PERSONALIZED BELIEF–ACTION GAME DATA AUGMENTATION: THREE-DIMENSIONAL EVALUATION AND RAGDA

Anonymous authors
Paper under double-blind review

ABSTRACT

Given a limited participant-specific history, effective belief–action augmentation should preserve history-conditioned responses and belief–action relationships beyond marginal behavior. We therefore evaluate this task along three dimensions: history–response compatibility, model-relative belief–action consistency, and synthetic substitution utility. To preserve these structures, we propose RAGDA, a retrieval-based framework that uses an LLM to author and verify retrieval programs offline and learns their combination from training data. During generation, RAGDA retrieves an admissible historical case from the same participant and reuses its paired belief–action record. Across two tasks from game-theoretic experiments, RAGDA outperforms seven generators on all history–response compatibility and model-relative belief–action consistency metrics. Under fixed-budget substitution with replacement probability, it reduces test prediction errors on first-price auction task and incurs the smallest degradation on finitely repeated prisoner's dilemma. These results demonstrate RAGDA's ability to preserve personalized behavioral structure during synthetic substitution.

1 INTRODUCTION

Human strategic-decision experiments are central to empirical game-theory research because they reveal how incentives, beliefs, and others' actions shape behavior. Joint action–belief observations show both what a participant chooses and expects, enabling analysis of history-dependent responses and action–belief relationships (Nyarko & Schotter, 2002; Kirchkamp & Reiß, 2011; Aoyagi et al., 2024). Given a limited participant-specific history, we study personalized behavioral augmentation: generating belief–action records at new legal contexts using only information available at each decision.

How should personalized belief–action data be evaluated? Existing work measures distributional or trajectory fidelity and downstream utility (Pitis et al., 2020; 2022; Lu et al., 2023; Lee et al., 2024), individual action predictability (Nay & Vorobeychik, 2016; Kolumbus & Noti, 2019; Zhu et al., 2025), or population- and game-level behavioral alignment (Aher et al., 2023; Argyle et al., 2023; Akata et al., 2025; Jia et al., 2024). These evaluations do not by themselves determine whether generated data reproduce a particular participant's history-conditioned response and preserve its belief–action pairing.

Existing generators operate at the level of transitions, trajectories, actions, or population-level responses. Counterfactual transition methods recombine state–action factors (Pitis et al., 2020; 2022); diffusion methods sample or guide replay trajectories (Lu et al., 2023; Lee et al., 2024); imitation and action-prediction methods learn policies or next-action predictors (Song et al., 2018; Nay & Vorobeychik, 2016; Kolumbus & Noti, 2019); and prompt-, persona-, or multi-agent LLM methods generate responses from profiles, exemplars, opponent models, or game histories (Aher et al., 2023; Argyle et al., 2023; Bui et al., 2025; Sreedhar & Chilton, 2025; Yu et al., 2025; Akata et al., 2025; Jia et al., 2024). These abstractions generally do not jointly model a participant's history-conditioned action and elicited belief, which are the paired objects evaluated here.

Our goal is to evaluate and generate participant-specific belief–action responses under each environment's information constraints. The evaluation challenge is to separate history–response compatibility, model-relative belief–action consistency, and synthetic substitution utility under this constraint.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The generation challenge is to retrieve a relevant same-participant precedent from the available history and handle queries with no usable precedent.

We address these challenges with a three-dimensional evaluation framework and *RAGDA* (*Retrieval-Based Agentic Framework for Game-Theoretic Data Augmentation*; Section 4). The framework measures history–response compatibility, model-relative belief–action consistency, and synthetic substitution utility using participant-normalized CMMD ratios, λ-MAE and Fisher-CI IoU, and held-out prediction after fixed-budget replacement, respectively. Controlled-mismatch experiments test the first two dimensions. RAGDA uses an offline LLM to generate and verify retrieval programs, whose combination weights are learned from training data. At generation time, it retrieves the most similar admissible historical case from the same participant and reuses its belief–action record. Figure 1 summarizes the resulting formulation: it connects observed and augmented belief–action trajectories to the three evaluation questions and contrasts direct LLM prompting with RAGDA's participant-specific historical retrieval.

