Manners Maketh MAN: Moral-Profile Diversity and Cooperative Dynamics in LLM-Based Multi-Agent Simulation

Keeheon Lee*
keeheon@yonsei.ac.kr
Yonsei University
Seoul, Republic of Korea

Kunhee Ryu
rgh00826@yonsei.ac.kr
Yonsei University
Seoul, Republic of Korea

Hogyun Yoo
yoohogyun@yonsei.ac.kr
Yonsei University
Seoul, Republic of Korea

ABSTRACT

How does the moral composition of a group shape cooperation in a shared space? We study this question with an LLM-based multi-agent simulation grounded in Moral Foundations Theory (MFT). Each agent is a persona-prompted Large Language Model whose moral priorities (Care, Fairness, Loyalty, Authority, and Purity) are instantiated from cross-national MFT survey scores. We use these scores as fixed, illustrative *moral profiles*; they are not intended to represent national populations or to support real-world grouping decisions. Agents maintain compressed episodic memories in which the same event can be framed through different moral priorities: for example, a dirty kitchen may be encoded as a harmony problem, a fairness imbalance, or a need for clearer procedure. A QMIX-based credit assignment module decomposes team reward into per-agent contribution signals, which are then fed back into the memory system. Across 15 four-agent profile compositions in a shared-dormitory environment, we find three simulation-level patterns. First, profiles with stronger behavioral restraint and commons-oriented maintenance achieve higher Nash Social Welfare. Second, explicit fairness monitoring can be associated with lower welfare when it amplifies complaints rather than contributions. Third, the same composition differences that affect welfare also affect credit-assignment stability: predictable low-externality behavior is easier for the QMIX module to decompose than complaint-driven, non-stationary behavior. We therefore present the framework as a sandbox for studying norm dynamics among artificial agents, not as a model of real human cultures.

KEYWORDS

Moral Foundations Theory; Large Language Models; Multi-Agent Simulation; Social Norms; Episodic Memory; Credit Assignment; Cooperative AI; AI Ethics

1 INTRODUCTION

Shared living spaces turn ordinary actions into social dilemmas. Leaving dishes in a sink, making noise late at night, or repeatedly occupying a bathroom are small individual choices, but they impose costs on others. People often disagree about such situations not because they lack rules, but because they attach different moral meanings to the same event: one person sees a hygiene problem, another sees unfair burden sharing, and a third sees a breakdown of agreed procedure.

The maxim "Manners maketh man," attributed to William of Wykeham in the fourteenth century, encodes a sociological insight: everyday norms do not merely reflect social order, but help constitute it [3]. In this paper, we operationalize "manners" as *internalized moral dispositions* that shape how artificial agents interpret and respond to shared-resource conflicts. The dormitory setting is used as a compact simulation environment because it contains familiar public-good, externality, and coordination problems. It is *not* proposed as a basis for assigning real roommates, designing housing policy, or ranking human cultures.

Prior computational work has studied norm emergence and cooperation in multi-agent systems [11, 12], but much of it assumes relatively homogeneous agents or treats heterogeneity as a behavioral parameter rather than as a difference in moral interpretation. We ask a narrower question: when artificial agents are given different MFT-based moral profiles, how does profile composition affect simulated cooperation, communication, and credit assignment?

We address this question using LLM-based multi-agent simulation grounded in Moral Foundations Theory (MFT) [5, 7]. Each agent is a persona-prompted Large Language Model whose moral priorities (Care, Fairness, Loyalty, Authority, and Purity) are initialized from cross-national survey scores. To reduce the risk of essentializing real populations, we refer to these as *JP-profile*, *US-profile*, and *UK-profile* agents: the labels indicate score vectors used as empirical anchors, not claims about Japanese, American, or British individuals. Two mechanisms drive adaptation. First, a *compressed episodic memory* stores action history, contribution estimates, and short lessons whose wording depends on the agent's moral profile. Second, a *QMIX-based credit assignment module* [16] decomposes team reward into per-agent contribution scores, which become part of each agent's memory. The LLM parameters remain frozen throughout; only the credit assignment module is trained.

Our investigation is guided by three research questions:

- **RQ1 (Composition Effect):** How does moral-profile composition—both homogeneous and mixed—affect collective welfare in a shared-resource simulation?
- **RQ2 (Behavioral Mechanisms):** Which action patterns and communication dynamics explain welfare differences across profile compositions?
- **RQ3 (Cooperative Learnability):** How does profile composition affect reward equity and credit-assignment stability, and what does this reveal about the learnability of cooperative structure?

Our contributions are:

(1) A **memory-augmented LLM agent architecture** with MFT-conditioned experience interpretation, enabling in-context behavioral adaptation without LLM parameter updates.
(2) A **hybrid LLM–MARL framework** combining persona-prompted LLM decisions with QMIX credit assignment, where

*Corresponding author.

QMIX is used as a contribution signal generator rather than as the agents' policy learner.

(3) A **systematic simulation study** of 15 moral-profile compositions, showing how differences in behavioral restraint, communication, and fairness monitoring can change welfare and reward equity.

(4) An **ethical and methodological reframing** of culture-based simulation: the profiles are fixed experimental treatments inspired by MFT data, not representations of real populations, and the framework is not intended for discriminatory or institutional sorting applications.

