 1–10.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American psychologist*, *54*(7), 493.

Greif, A. (2006). *Institutions and the path to the modern economy: Lessons from medieval trade*. Cambridge University Press.

Hardin, G. (1968). The tragedy of the commons: The population problem has no technical solution; it requires a fundamental extension in morality. *science*, *162*(3859), 1243–1248.

Heckhausen, J., & Heckhausen, H. (2018). *Motivation and action*. Springer.

Hobbes, T., & Popiel, J. (2004). *Leviathan*. Barnes & Noble, Incorporated. https://books.google.com/books?id=zpcP4AV_jPAC

Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: Why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences*, *112*(6), 1727–1732. https://doi.org/10.1073/pnas.1417904112

Hommes, C. (2013). *Behavioral rationality and heterogeneous expectations in complex economic systems*. Cambridge University Press.

Huang, Y., Liu, A., Kong, F., Yang, Y., Zhu, S.-C., & Feng, X. (2024). Efficient adaptation in mixed-motive environments via hierarchical opponent modeling and planning. *International Conference on Machine Learning*, 20004–20022.

Hughes, E., Leibo, J. Z., Phillips, M., Tuyls, K., Dueñez-Guzman, E., García Castañeda, A., Dunning, I., Zhu, T., McKee, K., Koster, R., et al. (2018). Inequity aversion improves cooperation in intertemporal social dilemmas. *Advances in neural information processing systems*, *31*.

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part i* (pp. 99–127). World Scientific.

Kanfer, R. (1990). Motivation theory and industrial and organizational psychology. *Handbook of industrial and organizational psychology*, *1*(2), 75–130.

Kleiman-Weiner, M., Vientós, A., Rand, D. G., & Tenenbaum, J. B. (2025). Evolving general cooperation with a Bayesian theory of mind. *Proceedings of the National Academy of Sciences*, *122*(25), e2400993122. https://doi.org/10.1073/pnas.2400993122

Kong, F., Huang, Y., Zhu, S.-C., Qi, S., & Feng, X. (2024). Learning to balance altruism and self-interest based on empathy in mixed-motive games. *Advances in Neural Information Processing Systems*, *37*, 135819–135842.

Köster, R., Hadfield-Menell, D., Everett, R., Weidinger, L., Hadfield, G. K., & Leibo, J. Z. (2022). Spurious normativity enhances learning of compliance and enforcement behavior in artificial agents. *Proceedings of the National Academy of Sciences*, *119*(3), e2106028118.

Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *physica*, *7*(4), 284–304.

Masuda, N. (2012). Evolution of cooperation driven by zealots. *Scientific Reports*, *2*, 646. https://doi.org/10.1038/srep00646

McClennen, E. F. (1990). *Rationality and dynamic choice: Foundational explorations*. Cambridge university press.

Michel-Mata, S., Kawakatsu, M., Sartini, J., Kessinger, T. A., Plotkin, J. B., & Tarnita, C. E. (2024). The evolution of private reputations in information-abundant landscapes. *Nature*, *634*(8035), 883–889.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge university press.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *science*, *314*(5805), 1560–1563.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291–1298.

Olson Jr, M. (1971). *The logic of collective action: Public goods and the theory of groups, with a new preface and appendix* (Vol. 124). harvard university press.

Ostrom, E. (2009). *Understanding institutional diversity*. Princeton university press.

Peysakhovich, A., & Lerer, A. (2018). Prosocial learning agents solve generalized stag hunts better than selfish ones. *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, 2043–2044.

Rabin, M. (1992). Incorporating fairness into game theory.

Rousseau, J., & Gourevitch, V. (1997). *Rousseau: 'the social contract' and other later political writings*. Cambridge University Press. https://books.google.com/books?id=kcvseZCgQKMC

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54–67.

Santos, F. C., Pacheco, J. M., & Lenaerts, T. (2006). Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences*, *103*(9), 3490–3494.

Schelling, T. C. (2006). *Micromotives and macrobehavior*. WW Norton & Company.

Schildberg-Hörisch, H. (2018). Are risk preferences stable? *Journal of Economic Perspectives*, *32*(2), 135–154.

Shum, M., Kleiman-Weiner, M., Littman, M. L., & Tenenbaum, J. B. (2019). Theory of minds: Understanding behavior in groups through inverse planning. *Proceedings of the AAAI conference on artificial intelligence*, *33*(01), 6163–6170.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129.

Tirole, J. (2017). Economics for the common good.

Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, *80*(1), 234–248.