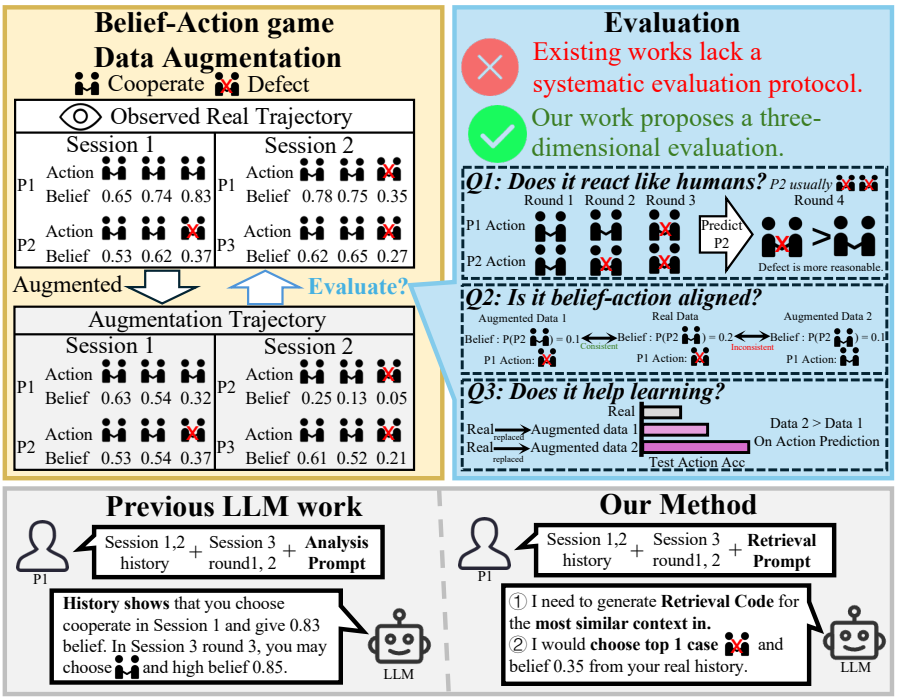


Figure 1: **Personalized belief–action augmentation: task, evaluation, and retrieval-based generation.** The top panels show observed and augmented trajectories and the three evaluation questions: history–response compatibility, model-relative belief–action consistency, and synthetic substitution utility. The bottom panels contrast direct LLM prompting with RAGDA's retrieval of a same-participant historical case.

We evaluate RAGDA against seven competing generators on two datasets from game-theoretic experiments. Under a matched LLM backbone, RAGDA outperforms all seven baselines on the history–response compatibility and model-relative belief–action consistency metrics across both datasets. With real training records replaced at probability $p = 0.5$ under a fixed budget, RAGDA reduces prediction errors on both ooExp26c tasks and incurs the smallest degradation on BiRG-finite, where no generator improves over matched real-only training.

Together, the framework and results yield three contributions:

- We introduce a three-dimensional evaluation protocol for personalized belief–action game data augmentation.
- We propose RAGDA, a retrieval-based agentic framework that uses an offline LLM to construct and verify retrieval programs for belief–action generation.
- We demonstrate strong performance across all three proposed evaluation dimensions.

2 Related Work

Strategic behavior models and generation. Strategic-behavior models describe how incentives, beliefs, and interaction histories shape choices through QRE, cognitive hierarchy, and experience-weighted attraction learning (McKelvey & Palfrey, 1995; Camerer et al., 2004; Camerer & Hua Ho, 1999). Belief-elicitation studies pair expectations with actions in learning, auctions, and repeated games (Nyarko & Schotter, 2002; Kirchkamp & Reiß, 2011; Aoyagi et al., 2024), while neural and foundation models predict individual choices (Nay & Vorobeychik, 2016; Kolumbus & Noti, 2019; Zhu et al., 2025; Binz et al., 2025). Sequential augmentation recombines dynamics or synthesizes and stitches trajectories (Pitis et al., 2020; 2022; Lu et al., 2023; Lee et al., 2024; Fu et al., 2025; Yuan et al., 2025). LLM simulation covers population and individual settings, including strategic interaction (Aher et al., 2023; Argyle et al., 2023; Bui et al., 2025; Park et al., 2024; Lei et al., 2026; Akata et al., 2025; Jia et al., 2024; Yu et al., 2025). These approaches primarily model observed choices, policy/replay trajectories, or population responses, rather than generate a known participant's belief–action record at a legal context given a limited participant-specific history. RAGDA uses an admissible same-participant record as the generation unit and preserves its paired outputs.

Evaluation of generated behavior and history–response compatibility. Evaluation work measures distributional or trajectory fidelity, downstream utility, and agreement with human responses (Pitis et al., 2020; 2022; Lu et al., 2023; Lee et al., 2024). Conditional mean-embedding methods support CMMD for history–response distributions (Ren et al., 2016). SimBench evaluates group-level, single-turn LLM responses across 20 datasets (Hu et al., 2026). QRE-based evaluation compares participant-level parameters (McKelvey & Palfrey, 1995). Existing evaluations address these properties separately; we evaluate all three jointly and calibrate the first two under controlled mismatch.

Personalized retrieval and program discovery. Personalized retrieval conditions language-model outputs on user history, as in LaMP (Salemi et al., 2024). Program-search methods use LLM proposals and external evaluation to discover programs or cognitive models, as in FunSearch and CogFunSearch (Romera-Paredes et al., 2024; Castro et al., 2025). RAGDA differs by using LLM-proposed, externally verified comparison programs to learn a participant-specific retrieval distance, with online LLM generation only at cold start.

3 Task Setting and Three-Dimensional Evaluation

We evaluate augmented data along three dimensions: history–response compatibility, model-relative belief–action consistency, and synthetic substitution utility. The first measures agreement with human history–response relationships, the second compares belief-conditioned choice fits under a specified model, and the third measures the effect of replacing real training records with synthetic ones on held-out prediction.