2 BACKGROUND AND RELATED WORK

2.1 Cultural Moral Psychology

Moral Foundations Theory [5, 7] proposes that moral judgment draws on several recurring foundations: Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion, and Purity/Sanctity, with Liberty/Oppression proposed subsequently [6]. Cross-cultural work suggests that these foundations are widely present but differently emphasized across contexts [1, 5]. At the same time, national averages necessarily hide substantial within-culture variation. We therefore use survey scores only to instantiate fixed simulation profiles. We do not claim that an MFT score vector captures a culture, a nationality, or an individual's moral psychology.

Despite MFT's influence in moral psychology, it remains underused in computational modeling. Work on value alignment [4] and moral reasoning in LLMs [9] has explored related questions, but fewer studies use MFT profiles to control how multi-agent memories are interpreted over time. Our work focuses on this interpretive layer: not only what agents do, but how they encode what happened and what lesson they carry forward.

2.2 Social Norms and Collective Welfare

The emergence and enforcement of social norms has been studied in game-theoretic [2], evolutionary [14], and MARL frameworks [11, 12]. Köster et al. [11], for example, show that punishment of norm violations can improve welfare in certain multi-agent settings. We extend this line by studying agents that differ not only in behavior but also in their interpretation of the same event. In our setting, a dirty kitchen can become a Care/Purity problem, a Fairness problem, or a procedure problem depending on the profile used to summarize memory.

2.3 LLM-Based Social Simulation

LLMs are increasingly used as social simulation agents [10, 13, 15]. Park et al. [15] demonstrate emergent social behavior with persistent memory and reflection. Our work shares the emphasis on memory-driven behavior, but adds two constraints. First, personas are parameterized by explicit MFT scores. Second, memories are compressed through profile-specific moral framing. This makes the simulation suitable for probing how identical events can generate different future tendencies under different moral profiles. However, LLM agents are not human participants and do not possess human moral intuitions; their outputs are role-conditioned text generation and should be interpreted accordingly.

2.4 Credit Assignment in Cooperative MARL

Value decomposition methods address the credit assignment problem in cooperative settings. QMIX [16] learns a monotonic mixing of per-agent values conditioned on global state. We use QMIX not as the primary decision policy, but as a *contribution signal generator*: its per-agent credit estimates are written into the memory system, providing quantitative feedback on how an action affected team welfare. A central question is whether profile composition changes how stable this decomposition becomes.

3 METHODOLOGY

Our framework has four components (Figure 1): (1) a shared-dormitory simulation environment implemented as a PettingZoo AEC game, (2) LLM agents with MFT-grounded moral profiles, (3) a compressed episodic memory system that generates profile-specific lessons, and (4) a QMIX credit assignment module that provides per-agent contribution signals.

3.1 Terminology and Scope

Throughout the paper, *JP-profile*, *US-profile*, and *UK-profile* refer to three fixed MFT score vectors used as experimental conditions. They are derived from cross-national survey averages, but they should be read as synthetic agent profiles rather than as direct models of real national groups. This distinction is important: our goal is to study how moral-priority vectors affect artificial multi-agent dynamics, not to make empirical claims about cultures or people.

3.2 Simulation Environment

We implement the dormitory as a PettingZoo AEC environment with N=4 agents, 100 timesteps per episode (approximately four simulated days at hourly resolution), and partial observability.

State and observation. The global state (20D) comprises environmental variables—kitchen cleanliness ($C_k \in [0, 100]$), bathroom occupancy, noise level, and time of day—and per-agent variables: hygiene, energy, hunger, and stress (4D $\times$ 4 agents). Each agent observes an 11D partial view: four environmental variables, its own four internal states, and the three roommates' stress levels.

Actions. Agents choose among nine discrete actions: `Sleep` (0), `Eat` (1), `Shower` (2), `Study` (3), `Party` (4), `Clean_Kitchen` (5), `Clean_Bathroom` (6), `Complain` (7), and `Idle` (8). These actions create social dilemmas: cleaning is a public good, noise is a negative externality, and bathroom use is a contested resource. Invalid or unparsable LLM outputs are mapped to `Idle`.

3.3 MFT-Grounded LLM Agents

Each agent wraps GPT-4o-mini with a moral profile, an observation interface, and a memory prompt. The LLM is used only for action and message generation; its parameters are not updated.

Moral Foundation profiles. Agent personas are parameterized by MFT scores drawn from cross-national survey data [1, 5]. Table 1 reports the raw scores used in the simulation. We intentionally retain the country-derived labels because they identify the empirical

