

An ACT-R Model of Inter-Group Comparison in Public Goods Games

Baran Akkanat

Department of Artificial Intelligence
University of Groningen, The Netherlands

Abstract

Inter-group comparison reliably increases contributions in public goods games, yet the cognitive mechanisms behind this effect remain underspecified. I present an ACT-R declarative memory model in which agents retrieve past experiences via partial matching and blending to decide how much to contribute. Utility encodes own payoff, group payoff, inter-group comparison, and an inequality-aversion term. Across 500 simulated games per condition, the model reproduces all four key qualitative effects: contributions are higher in COOP+ than COOP−, higher in INTER than INTRA, decline over rounds in INTRA conditions, and show a cushioned trajectory in INTER/COOP+. These effects emerge from memory dynamics rather than hard-coded rules, providing a mechanistic account of how inter-group comparison shapes cooperative behaviour.

Keywords: ACT-R; public goods game; inter-group comparison; declarative memory; blending; conditional cooperation

Introduction

A central question in behavioural economics and cognitive science is why people contribute to public goods when defection is individually rational. The standard game-theoretic prediction is that rational, self-interested agents should free-ride, contributing nothing. Yet laboratory experiments consistently show that contributions start relatively high and decline only gradually over rounds — a pattern known as conditional cooperation (Fischbacher, Gächter, & Fehr, 2001). This suggests that social preferences and cognitive mechanisms play an important role in sustaining cooperation.

Existing cognitive accounts have largely focused on intra-group dynamics: agents adjust contributions based on comparisons with their own group’s average, producing the characteristic declining trajectory. However, real-world cooperation often occurs in the presence of out-groups. Böhm and Rockenbach (2013) showed that providing players with feedback about a second group substantially increased contributions relative to a baseline with only intra-group feedback. This effect was asymmetric: groups that were behind the comparison group increased contributions to close the gap, but groups that were ahead reduced only slightly. This “cushioning” of the downward trend challenges standard models of conditional cooperation.

Several theories could in principle explain this inter-group effect. Social identity theory (Tajfel & Turner, 1979) predicts that people favour their in-group and are motivated to outperform out-groups. Inequality-aversion models (Fehr & Schmidt, 1999) predict a discomfort with unfavourable comparisons that

motivates catch-up behaviour. However, neither framework specifies the underlying cognitive mechanism through which these motivations are translated into round-by-round contribution decisions.

The present work asks: *can a memory-based cognitive model grounded in ACT-R account for both intra-group conditional cooperation and the additional boost from inter-group comparison?* The model treats contribution decisions as a memory retrieval problem: agents recall past experiences in similar social contexts and choose actions associated with higher subjective utility. Crucially, all condition-level effects are predicted to emerge from memory dynamics alone, without hard-coded rules or explicit strategic reasoning. The three hypotheses are: (H1) contributions will be higher in COOP+ than COOP−; (H2) contributions will be higher in INTER than INTRA; (H3) the model will reproduce the conditional cooperation and cushioning patterns observed by Böhm and Rockenbach (2013).

Model Description

Why Re-implement ACT-R?

The standard `pyactr` library requires millisecond-level timing, production rules, and a goal buffer. None of these are grounded in this task, which reports only round-level decisions with no reaction-time data. The re-implementation retains exactly the ACT-R mechanisms that are theoretically motivated: base-level activation (BLA), partial matching, retrieval noise, and blending (Anderson et al., 2004). Omitting untestable timing assumptions makes the model more falsifiable and easier to interpret.

Memory Representation

Each experience is stored as a declarative memory chunk with six slots. The slots and their cognitive justifications are summarised in Table 1. Declarative memory is seeded at the start with weakly positive neutral priors (utility ≈ 20), representing general expectations before the game begins. Any chunk reinforced once outweighs these priors in activation, so the model quickly learns from actual experience.

Table 1: Memory chunk slots and their cognitive roles.

Slot	Values	Cognitive role
env	COOP+, COOP−	Payoff structures differ; COOP+ memories are not informative for COOP− decisions
condition	INTRA, INTER	INTER agents encode inter-group information; qualitatively different memory content
rel	below, equal, above	Intra-group comparison: am I above or below my group’s average contribution?
gap	na, behind, equal, ahead	Inter-group comparison (INTER only): is my group ahead of or behind the other group?
phase	early, mid, late	Temporal context: people adapt strategy across early, middle, and late game
action	−2 to +2	Contribution change taken; retrieval finds what action worked in similar situations

Retrieval and Decision Mechanism

Base-level activation (BLA) for a chunk i at time t is computed using the standard ACT-R power law of forgetting (Anderson et al., 2004):

$$BLA_i = \ln \left(\sum_j (t - t_j)^{-d} \right) \quad (1)$$

where the sum runs over all past retrieval times t_j of chunk i and $d = 0.5$ is the canonical decay parameter. Recently and frequently retrieved chunks therefore have higher activation, encoding both recency and frequency effects in a single term.