3.1 Task Setting

Given a limited participant-specific history, we generate a belief–action sequence for that participant using only information available at each decision point.

For participant i, let $\tau_i = \{(x_{it}, b_{it}, a_{it})\}_{t=1}^{T_i}$ denote the observed sequence, where x_{it}, b_{it}, and a_{it} denote the available information, elicited belief, and observed action. For target participant i, let H_i denote the observed personal history over indices $\mathcal{O}_i$. Given a legal query q_{it}, the generator produces an augmented belief–action record

$$\tilde{y}_{it} = (\tilde{b}_{it}, \tilde{a}_{it}) \sim G(\cdot \mid H_i, q_{it}).$$

Each generated record is incorporated into the augmented data according to the environment's temporal structure using only information available at the corresponding query; construction details are given in Appendix A.1.

3.2 History–Response Compatibility

CMMD (Ren et al., 2016) assesses history–response compatibility by comparing the human and augmented conditional response distributions, denoted by $P_i^{\mathrm{H}}(Y^{c,h} \mid X)$ and $P_i^{\mathrm{aug}}(Y^{c,h} \mid X)$. Given

history length L and prediction horizon h, we construct X_{it} from the pre-query history window and game context, and define action and belief responses as in equation 1.

$$\begin{aligned} Y_{it}^{\text{action},h} &= (A_{i,t}, \ldots, A_{i,t+h-1}), \\ Y_{it}^{\text{belief},h} &= (b_{i,t}, \ldots, b_{i,t+h-1}). \end{aligned} \tag{1}$$

$$R_{i,c,h} = \frac{\text{CMMD}^2\left(P_i^{\text{H}}(Y^{c,h} \mid X), P_i^{\text{aug}}(Y^{c,h} \mid X)\right)}{B_{i,c,h}^{\text{RR}}}, \tag{2}$$

Here $B_{i,c,h}^{\text{RR}}$ is the participant-level real–real split-half reference. We report macro-averaged action and belief ratios over eligible participants; lower values indicate closer distributional history–response compatibility under this metric.

3.3 Model-Relative Belief–Action Consistency

We assess model-relative belief–action consistency using the same belief-conditioned choice model on human and generated records, comparing the resulting payoff-sensitivity parameters. The model uses the environment's action and payoff spaces; human and generated fits share this specification within each environment. For participant i at decision t, we use a logit quantal-response model:

$$p_{\lambda_i}(a \mid b_{it}, x_{it}) = \frac{\exp\{\lambda_i U(a \mid b_{it}, x_{it})\}}{\sum_{a' \in \mathcal{A}} \exp\{\lambda_i U(a' \mid b_{it}, x_{it})\}}, \tag{3}$$

where λ_i captures sensitivity to belief-implied expected payoffs. We estimate $\hat{\lambda}_i$ from human trajectories and $\hat{\tilde{\lambda}}_i$ from generated trajectories using the same model specification.

We measure parameter agreement using the participant-level MAE:

$$\text{MAE}_\lambda = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} |\hat{\tilde{\lambda}}_i - \hat{\lambda}_i|, \tag{4}$$

and confidence-interval agreement using Fisher-CI IoU:

$$\text{IoU}_{\text{Fisher}} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \frac{\text{len}(C_i \cap \tilde{C}_i)}{\text{len}(C_i \cup \tilde{C}_i)}. \tag{5}$$

Lower MAE_λ and higher $\text{IoU}_{\text{Fisher}}$ indicate stronger model-relative belief–action consistency.

3.4 Synthetic Substitution Utility

Synthetic substitution utility measures how prediction changes when generated records replace a fraction of real training records, with the learner, training-set size, and training budget held fixed. We compare the resulting predictors against matched real-only predictors on human test sets, separately for action and belief prediction.

Let $f_{c,p}^G$ denote the predictor trained by replacing each eligible real training unit with its matched synthetic record with probability p at each training epoch, using fresh replacement draws each epoch, and let $f_{c,0}$ denote the corresponding real-only predictor; in the main experiments, $p = 0.5$. We report the signed difference in the task-specific test metric:

$$\Delta_{p,c}(G) = \mathcal{M}_c(f_{c,p}^G) - \mathcal{M}_c(f_{c,0}), \tag{6}$$

where $\mathcal{M}_c$ is action accuracy for BiRG-finite's action task and MAE for the other tasks. Thus a positive Δ is favorable for BiRG-finite action accuracy, whereas a negative Δ is favorable for MAE.

4 RAGDA: Personalized Belief–Action Data Augmentation

4.1 Overview

RAGDA is a retrieval-based agentic framework for personalized belief–action augmentation (Figure 2). It has three stages: *Search Space* defines query and candidate fields, admissible histories, and retrieval families; *Program Discovery* uses an LLM to generate and verify retrieval programs, then learns their combination by minimizing training retrieval loss; and *Select and Augment* selects and freezes a retriever on validation data, then reuses retrieved cases to generate participant-specific sequences under the target environment's transition rules. Generation uses an LLM only for cold-start queries with no usable historical case.