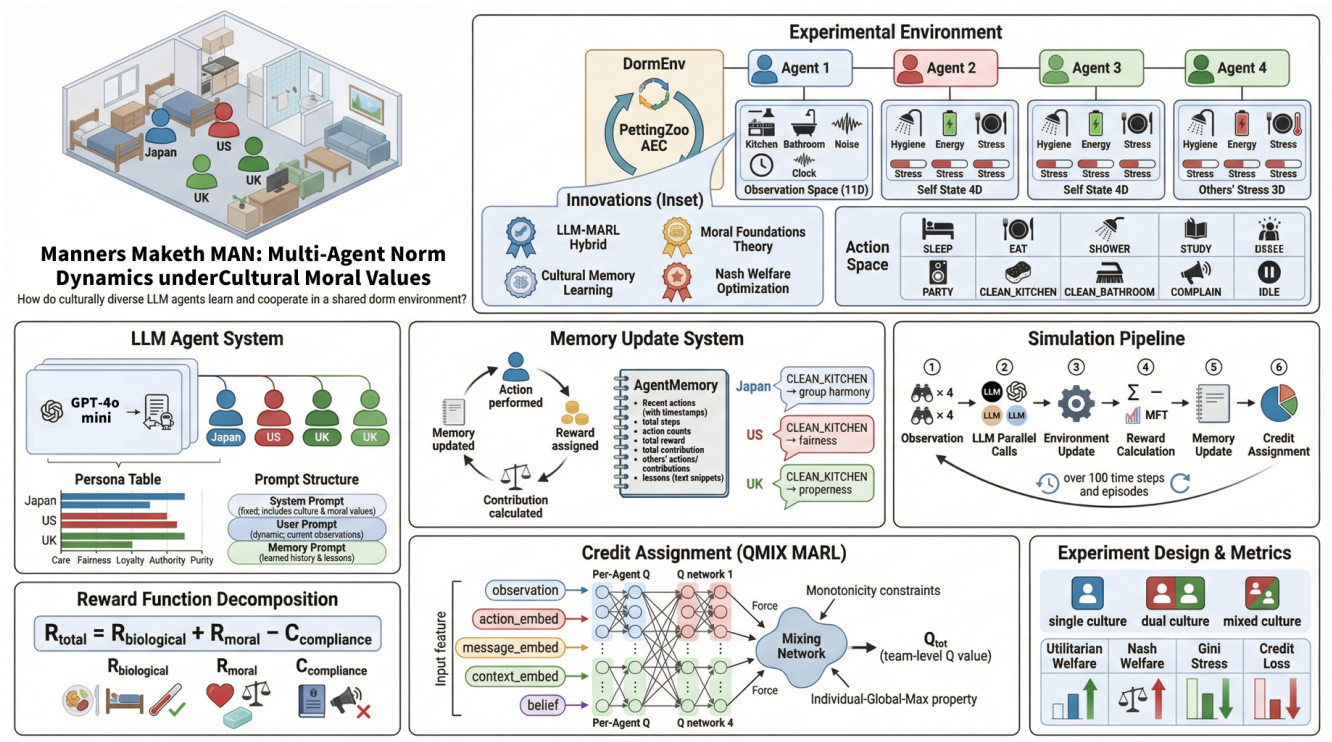

Figure 1: System architecture. Four LLM agents with fixed MFT-based moral profiles share a simulated dormitory. The PettingZoo AEC environment provides 11D partial observations and 9 discrete actions. GPT-4o-mini generates actions through a three-layer prompt consisting of persona, current observation, and compressed memory. The memory system stores action history, rewards, contribution scores, and profile-specific lessons. QMIX decomposes team value Q_{tot} into per-agent credit signals via monotonic mixing. We evaluate 15 profile compositions over 20 episodes using Nash welfare, mean reward, reward variance, message volume, and QMIX TD loss.

Table 1: Moral Foundation scores used to instantiate the three fixed simulation profiles. Labels indicate empirical anchors, not population claims.

Profile	Care	Fair.	Loy.	Auth.	Pur.
JP-profile	22.74	18.57	14.36	15.95	16.60
US-profile	22.00	22.00	15.00	15.00	12.00
UK-profile	21.00	20.00	16.00	15.00	14.00

source of the score vectors, but all results are stated at the level of profile behavior.

The profiles differ in relative emphasis. JP-profile has the highest Care, Authority, and Purity scores; US-profile has the highest Fairness score and the lowest Purity score; UK-profile is intermediate and has the highest Loyalty score among the three. These differences are used to condition both rewards and memory summaries.

Prompt architecture. Each agent receives a three-layer prompt. The *system prompt* contains the fixed profile description, MFT scores, available actions, and required JSON output format. The *user prompt* contains the current environmental state, the agent's internal state, roommates' stress levels, and recent messages. The *memory prompt* contains compressed history and profile-specific lessons. The LLM returns a JSON object with a short rationale, an action index, and an optional message. All four LLM calls are executed in parallel at each step.

3.4 Compressed Episodic Memory

The memory system enables in-context adaptation while bounding prompt length at approximately 700 tokens per agent. Each agent maintains four memory components:

(1) **Aggregate statistics**: total steps, per-action counts, total reward, and contribution rank.
(2) **Lessons learned**: up to five short natural-language lessons persisting across episodes.
(3) **Recent history**: the five most recent action–reward–contribution records.
(4) **Roommate observations**: per-roommate action counts, message counts, and approximate contribution estimates.

Memory updates occur at two timescales. At each timestep, the agent logs its action, reward, QMIX contribution estimate, and salient observed events. At the end of each episode, older records

Table 2: Illustrative memory interpretation. The same situation can be summarized differently because foundation weights select different event features.

Profile	Dominant framing	Example lesson
JP-profile	Care, Purity	"Cleaning the kitchen improved the shared environment. Reducing burden on others helps the group."
US-profile	Fairness	"Cleaning helped everyone, but contribution levels are uneven. Ask others to share the work."
UK-profile	Fairness, Loyalty, Authority	"Cleaning helped the group. A clearer rotation may prevent repeated imbalance."

are compressed into aggregate statistics and a small set of lessons. This prevents prompt length from growing with episode count.

3.5 Profile-Specific Memory Interpretation

The main interpretive mechanism is the generation of profile-specific lessons from shared events. This process is not derived from a separate "national character" theory. Instead, it is a deterministic operationalization of MFT weights within the simulation.