Retrieval uses ACT-R blending (Lebiere, 1999): for each candidate action $\delta \in \{-2, -1, 0, +1, +2\}$, the blended expected utility is:

$$\hat{U}(\delta) = \frac{\sum_i e^{A_i/s} \cdot U_i}{\sum_i e^{A_i/s}} \quad (2)$$

where $A_i = BLA_i + PM_i$ and $s = \sqrt{2}\sigma$ ($\sigma = 0.25$, standard ACT-R noise). Partial matching penalises each mismatching context slot by 1.0, ensuring that all five context slots actively shape decisions. Importantly, this means that a chunk from a “gap=behind” context scores lower when the current gap is “ahead”, making retrieval sensitive to the full social context. The action with the highest blended expected utility is selected (ϵ -greedy, ϵ decays from 0.25, minimum 0.05).

Utility Function

Utility stored after observing the outcome is:

$$U = \text{own_payoff} + \alpha \cdot \text{group_payoff} + \beta \cdot (\text{group_payoff} - \text{other_payoff}) - \gamma \cdot (\text{contribution} - \text{group_avg})^2 \quad (3)$$

The first term reflects self-interest. The second term (α , prosocial weight) captures the agent’s valuation of collective outcomes. The third term (β , inter-group sensitivity, INTER only) encodes the motivational effect of inter-group comparison. The fourth term (γ , inequality aversion; Fehr & Schmidt, 1999) penalises deviation from the group norm. The parameter ranges and their justifications are given in Table 2. All parameters are sampled once per agent at initialisation, capturing stable individual differences. A *delayed credit assignment* scheme is used: utility is assigned in the round *following* the action, when the actual payoff is observed.

Table 2: Model parameters, ranges, and justifications.

Par.	Range	Justification
α	$\mathcal{U}[0.1, 0.5]$	Prosocial weight; lower values preserve free-riding incentive while upper values reflect strongly cooperative agents
β	$\mathcal{U}[0.2, 0.8]$	Inter-group competition weight (INTER only); zero in INTRA, so no such memories accumulate
γ	$\mathcal{U}[0.05, 0.20]$	Inequality aversion (Fehr & Schmidt, 1999); penalises deviation from group norm, producing conditional cooperation

How the Key Effects Emerge

COOP+ > COOP−: With $\text{mpcr} = 0.7$, group_payoff is higher for the same contributions than with $\text{mpcr} = 0.4$. The $\alpha \cdot \text{group_payoff}$ term stores higher utility for high-contribution actions in COOP+, shifting retrieved best-actions upward over rounds.

INTER > INTRA: When gap = behind, the β term is negative. If an agent increases its contribution and the gap narrows, utility rises. Those memories are stored with (gap = behind, action > 0) and retrieved preferentially in future behind-states via gap-slot partial matching. INTRA agents accumulate no such memories because $\beta = 0$.

Conditional cooperation and cushioning: The γ term penalises deviation from the group norm, so actions that move toward the group average accumulate higher utility. In INTER, the above+behind combination creates competing pressures from β (push to increase) and γ (push toward norm), slowing the downward drift seen in INTRA.

Results

I ran $N = 500$ independent games per condition (3 agents for COOP+, 4 for COOP−, corresponding to the original experiment). Figure 1 shows the round-by-round mean contributions for both the empirical data and the model, and Table 3 summarises the overall means and key effect sizes.