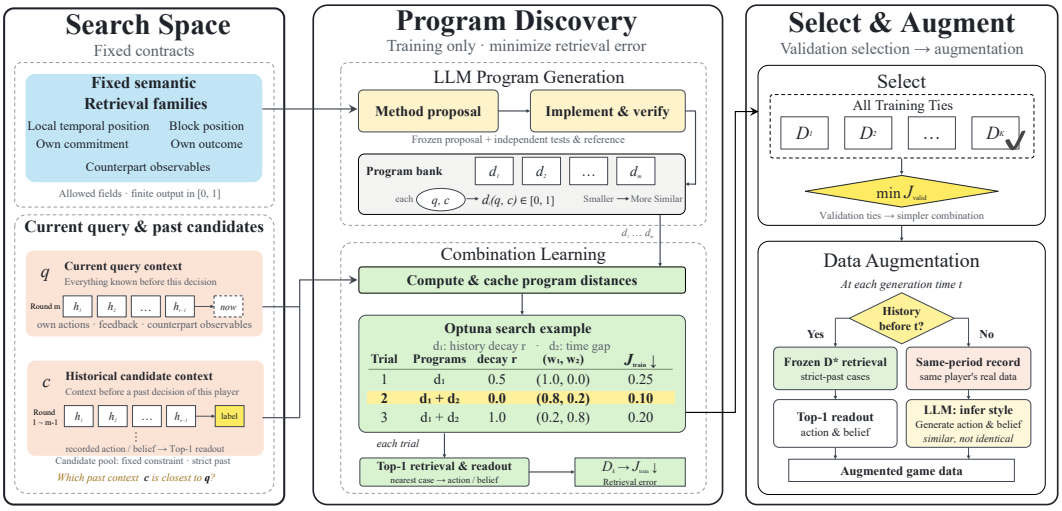


Figure 2: **Three-stage RAGDA pipeline. Search Space** defines the observable comparison contracts and the admissible historical pool. **Program Discovery** authors, verifies, and combines comparison programs by minimizing the training retrieval loss. **Frozen Evaluation** selects a retriever from the complete training tie pool using validation and then reads the output of one Top-1 case during generation. The cold-start LLM is used only when no admissible historical case exists.

4.2 Search Space: Fixed Semantic Contracts

For query $q = \tilde{x}_{it}$ and earlier case $c = x_j$, the candidate pool $\mathcal{C}(q)$ contains same-participant cases admissible under the shared generation setting and providing the outputs required for replay. Admissibility is environment-defined; retrieval programs score similarity only among these cases.

We consider five semantic retrieval families: local temporal position, block position, own commitment, own outcome, and counterpart observables. Each family specifies the permitted query and candidate fields, while a retrieval program defines the concrete comparison rule:

$$d_k(q, c) \in [0, 1], \tag{7}$$

where smaller values indicate greater similarity.

4.3 Program Discovery on Training Data

Once the program bank $\mathcal{B} = d_1, \ldots, d_m$ is constructed, RAGDA evaluates each program on all training query–candidate pairs and caches the resulting distances. It then searches for a subset of programs $S \subseteq \mathcal{B}$ and nonnegative weights $\alpha_{k\,k\in S}$ that minimize the training retrieval loss. The combined distance is

$$D(q, c) = \frac{\sum_{k\in S} \alpha_k d_k(q, c)}{\sum_{k\in S} \alpha_k}, \qquad \alpha_k \geq 0. \tag{8}$$

For each training query q, the retrieved case is

$$j^*(q; D) = \arg\min_{c\in\mathcal{C}(q)} \big(D(q, c), \mathrm{id}(c)\big), \tag{9}$$

and the combination is optimized by

$$J_{\mathrm{train}}(D) = \sum_{q\in\mathcal{Q}_{\mathrm{tr}}} \omega_q, e(y_{j^*}, y_q). \tag{10}$$

Here, y_q is the target output of query q, y_{j^*} is the recorded output of its retrieved case, and e is the dataset-specific retrieval error. The subset and weights with the lowest training loss define the learned retrieval distance.

4.4 Retriever Selection and Generation

After program discovery, RAGDA selects the final retriever on the validation set while keeping the program bank and candidate configurations fixed. Let $\widehat{J}_{\text{train}}$ denote the minimum training retrieval loss. We retain the near-optimal training configurations

$$\mathcal{T}_{\text{train}} = \left\{ D : J_{\text{train}}(D) \leq \widehat{J}_{\text{train}} + \tau_{\text{tie}} \right\}, \tag{11}$$

and select

$$D^* = \arg \min_{D \in \mathcal{T}_{\text{train}}} J_{\text{valid}}(D). \tag{12}$$

No programs or weights are modified during validation. Appendix A.2 reports the search settings, program-bank sizes, verification outcomes, Optuna trial counts, and tie tolerance.