For each candidate event e, the system computes foundation-relevant features $\phi_m(e)$, such as whether the event reduced another agent's stress (Care), changed the equality of cleaning burden (Fairness), followed or violated an implicit rule such as quiet hours (Authority), or improved cleanliness (Purity). The event salience for an agent with profile c is:

$$S_c(e) = \lambda_q \cdot \widehat{q}(e) + \sum_{m \in \mathcal{M}} \widehat{w}_m^{(c)} \phi_m(e) + \lambda_r \cdot \text{recency}(e), \quad (1)$$

where $\widehat{q}(e)$ is the normalized QMIX contribution signal, $\widehat{w}_m^{(c)}$ is the normalized MFT weight for foundation m, and λ_q and λ_r weight contribution and recency. The top salient events are then rendered as short lessons using foundation-tagged templates. Thus, a "procedural" lesson is not an additional cultural construct; it is a descriptive shorthand for lessons in which Fairness, Loyalty, and Authority features make rule or rotation language salient.

This mechanism creates a feedback loop: QMIX evaluates actions, contribution estimates enter memory, memory is compressed through the profile's moral weights, and the next LLM decision is conditioned on the resulting lessons.

3.6 QMIX Credit Assignment

QMIX [16] serves as a contribution signal generator. It does not replace the LLM policy. Each agent's Q-network takes the agent's observation, selected action, message embedding, context embedding, and recurrent belief state as input and produces Q-values for the nine actions. The mixing network computes

$$Q_{\text{tot}} = f_{\text{mix}}(Q_1, \ldots, Q_4;\ s_{\text{global}}), \quad (2)$$

with non-negative mixing weights to enforce the monotonicity constraint and the Individual-Global-Max property. Training uses the temporal difference loss:

$$\mathcal{L}_{\text{TD}} = \mathbb{E}\left[\left(Q_{\text{tot}} - (r + \gamma \max_{\mathbf{a}'} Q'_{\text{tot}})\right)^2\right]. \quad (3)$$

For memory feedback, per-agent contribution is computed as the change in Q_{tot} attributable to an agent's action relative to a neutral baseline action. These contribution scores are not treated as ground truth moral judgments; they are approximate learning signals used to study whether different profile compositions produce more or less stable credit assignment.

3.7 Reward Structure

Agent i's reward at time t has two components:

$$R_i(t) = R^i_{\text{bio}}(t) + R^i_{\text{moral}}(t). \quad (4)$$

Biological reward. R_{bio} captures individual need satisfaction and penalties for unmet needs such as hunger, fatigue, poor hygiene, and stress. These terms ensure that agents cannot maximize group welfare by ignoring their own basic state.

Moral reward. R_{moral} is an MFT-weighted sum:

$$R^i_{\text{moral}}(t) = \sum_{m \in \mathcal{M}} w_m^{(c_i)} R^i_m(t), \quad (5)$$

where $w_m^{(c_i)}$ is agent i's profile-specific weight for foundation m. Care rewards reductions in others' stress and avoidance of harmful externalities; Fairness rewards more even contribution distributions; Purity rewards cleaner shared spaces; Authority and Loyalty reward compliance with shared routines and repeated cooperative commitments. The same environmental event can therefore have different reward salience for different profiles.

3.8 Outcome Metrics

We report five metrics, matching the quantities shown in the architecture figure and result tables.

Nash Social Welfare:

$$SW_{\text{Nash}} = \left(\prod_{i=1}^{N} (R_i - R_{\min} + \epsilon)\right)^{1/N}. \quad (6)$$

Nash welfare increases only when both total welfare and equity improve, making it appropriate for shared-resource settings.

Mean Reward:

$$\bar{R} = \frac{1}{N} \sum_i R_i. \quad (7)$$

All configurations produce negative mean rewards because communal living contains unavoidable biological and social costs; higher values therefore indicate better outcomes.

Reward Variance:

$$\text{Var}(R_i) = \frac{1}{N} \sum_i (R_i - \bar{R})^2. \quad (8)$$

Lower variance indicates a more equitable distribution of burden.

Message Volume: average number of messages sent per episode, used to compare implicit and explicit coordination.

Table 3: Moral-profile compositions evaluated in the simulation.

Type	Compositions
Homogeneous (3)	JP_4, US_4, UK_4
3+1 Split (6)	JP_3US_1, JP_3UK_1, US_3JP_1, US_3UK_1, UK_3JP_1, UK_3US_1
2+2 Split (3)	JP_2US_2, JP_2UK_2, US_2UK_2
2+1+1 Split (3)	$JP_2US_1UK_1$, $US_2JP_1UK_1$, $UK_2JP_1US_1$

Table 4: Main simulation settings.

Component	Setting
Agents per episode	4
Compositions	15
Episodes per composition	20
Timesteps per episode	100
Observation dimension	11 per agent
Global state dimension	20
Action space	9 discrete actions
LLM policy	GPT-4o-mini
LLM temperature	0.7
Memory persistence	Enabled across episodes
Memory budget	Approximately 700 tokens per agent
Lessons retained	Up to 5
Recent records retained	5 action–reward–contribution records
Invalid LLM action	Mapped to `Idle`
Trained component	QMIX credit assignment module only

QMIX Credit Loss: final TD loss of the mixing network, used as a measure of credit-assignment stability rather than as a direct measure of human-like cooperation.

4 EXPERIMENTAL DESIGN

4.1 Profile Compositions

We systematically vary moral-profile heterogeneity across 15 four-agent compositions:

Homogeneous groups establish profile-specific baselines. The 3+1 splits test whether a single minority profile changes a majority group's dynamics. The 2+2 and 2+1+1 splits test increasingly heterogeneous configurations.