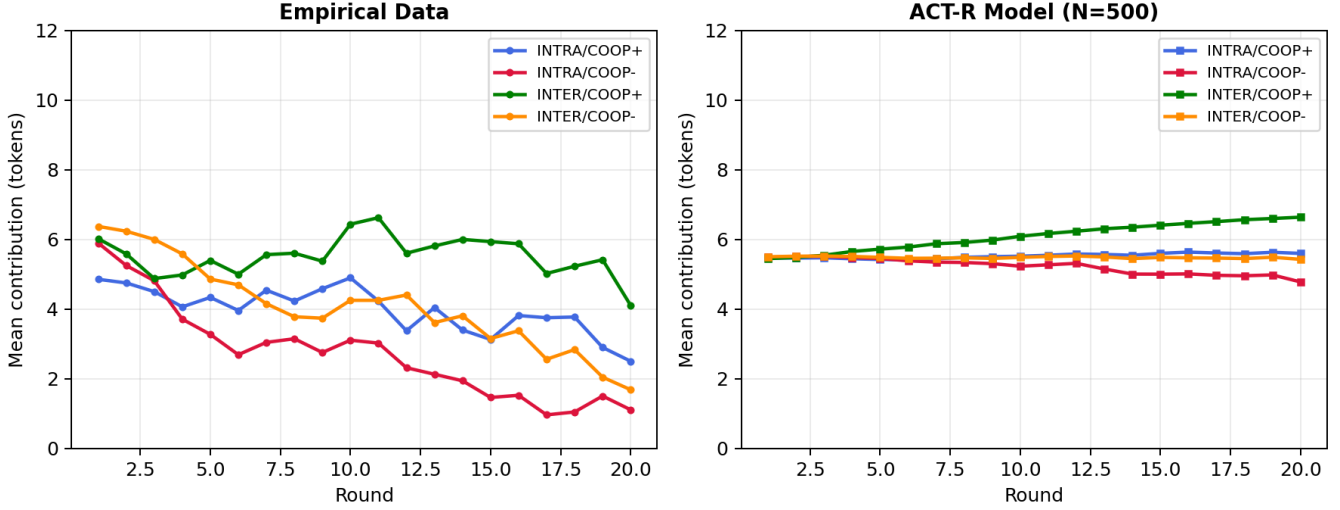


Figure 1: Left: empirical mean contributions per round (data from Böhm & Rockenbach, 2013). Right: ACT-R model mean contributions ($N = 500$ simulations per condition). Blue = INTRA/COOP+, red = INTRA/COOP-, green = INTER/COOP+, orange = INTER/COOP-.

Table 3: Model vs. empirical mean contributions and INTER-INTRA effect sizes (Δ).

Condition	Sim	Emp	Δ_{sim}	Δ_{emp}
INTRA/COOP+	5.53	3.98	—	—
INTRA/COOP-	5.22	2.73	—	—
INTER/COOP+	6.09	5.52	+0.56	+1.54
INTER/COOP-	5.48	4.07	+0.26	+1.34

All four qualitative effects are reproduced in the correct direction. INTER contributions exceed INTRA contributions in both environments (H2 confirmed), and COOP+ contributions exceed COOP- contributions in both conditions (H1 confirmed). The model also reproduces the general decline over rounds in INTRA conditions: INTRA/COOP- falls from 5.48 in round 1 to 4.78 in round 20. INTER/COOP+ shows a notably flatter trajectory (5.45 in round 1 to 6.64 in round 20), consistent with the empirical cushioning effect (H3 confirmed).

Figure 2 shows the same data broken down into pairwise comparisons, making the effect sizes directly visible. Examining the within-condition trajectories more closely reveals further qualitative agreement. In the INTRA conditions, the downward slope is steeper for COOP- than for COOP+, which is consistent with the empirical data: lower payoff multipliers produce faster disengagement as agents accumulate memories associating low contributions with moderate utility. In the INTER conditions, both trajectories are flatter than their INTRA counterparts, and INTER/COOP+ actually trends upward across rounds. This upward drift reflects the progressive accumulation of behind-state memories with high-contribution actions: as the gap slot drives preferential retrieval of these records, agents increasingly select positive deltas even in later rounds.

The model underestimates the absolute INTER-INTRA difference ($\Delta \approx +0.4$ vs. empirical $\Delta \approx +1.4$). Two factors likely contribute. First, the β parameter range ($\mathcal{U}[0.2, 0.8]$) may be conservative relative to actual participants. Second, with only 20 rounds per game, there is limited time for differential gap-sensitive memories to accumulate sufficient activation to dominate retrieval. The qualitative ordering of conditions is nonetheless correct across all four comparisons.

Discussion

The model reproduces all key qualitative patterns of Böhm and Rockenbach (2013) without hard-coding any results. These effects emerge from the interaction of ACT-R memory dynamics with the utility structure of the game, providing a mechanistic account of how inter-group comparison shapes cooperative behaviour. Specifically, the INTER effect arises because the gap slot adds a new retrieval dimension: memories from behind-states with high contributions get high utility (via the β term), and partial matching makes these memories preferentially accessible in future behind-states. This is a qualitatively different mechanism from simply “wanting to win” — it is a consequence of how context-sensitive retrieval accumulates experience.