The selected retriever D^* is then fixed for behavioral generation. For each query $\tilde{x}_{it}$,

$$\tilde{y}_{it} = \begin{cases} y_{j^*}, & j^* = \arg \min_{c \in \mathcal{C}(\tilde{x}_{it})} D^*(\tilde{x}_{it}, c), \\ G_0(\cdot \mid \tilde{x}_{it}), & \mathcal{C}(\tilde{x}_{it}) = \emptyset. \end{cases} \tag{13}$$

All output components are taken from the same retrieved case to preserve their joint behavioral structure. The LLM-based generator G_0 is used only for cold-start queries with no admissible historical case.

5 Experiments

5.1 Experimental Setup

Datasets

Given a limited participant-specific history, we evaluate RAGDA on two processed belief–action datasets with different game structures: ***BiRG-finite***, a finite repeated Prisoner's Dilemma, and *ooExp26c*, a sequence of one-shot auctions.

BiRG-finite. Using public data from Aoyagi et al. (2024), we retain 30 participants in three participant-level folds, each with multiple eight-round supergames and complete participant trajectories and opponent feedback. Within each fold, supergames 1–7 are used for training, 8 for validation, and later supergames for testing.

ooExp26c. We use the *expectations* treatment of Kirchkamp & Reiß (2011), which contains 74 participants over twelve ordered auction periods. Each period is a separate one-shot game, so earlier periods provide personal history and partially observed opponent feedback. Expanding each period to six private-value nodes, $\{0, 10, 20, 30, 40, 50\}$, with a bid and elicited belief at each node yields 888 subject–period episodes and 5,328 decision points. For CMMD, each generated period replaces the preceding history while the subsequently observed human response remains fixed. This history–response compatibility test measures whether the resulting pairs remain close to empirical human transition patterns; it does not estimate how the participant would causally respond to the generated history.

All methods, including the LLM baselines, use the same strict-past query-time observability, generation, and metric protocols described in Appendices A.1 and A.3; all main-text results use these fixed settings.

5.2 Metric Validation under Controlled Mismatch

Figure 3 evaluates whether the metrics respond to controlled history–response mismatch. For BiRG-finite, we keep each history fixed and reassign a fraction of complete future responses to same-participant windows, using 20 paired seeds for every (L, h) setting. Each setting uses its own human-fitted kernels and a size-matched real–real denominator. The action and belief CMMD ratios are non-decreasing with mismatch: at $(L, h) = (1, 1)$, the BiRG ratios rise from 1.033 to 1.386 and from 1.027 to 1.603, respectively, while both ooExp26c ratios rise from roughly 1.002 to 1.012. Because lower CMMD ratios indicate closer history–response compatibility, these curves validate the metric's response to history–response mismatch.

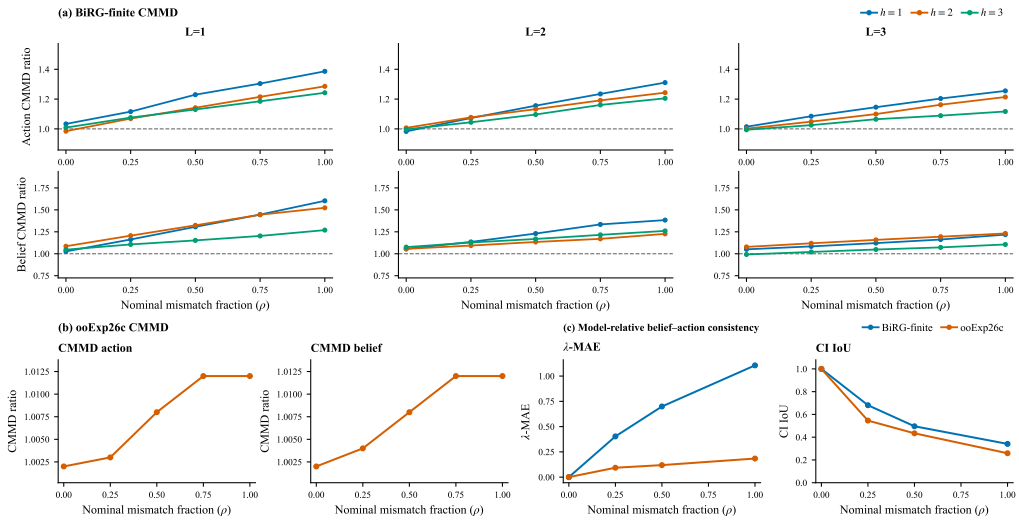


Figure 3: **Controlled-mismatch calibration.** Panel (a) is a six-panel BiRG-finite 2×3 matrix: rows are action and belief CMMD, columns are $L \in \{1, 2, 3\}$, and each panel contains the three $h \in \{1, 2, 3\}$ curves shown by the shared color key. Panel (b) reports the ooExp26c CMMD curves; panel (c) reports λ-MAE and Fisher-CI IoU for both datasets.