4.2 Simulation Settings

All 15 compositions are evaluated over 20 episodes of 100 timesteps each. Memory persists across episodes via `keep_memory=True`, enabling cumulative in-context adaptation. The main settings are summarized in Table 4; this replaces the earlier appendix reference and keeps implementation-critical details in the main paper.

Table 5: Aggregate results for all 15 compositions over 20 episodes. Nash = Nash Social Welfare; $\bar{R}$ = mean reward; Msg = messages per episode; Loss = final QMIX credit loss.

Rk	Config	Nash	$\bar{R}$	Msg/Ep	Loss
1	JP_4	$45.52_{\pm 0.77}$	$-4.18_{\pm 0.78}$	13.2	33.9
2	JP_3UK_1	$42.51_{\pm 2.0}$	$-6.75_{\pm 1.75}$	15.8	111.8
3	US_3JP_1	$42.07_{\pm 3.6}$	$-7.50_{\pm 3.38}$	25.3	233.4
4	JP_3US_1	$41.78_{\pm 4.7}$	$-7.65_{\pm 4.03}$	19.5	292.8
5	UK_3JP_1	$41.36_{\pm 7.4}$	$-7.60_{\pm 5.88}$	22.8	252.2
6	US_2UK_2	$40.47_{\pm 2.6}$	$-8.68_{\pm 2.30}$	26.2	201.2
7	JP_2UK_2	$40.40_{\pm 3.8}$	$-8.68_{\pm 3.15}$	23.2	439.8
8	UK_4	$39.38_{\pm 7.6}$	$-9.55_{\pm 6.70}$	25.8	386.9
9	$JP_2US_1UK_1$	$38.80_{\pm 3.6}$	$-9.65_{\pm 2.63}$	21.5	322.3
10	JP_2US_2	$38.18_{\pm 2.4}$	$-10.55_{\pm 1.68}$	23.0	373.4
11	$US_2JP_1UK_1$	$36.62_{\pm 2.7}$	$-12.03_{\pm 2.45}$	31.0	512.8
12	$UK_2JP_1US_1$	$36.15_{\pm 4.7}$	$-11.65_{\pm 2.88}$	19.5	323.2
13	US_3UK_1	$33.63_{\pm 5.7}$	$-13.75_{\pm 4.08}$	27.5	694.7
14	US_4	$32.33_{\pm 5.9}$	$-16.13_{\pm 5.38}$	44.5	660.1
15	UK_3US_1	$31.97_{\pm 6.1}$	$-14.28_{\pm 2.95}$	22.5	430.2

4.3 Diagnostic Controls

We include three diagnostic controls to check whether the reported patterns depend on memory and profile conditioning:

- **Memoryless**: no accumulated history is inserted into the prompt.
- **Shuffled-profile**: MFT weights are randomly permuted across profile labels, probing whether behavior is driven by the supplied moral profile rather than the label alone.
- **Profile-blind memory**: the same lesson framing is used for all profiles, isolating moral framing from memory persistence.

These controls are not a complete ablation of LLM moral reasoning. In particular, they do not prove how much the model relies on numeric MFT scores versus textual persona descriptions, which remains a limitation.

5 RESULTS

All results below refer to simulated agents with fixed MFT-based profiles. They should not be read as claims about real national groups. We report aggregate outcomes first, then analyze behavioral mechanisms and credit assignment stability.

5.1 RQ1: Moral-Profile Composition and Collective Welfare

Table 5 shows that composition has a substantial effect on simulated welfare. Three patterns are most salient.

High-restraint profiles lead in this environment. JP_4 achieves the highest Nash Welfare (45.52), the smallest standard deviation (±0.77), and the best mean reward (−4.18). Compositions with three JP-profile agents also rank near the top. Within this dormitory environment, the JP-profile's combination of high Care, high Purity, and relatively strong Authority produces low-externality behavior that benefits the group.

Table 6: Nash Welfare by number of JP-profile agents in the group.

JP-profile agents	Mean Nash	n
4	45.52	1
3	42.15	2
2	39.13	3
1	39.05	4
0	35.56	5

High fairness monitoring alone is not sufficient. US_4 ranks 14th (Nash 32.33, mean reward −16.13). This does not imply that fairness is undesirable in general. Rather, under our prompt, memory, and reward design, the profile with the highest Fairness score often used fairness as a monitoring and complaint signal, while also producing more negative externalities. The result is a simulation-level interaction between Fairness emphasis, communication, and action choice.

JP-profile presence is associated with higher welfare. Aggregating by the number of JP-profile agents shows a monotonic relationship in these experiments:

The Spearman rank correlation between JP-profile count and Nash Welfare is strong ($\rho \approx 0.95$ when grouped by count), but the pattern should be interpreted as a property of this simulation. We describe the single-JP effect as a *coordination-anchor* effect: one predictable, low-externality agent can give others a stable behavior around which to coordinate. This should not be generalized to human populations or used for institutional selection.

Answer to RQ1: Moral-profile composition strongly affects simulated welfare. Profiles that combine commons maintenance, restraint, and predictable routines perform best in this shared-resource setting, while profiles that produce more negative externalities and more complaint-oriented communication perform worse.

5.2 RQ2: Behavioral Mechanisms and Communication

To understand the welfare differences, we examine action frequencies and message patterns in the homogeneous groups.