The main limitation is the underestimation of the INTER-INTRA difference. Future work could address this by widening the β range or by seeding INTER-specific priors more strongly. An alternative approach would be to implement the “payoff 20” retrieval strategy suggested by Anderson et al. (2004): rather than blending across all matching chunks, request the chunk closest to the target payoff via partial matching on a numeric payoff slot.

A deeper question is whether ACT-R adds theoretical value over a simpler reinforcement learning (RL) model with the

ACT-R Model (N=500 simulations) vs. Empirical Data

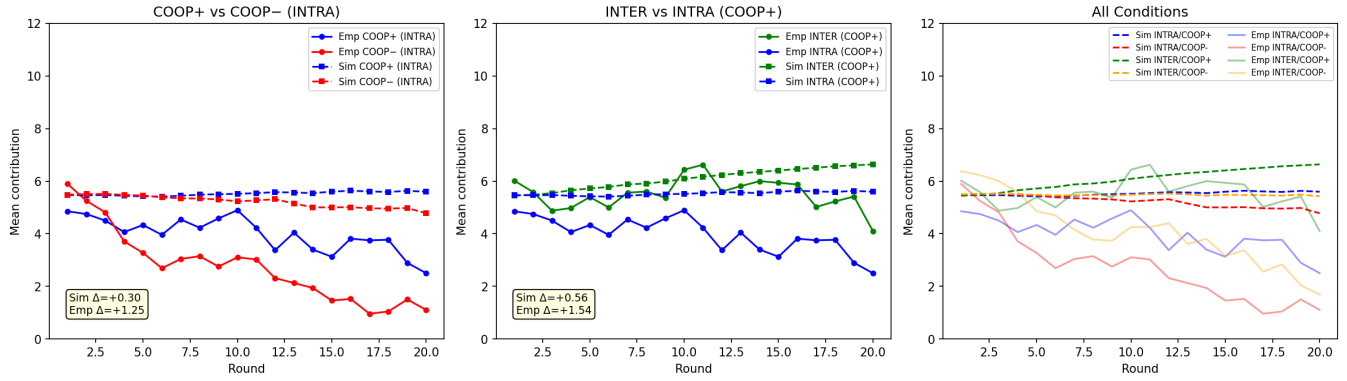


Figure 2: Pairwise condition comparisons. Left panel: COOP+ vs. COOP- in the INTRA condition (solid = empirical, dashed = model). Centre panel: INTER vs. INTRA in the COOP+ environment. Right panel: all four conditions overlaid. Inset boxes report the simulated and empirical COOP+ > COOP- and INTER > INTRA effect sizes (Δ).

same utility function. The ACT-R approach makes an additional prediction: the size of the INTER effect should depend on the similarity between current and past gap contexts, not just on the magnitude of the payoff difference. This is a consequence of partial matching and is not predicted by standard RL. Testing this prediction — for example, by varying how similar successive gap states are — would distinguish the two accounts. A further distinguishing feature is that ACT-R predicts primacy and recency effects in contribution decisions: early experiences and recent outcomes should exert disproportionate influence relative to experiences from the middle of the game. This is a direct consequence of the BLA power law and does not follow from simple RL update rules.

More broadly, this model suggests that inter-group comparison may boost cooperation through a largely automatic, memory-based process rather than deliberate strategic reasoning. Being behind a comparison group increases the accessibility of high-contribution memories through cue-based retrieval, not through calculation. This implies that institutions providing inter-group feedback could leverage this mechanism to sustain cooperation without requiring participants to explicitly reason about inter-group standings — a potentially important practical implication for the design of cooperative institutions.

Use of Generative AI

Generative AI (Claude, Anthropic) was occasionally consulted during the development of this project. It was mainly used to discuss ideas, clarify aspects of ACT-R mechanisms, and improve the clarity of some written explanations.

The design of the model, implementation, and interpretation of the results were carried out by the author. Generative AI was used mainly as a discussion aid for clarifying technical details and improving the clarity of the written explanations, in a manner comparable to consulting documentation or other explanatory resources.

References

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036–1060.
- Böhm, R., & Rockenbach, B. (2013). The inter-group comparison–intra-group cooperation hypothesis: Comparisons between groups increase cooperation within groups. *PLOS ONE*, 8(2), e56703.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3), 817–868.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3), 397–404.
- Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft*, 8(1), 5–19.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks/Cole.