For model-relative belief–action consistency, we exchange BiRG cooperation decisions on belief-observed rows and ooExp26c six-value bid curves across participants while holding beliefs fixed. Over 20 nested seeds, λ-MAE increases and Fisher-CI IoU decreases with mismatch. At full mismatch, BiRG-finite λ-MAE rises from 0 to 1.106 and IoU falls from 1 to 0.341; on ooExp26c, they reach 0.184 and 0.259, respectively. The metrics thus detect participant-level belief–action mismatch under the specified model.

5.3 Main Comparison across Three Evaluation Dimensions

We compared RAGDA with seven baselines across history–response compatibility, model-relative belief–action consistency, and synthetic substitution utility on BiRG-finite and ooExp26c. The baselines included neural generators (NAY-MLP (Nay & Vorobeychik, 2016), CNN (Kolumbus & Noti, 2019), and MAGAIL-Song (Song et al., 2018)) and LLM-based generators (LLM-MOP (Bui et al., 2025), Strategic-MAS (Sreedhar & Chilton, 2025), EMO (Yu et al., 2025), and Repeated Games LLM (Akata et al., 2025)). To isolate generator design from language-model effects, all LLM baselines and the DeepSeek RAGDA configuration used for the matched comparison shared the DeepSeek-V4-Flash backbone (DeepSeek, 2026). All methods followed the dataset-specific generation and evaluation protocols in Appendix A.1 and Appendix A.3. We additionally reported RAGDA with GPT-5.6 Luna (OpenAI, 2026) and Claude Haiku 4.5 (Anthropic, 2025) to assess the robustness of the pipeline across authoring backbones. Table 1 summarizes the resulting comparisons.

Overall comparison. Under the matched DeepSeek configuration, RAGDA achieved lower λ-MAE and CMMD ratios and higher Fisher-CI IoU than all baselines on both datasets. This consistent pattern in the repeated-game and one-shot settings supported the evaluation objective: the RAGDA pipeline achieved closer history–response compatibility and model-relative belief–action consistency under both temporal structures. The substitution results were dataset-dependent: RAGDA achieved the most favorable metric differences in ooExp26c, whereas no method improved either target over matched real-only training in BiRG-finite under the $p = 0.5$ replacement setting; the best RAGDA configuration incurred the smallest degradation.

Backbone robustness. Across the three evaluated authoring backbones, RAGDA attains closely aligned scores on both datasets, showing that its measured performance is robust to the choice of authoring LLM.

Table 1: Main comparison on BiRG-finite (three folds; locked strict-bilateral protocol) and ooExp26c (locked bilateral history-state protocol). CMMD uses $(L, h) = (1, 1)$. Substitution columns report differences from matched real-only training at $p = 0.5$ under fixed learner and training budgets: BiRG-finite $\Delta\mathrm{Act}$ is an accuracy difference (higher is better), while its $\Delta\mathrm{Bel}$ and both ooExp26c differences are MAE differences (lower is better). Other metric directions appear in the headers. Bold marks the best value within each dataset; ties at the reported precision are jointly bold.

BiRG-finite

Method	λ-MAE↓	Fisher-CI IoU↑	CMMD-A↓	CMMD-B↓	$\Delta\mathrm{Act}$ ↑	$\Delta\mathrm{Bel}$ ↓
Machine-learning generators						
CNN	0.9494	0.4045	1.9059	2.8478	-0.0503	+0.0971
MAGAIL-Song	1.0854	0.4026	1.9459	2.7535	-0.0528	+0.0925
NAY-MLP	1.1753	0.3799	1.7635	2.5037	-0.0431	+0.0879
LLM generators (DeepSeek-V4-Flash backbone)						
LLM-MOP	1.0216	0.4189	2.8740	3.1055	-0.0507	+0.0760
LLM-Strategic-MAS	1.0706	0.3781	2.8979	3.1364	-0.0493	+0.0774
LLM-EMO	0.9819	0.4433	2.9251	3.1673	-0.0469	+0.0780
LLM-Repeated Games	1.0208	0.4223	2.8708	3.1493	-0.0521	+0.0762
RAGDA (DeepSeek-V4-Flash)	0.6146	0.5364	**1.5097**	**1.9992**	-0.0451	+0.0556
RAGDA (GPT-5.6 Luna)	0.6241	0.5073	1.6927	2.1312	-0.0483	+0.0661
RAGDA (Claude Haiku 4.5)	**0.5938**	**0.5523**	1.5641	2.1334	**−0.0424**	**+0.0485**