Negative externalities dominate positive contributions. The largest behavioral difference is Party frequency: JP-profile agents party at 0.6% of timesteps, UK-profile agents at 5.9%, and US-profile agents at 16.0%. Since Party increases noise, degrades sleep quality, and raises stress for roommates, this action has broad negative externalities. The top-ranked profile does not achieve high welfare by cleaning dramatically more; it avoids creating many of the problems that would require later repair.

Reactive cleaning is less efficient than restraint. US_4 devotes a higher proportion of actions to cleaning (5.8%) than JP_4 (4.6%), yet has lower welfare. This apparent paradox is explained by timing and context. In US_4, cleaning often appears reactive, compensating for prior environmental degradation. In JP_4, lower party frequency and more regular low-externality actions reduce the need for remedial cleaning. Thus, the efficiency advantage comes from fewer harmful actions, not simply more prosocial actions.

Self-regulation supports group welfare. JP-profile agents devote 41.5% of actions to Study, compared with 33.0% for US-profile agents and 34.9% for UK-profile agents. Combined with the highest Sleep rate, this produces a low-conflict routine that maintains individual wellbeing while avoiding shared-resource disruption. In this environment, self-regulation indirectly supports collective welfare.

The communication paradox. Message volume is negatively associated with welfare across the main results. JP_4 sends the fewest messages (13.2 per episode) and achieves the highest welfare, whereas US_4 sends the most messages (44.5 per episode) and ranks near the bottom. This is not evidence that communication is intrinsically harmful. The causal direction is ambiguous: conflict can increase messages, and messages can also escalate conflict. Qualitatively, however, the message types differ. JP-profile agents tend to send declarative contribution messages such as offering to clean. US-profile agents more often issue fairness demands, such as asking others to contribute more. UK-profile agents more often propose schedules or rotations. The harmful pattern is not communication itself, but complaint-oriented communication that increases stress without changing behavior.

Fairness as monitoring versus motivation. The Fairness result is best understood as a distinction between Fairness as a *monitoring* function and Fairness as a *contribution* function. In our simulation, the highest-Fairness profile frequently detects imbalances and comments on them. This can improve coordination when others respond constructively, but it can also create a grievance cycle when complaints raise stress and reinforce negative memories. Profiles with stronger Care/Purity salience instead tend to frame cleaning and restraint as intrinsically good for the shared environment, which produces fewer explicit disputes.

Memory amplifies initial tendencies. Profile-specific memory turns small early differences into persistent behavioral patterns. A Care/Purity-framed lesson records cleaning as supporting the shared environment; a Fairness-framed lesson records the same event as evidence of unequal burden; a rule-framed lesson records it as a reason to create a rotation. In the profile-blind memory diagnostic, this interpretive divergence is reduced, suggesting that memory framing rather than memory persistence alone contributes to the observed behavioral differences. This diagnostic does not fully isolate the effect of numeric MFT scores, but it supports the importance of the interpretive memory layer.

Answer to RQ2: Welfare differences are driven primarily by negative-externality avoidance and by how agents interpret contribution imbalances. Implicit coordination through predictable low-conflict actions can outperform explicit fairness negotiation when negotiation becomes complaint-oriented.

5.3 RQ3: Reward Equity and Credit Assignment

Reward distribution and focal blame. Table 7 shows that welfare differences are accompanied by large equity differences. JP_4 distributes rewards relatively evenly (variance 2.0). US_4 has much higher variance (54.8), with one agent receiving a substantially

Table 7: Per-agent mean reward and inter-agent variance for selected compositions. Lower variance indicates a more equitable distribution.

Config	Ag. 0	Ag. 1	Ag. 2	Ag. 3	Var.
JP_4	−6.4	−2.7	−3.2	−4.3	2.0
US_3JP_1	−9.3	−5.9	−6.0	−8.9	2.6
US_4	−11.0	−28.9	−11.4	−13.2	54.8
JP_2US_2	−3.4	−4.7	−12.6	−21.6	53.1

Table 8: Final QMIX credit loss after 20 episodes. Lower values indicate more stable credit assignment learning.

Config	Final Loss	Interpretation
JP_4	33.9	Stable
JP_3UK_1	111.8	Moderate
US_2UK_2	201.2	Elevated
US_4	660.1	Unstable
US_3UK_1	694.7	Unstable

worse reward than the others. We call this a *focal-blame* pattern: early stochastic differences in contribution can become encoded as fairness grievances, making one agent the repeated target of complaints. The targeted agent then accumulates stress, which further worsens its outcome.

The mixed composition JP_2US_2 shows a profile fault line: JP-profile agents receive relatively high rewards (−3.4, −4.7), while US-profile agents receive lower rewards (−12.6, −21.6). This indicates that mixing profiles can create inequity when agents differ in how they interpret burden sharing and how they respond to complaints. Again, this is a property of the artificial profiles and environment, not a claim about real groups.

Credit assignment stability. Table 8 shows a stark difference in credit-assignment loss. JP_4 reaches the lowest final loss (33.9), while US_4 and US_3UK_1 reach losses approximately 19 times higher. We interpret this as a learnability pattern: predictable routines and low-externality actions create a stable target for value decomposition, while complaint-driven and retaliatory dynamics create non-stationarity. Longer training horizons would be needed to distinguish fundamental unlearnability from slower convergence.

The learnability–autonomy tradeoff. The results reveal a tradeoff between behavioral flexibility and cooperative learnability. A profile that follows stable routines is easier for the credit assignment module to model, but may express less strategic variety. A profile that reacts strongly to perceived inequity may preserve more autonomy, but creates a less stable learning target for cooperative credit decomposition. This tradeoff matters for heterogeneous cooperative AI systems because agent diversity can increase representational richness while also making coordination harder to learn.

Answer to RQ3: Profile composition affects not only welfare but also the stability of credit assignment. Predictable low-externality profiles produce more equitable reward distributions and lower QMIX loss, whereas complaint-amplifying dynamics produce focal blame and unstable credit learning.

5.4 Mixed-Group Dynamics

The mixed-composition results suggest three interaction patterns.

Majority-profile stabilization. In JP_3US_1 (Nash 41.78) and JP_3UK_1 (Nash 42.51), the JP-profile majority maintains a low-externality routine, and the minority agent partially adapts to the group pattern. JP_3UK_1 performs slightly better than JP_3US_1, suggesting that the UK-profile's rotation-oriented messages are more compatible with the majority routine than complaint-oriented fairness demands.

Coordination anchor. US_3JP_1 (Nash 42.07) substantially outperforms US_4 (Nash 32.33). The single JP-profile agent tends to provide consistent maintenance behavior, reducing the instability that otherwise appears in the US_4 condition. We use the term *anchor* rather than *catalyst* to emphasize that the effect is a simulated coordination pattern, not a claim about a real cultural group.

Compatibility limits. UK_3US_1 (Nash 31.97) performs similarly to US_4 (Nash 32.33), with overlapping standard deviations. This suggests that a minority profile can fail to improve welfare when its action pattern and communication style are poorly aligned with the majority profile. We do not interpret this as "disruption" by a real-world culture; it is a compatibility result for fixed artificial profiles.

6 DISCUSSION

6.1 Behavioral Care versus Verbal Care

A central result is that similar Care scores can lead to different outcomes when combined with different surrounding foundations. JP-profile and US-profile have similar raw Care scores (22.74 vs. 22.00), but they behave differently because Care is interpreted through different adjacent priorities. In JP-profile, higher Purity and Authority make restraint, cleanliness, and routine salient. In US-profile, higher Fairness makes burden monitoring and verbal accountability salient. The result is not that one group of humans is more caring than another, but that the *form* of prosociality in an artificial agent matters: preventing harmful externalities and verbally objecting to inequity are different coordination strategies.

6.2 Implicit Coordination and Communication Bandwidth

The negative association between message volume and welfare suggests that more communication does not automatically improve multi-agent coordination. When agents already follow predictable low-conflict routines, few messages may be needed. When agents repeatedly complain about imbalance, communication can become part of the conflict dynamics. This observation is consistent with the distinction between high-context and low-context coordination styles [8], but our simulation does not establish a general cultural law. The safer design lesson is that communication protocols for cooperative AI should consider message content, timing, and stress effects, not only message capacity.

6.3 Fairness as a Double-Edged Signal

Fairness is a desirable value, but its operational form matters. In our environment, Fairness can motivate contribution, but it can also become a monitoring signal that amplifies blame. This is relevant to AI alignment because systems optimized for explicit equity checks may still fail to produce cooperative welfare if they do not also support repair, forgiveness, or low-conflict contribution. Fairness should therefore be modeled together with the interaction dynamics it induces.

6.4 Implications for Cooperative AI

The learnability results suggest that heterogeneous agent values can change the difficulty of credit assignment. Stable routines make cooperative structure easier to learn, while reactive complaint cycles make it harder. This does not imply that homogeneous or low-autonomy agents are always preferable. Rather, it suggests that cooperative AI systems need mechanisms that preserve diversity while preventing non-stationary blame dynamics, such as explicit repair protocols, shared commitments, or memory designs that avoid accumulating grievances without resolution.

7 ETHICAL SCOPE AND INTENDED USE

This work studies artificial agents, not human populations. The profile labels are empirical anchors for MFT score vectors, and the simulated results must not be interpreted as statements about Japanese, American, British, or any other real people. National averages do not capture within-culture variation, migration histories, individual differences, or contextual adaptation.

The dormitory environment is a research sandbox, not a deployment proposal. We explicitly reject using this framework to assign roommates, select team members, screen applicants, restrict exposure to diversity, or make institutional decisions about people on the basis of nationality, culture, or inferred moral profile. Such uses would risk discrimination and would exceed what LLM-based simulation can justify.

The LLM agents also do not possess moral intuitions. They generate text and actions conditioned on prompts, memory, and training data. The simulation can help formulate hypotheses about artificial norm dynamics, but it cannot validate claims about human moral behavior without human subjects data and careful ethical review.

8 LIMITATIONS AND FUTURE WORK

Profile simplification. The three profiles are fixed score vectors derived from national averages. This obscures within-culture variation and can invite overinterpretation. Future work should use continuous MFT parameter distributions, individual variation, and synthetic profile labels that are fully decoupled from national categories.

LLM persona fidelity. GPT-4o-mini may not consistently use numeric MFT scores as intended. The shuffled-profile and profile-blind memory diagnostics provide partial checks, but they are not sufficient to determine whether the model relies more on numbers, textual profile descriptions, or latent stereotypes from pretraining. A stronger ablation would remove labels, remove numeric scores, vary textual descriptions, and compare multiple LLM providers.

Single-model evaluation. All main experiments use GPT-4o-mini. This choice keeps the simulation computationally tractable and makes runs comparable, but it limits robustness. Future work should test stronger and more diverse LLMs, including open-weight models, to determine whether the patterns are model specific.