ooExp26c

Method	λ-MAE↓	Fisher-CI IoU↑	CMMD-A↓	CMMD-B↓	$\Delta\mathrm{Act}$ ↓	$\Delta\mathrm{Bel}$ ↓
Machine-learning generators						
CNN	0.1159	0.4373	1.5819	1.5795	−0.4170	−0.4390
MAGAIL-Song	0.1251	0.4692	1.6250	1.6222	−0.3580	−0.3880
NAY-MLP	0.1271	0.3575	1.6952	1.6945	−0.4190	−0.4810
LLM generators (DeepSeek-V4-Flash backbone)						
LLM-MOP	0.1910	0.4360	2.2250	2.2246	−0.3620	−0.0710
LLM-Strategic-MAS	0.2570	0.3690	2.3295	2.3296	−0.3390	+0.0740
LLM-EMO	0.1160	0.5470	2.2598	2.2600	−0.4210	−0.2670
LLM-Repeated Games	0.0950	0.5840	2.2877	2.2872	−0.3520	−0.2360
RAGDA (DeepSeek-V4-Flash)	0.0264	0.8153	1.1153	1.1082	−0.4790	−0.5820
RAGDA (GPT-5.6 Luna)	0.0261	0.8212	**1.1032**	**1.0979**	**−0.5000**	**−0.6020**
RAGDA (Claude Haiku 4.5)	**0.0259**	**0.8229**	1.1158	1.1101	**−0.5000**	−0.5830

5.4 RAGDA-Generated Retrieval Programs

To examine whether RAGDA uses retrieval cues that reflect the behavioral regularities of individual participants in each game, we inspect the selected retrieval programs and learned weights on both datasets. The selected retrieval items, their comparison logic, and final coefficients are reported in Appendix A.6.

On BiRG-finite, the selected retrieval logic emphasizes counterpart-action history and within-supergame position. This combination is consistent with experimental evidence from finite repeated prisoner's-dilemma games: participants use conditionally responsive strategies such as tit-for-tat and Grim trigger, while threshold strategies change behavior at a round-dependent point (Aoyagi et al., 2024). Thus, the retrieved cues capture both responses to the counterpart and end-of-supergame behavioral changes.

Across all three authoring backbones, the selected retrieval logic emphasizes period position. This emphasis is consistent with the temporal persistence observed in ooExp26c: participants' bid and belief functions are closely related across adjacent rounds and evolve through gradual adjustment, as also documented in prior auction evidence (Kirchkamp & Reiß, 2011; Neugebauer & Selten, 2006).

5.5 Retrieval Structure Analysis

Participant-specific retrieval. To examine the effect of participant-specific retrieval, we replace each target participant's admissible pool with an all-player pool while keeping the retrieval logic, observability, and rollout. We refit the retrieval weights and regenerate data. The all-player pool degrades λ-MAE, Fisher-CI IoU, and both CMMD metrics on both datasets (Table 2). The downstream gap is small on BiRG-finite but larger on ooExp26c, supporting participant-specific retrieval for preserving personalized behavioral structure (see Appendices A.7–A.8).

Table 2: Retrieval structure analysis on BiRG-finite (three-fold average) and ooExp26c. Rows specify pool/readout: same-participant or all-player, with shared or separate retrieval. CMMD uses $(L, h) = (1, 1)$; ooExp26c fixes real-first-period cold start. For the pool comparison, retrieval weights are refit with observability, output availability, rollout, replacement, and learner budget fixed. Lower λ-MAE and CMMD and higher CI IoU are better; BiRG ΔAct favors higher values, other Δ columns lower. Bold marks the best reported values per dataset.

Pool/readout	λ-MAE	CI IoU	CMMD-A	CMMD-B	ΔAct	ΔBel
BiRG-finite						
Same/shared	0.6241	0.5073	**1.6927**	**2.1312**	**-0.0483**	+0.0661
All/shared	1.1480	0.3624	1.8175	2.6529	-0.0486	**+0.0626**
Same/separate	**0.5895**	**0.5369**	1.6955	2.1519	-0.0503	+0.0602
ooExp26c						
Same/shared	**0.0260**	**0.8228**	**1.1032**	**1.0979**	**-0.5004**	**-0.5840**
All/shared	0.0684	0.7010	1.2057	1.1996	-0.4464	-0.4010
Same/separate	**0.0260**	**0.8228**	**1.1032**	**1.0979**	**-0.5004**	**-0.5840**

Paired-replay diagnostic. To test whether favorable λ-MAE is mechanically induced by replaying an intact human pair, we fit separate action and belief retrievers over the same admissible pool. On BiRG-finite, they choose different donors for 54/480 belief-observed decisions, yet λ-MAE improves from 0.6241 to 0.5895 and Fisher-CI IoU from 0.5073 to 0.5369, while CMMD changes only slightly (Table 2). Thus, strong model-relative consistency is not mechanically guaranteed by intact-pair replay. On ooExp26c, both retrievers select the same donor for all 370 retrieved actor–periods, leaving all metrics unchanged.

6 Conclusion

This work establishes a three-dimensional evaluation of personalized belief–action augmentation, covering history–response compatibility, model-relative belief–action consistency, and synthetic substitution utility. RAGDA operationalizes this evaluation through participant-specific historical retrieval: an offline LLM authors and verifies retrieval programs, and generation reuses the retrieved belief–action record. Across BiRG-finite and ooExp26c, RAGDA outperforms seven generators on history–response compatibility and model-relative belief–action consistency under a matched backbone. Under fixed-budget substitution, it reduces prediction errors on both ooExp26c tasks and incurs the smallest degradation on BiRG-finite, where synthetic replacement does not improve over real-only training. These results support historical retrieval for within-game and cross-period augmentation given a limited participant-specific history.