Memory compression. The memory system necessarily discards information. Although bounded memory is useful for repeated simulation, subtle interaction histories may be lost. Retrieval-augmented memory or structured event graphs could preserve more context while controlling prompt length.

Environment scope. The dormitory environment has four agents, 11D observations, 9 actions, and 20 episodes. It is useful for isolating public-good and externality dynamics, but it cannot capture the complexity of real communal life, friendship, negotiation, power asymmetry, or institutional support.

Absence of explicit norm interventions. The current study evaluates internalized profile differences without systematically varying external rules such as curfews, cleaning mandates, or mediation protocols. A natural extension is to cross moral-profile composition with explicit norm regimes.

Reproducibility. The paper specifies the environment, profile scores, prompt structure, memory mechanism, reward decomposition, and evaluation metrics, but full replication would benefit from released source code and prompts. Future versions should include an artifact package to support independent verification.

Future directions. Three extensions are especially important: (1) **synthetic profile distributions**, replacing national labels with sampled MFT vectors; (2) **norm × profile interaction**, testing how explicit rules interact with internalized moral priorities; and (3) **repair and mediation mechanisms**, studying whether grievance accumulation can be converted into constructive coordination.

9 CONCLUSION

This paper turns "Manners maketh man" into a computational question: how do moral-profile differences shape cooperation among artificial agents sharing a constrained environment? Using persona-prompted LLM agents, compressed profile-specific memory, and QMIX-based credit assignment, we evaluate 15 four-agent compositions in a dormitory simulation.

The results show three simulation-level patterns. First, profiles with stronger commons maintenance and behavioral restraint achieve higher Nash Social Welfare in this environment. Second, explicit fairness monitoring can reduce welfare when it becomes complaint-oriented and increases stress rather than changing contributions. Third, composition affects not only welfare but also learnability: predictable low-externality behavior produces more stable credit assignment, whereas reactive complaint cycles produce higher loss and less equitable reward distributions.

The main conceptual lesson is that prosociality is not only a matter of moral intensity but also of moral form. Similar Care scores can generate different outcomes when paired with different surrounding foundations and memory framings. At the same time, the study's claims are deliberately limited: these are artificial profiles in a simplified environment, not measurements of real cultures.

Used with this scope, MFT-grounded LLM–MARL simulation can serve as a useful sandbox for studying how moral interpretation, memory, and credit assignment interact in cooperative AI.

ACKNOWLEDGMENTS

This research was supported by the Basic Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2710004669; and by the Institute of Convergence Science (ICONS), Yonsei University, and (in part) by the Yonsei University Office of Research Affairs through the Yonsei T.R.U.S.T. (Yonsei Transdisciplinary Research Union for a Sustainable Tomorrow) program (Project Y) under Grant project no.: 2025-22-0461; and by the Yonsei Frontier Lab (YFL) Program for Distinguished Overseas Faculty of Yonsei University.

REFERENCES

[1] Mohammad Atari, Jonathan Haidt, Jesse Graham, Sena Koleva, Sean T Stevens, and Morteza Dehghani. 2023. Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology* 125, 5 (2023), 1157–1188.

[2] Robert Axelrod. 1986. An evolutionary approach to norms. *American Political Science Review* 80, 4 (1986), 1095–1111.

[3] Norbert Elias. 1994. *The Civilizing Process: Sociogenetic and Psychogenetic Investigations.* Blackwell, Oxford. Revised edition. Originally published in German, 1939.

[4] Iason Gabriel. 2020. Artificial intelligence, values, and alignment. *Minds and Machines* 30, 3 (2020), 411–437.

[5] Jesse Graham, Brian A Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H Ditto. 2011. Mapping the moral domain. *Journal of Personality and Social Psychology* 101, 2 (2011), 366–385.

[6] Jonathan Haidt. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion.* Vintage Books, New York.

[7] Jonathan Haidt and Craig Joseph. 2004. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus* 133, 4 (2004), 55–66.

[8] Edward T Hall. 1976. *Beyond Culture.* Anchor Books, Garden City, NY.

[9] Dan Hendrycks, Collin Burns, Steven Basart, Andrew Critch, Jerry Li, Dawn Song, and Jacob Steinhardt. 2021. Aligning AI with shared human values. In *Proceedings of the 9th International Conference on Learning Representations.*

[10] Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Ceyao Zhang, Jinlin Wang, Zili Wang, Steven Ka Zhong Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. 2024. MetaGPT: Meta programming for a multi-agent collaborative framework. In *Proceedings of the 12th International Conference on Learning Representations.*

[11] Raphaël Köster, Dylan Hadfield-Menell, Richard Everett, Laura Weidinger, Gillian K Hadfield, and Joel Z Leibo. 2022. Spurious normativity enhances learning of compliance and enforcement behavior in artificial agents. *Proceedings of the National Academy of Sciences* 119, 3 (2022), e2106028118.

[12] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. 2017. Multi-agent reinforcement learning in sequential social dilemmas. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems.* IFAAMAS, 464–473.

[13] Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023. CAMEL: Communicative agents for "mind" exploration of large language model society. *Advances in Neural Information Processing Systems* 36 (2023), 51991–52008.

[14] Martin A Nowak. 2006. Five rules for the evolution of cooperation. *Science* 314, 5805 (2006), 1560–1563.

[15] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology.* ACM, 1–22.

[16] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder de Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *Proceedings of the 35th International Conference on Machine Learning.* PMLR, 4295–4304.