Future work will extend the evaluation to a broader range of game-theoretic environments with different interaction horizons, information structures, and payoff incentives. This will test whether the three-dimensional protocol and participant-specific retrieval transfer beyond the repeated-dilemma and one-shot-auction settings studied here.

AI use statement

We used AI assistants for literature retrieval and discovery, research ideation and experimental planning, methodological and experimental discussion, and to draft, edit, and polish parts of the manuscript. The authors independently verified the relevant sources, code, figures, analyses, and reported results, and take full responsibility for the final content and claims of this paper.

Ethics statement

This work analyzes previously collected experimental data from the finite repeated Prisoner's Dilemma and one-shot auction studies cited in the paper; it does not recruit participants or collect new human-subject data. We use processed behavioral records for research analysis only, do not attempt to identify or contact participants, and follow the reuse conditions of the source datasets and studies.

Reproducibility statement

The task definition and evaluation metrics are specified in Section 3, the RAGDA pipeline in Section 4, and the experimental protocols in Section 5 and Appendices A.1 and A.3. Appendix A.6 lists the selected retrieval items and weights, and Appendices A.7 and A.8 provide the qualitative cases. An anonymous supplementary code package contains the retrieval-program search, rollout and evaluation entry points, configuration files, processed inputs, dependency specification, and execution instructions.

References

Gati V Aher, Rosa I Arriaga, and Adam Tauman Kalai. Using large language models to simulate multiple humans and replicate human subject studies. In *International conference on machine learning*, pp. 337–371. PMLR, 2023.

Elif Akata, Lion Schulz, Julian Coda-Forno, Seong Joon Oh, Matthias Bethge, and Eric Schulz. Playing repeated games with large language models. *Nature Human Behaviour*, 9(7):1380–1390, 2025.

Anthropic. Introducing claude haiku 4.5, 2025.

Masaki Aoyagi, Guillaume R Fréchette, and Sevgi Yuksel. Beliefs in repeated games: An experiment. *American Economic Review*, 114(12):3944–3975, 2024.

Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351, 2023.

Marcel Binz, Elif Akata, Matthias Bethge, Franziska Br"andle, Fred Callaway, Julian Coda-Forno, Peter Dayan, Can Demircan, Maria K Eckstein, Noémi Éltető, et al. A foundation model to predict and capture human cognition. *Nature*, 644(8078):1002–1009, 2025.

Ngoc Bui, Hieu Trung Nguyen, Shantanu Kumar, Julian Theodore, Weikang Qiu, Viet Anh Nguyen, and Rex Ying. Mixture-of-personas language models for population simulation. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 24761–24778, 2025.

Colin Camerer and Teck Hua Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874, 1999.

Colin F Camerer, Teck-Hua Ho, and Juin-Kuan Chong. A cognitive hierarchy model of games. *The quarterly journal of economics*, 119(3):861–898, 2004.

Pablo Samuel Castro, Nenad Tomasev, Ankit Anand, Navodita Sharma, Rishika Mohanta, Aparna Dev, Kuba Perlin, Siddhant Jain, Kyle Levin, Noémi Éltető, et al. Discovering symbolic cognitive models from human and animal behavior. *bioRxiv*, pp. 2025–02, 2025.

DeepSeek. Deepseek-v4 preview: Entering the era of affordable million-token context. DeepSeek, 2026. URL `https://deepseek.com/en/news/v4-preview/`. Accessed: September 20, 2026.

Yuqian Fu, Yuanheng Zhu, Jian Zhao, Jiajun Chai, and Dongbin Zhao. Ins: Interaction-aware synthesis to enhance offline multi-agent reinforcement learning. In *International Conference on Learning Representations*, volume 2025, pp. 10703–10727, 2025.

Tiancheng Hu, Joachim Baumann, Lorenzo Lupo, Nigel Collier, Dirk Hovy, and Paul R"ottger. Simbench: Benchmarking the ability of large language models to simulate human behaviors. In *International Conference on Learning Representations*, volume 2026, pp. 28290–28341, 2026.

Feiran Jia, Ziyu Ye, Shiyang Lai, Kai Shu, Jindong Gu, Adel Bibi, Ziniu Hu, David Jurgens, James Evans, Philip H Torr, et al. Can large language model agents simulate human trust behavior? *Advances in neural information processing systems*, 37:15674–15729, 2024.

Oliver Kirchkamp and J Philipp Reiß. Out-of-equilibrium bids in first-price auctions: Wrong expectations or wrong bids. *The Economic Journal*, 121(557):1361–1397, 2011.

Yoav Kolumbus and Gali Noti. Neural networks for predicting human interactions in repeated games. *arXiv preprint arXiv:1911.03233*, 2019.

Jaewoo Lee, Sujin Yun, Taeyoung Yun, and Jinkyoo Park. Gta: Generative trajectory augmentation with guidance for offline reinforcement learning. *Advances in neural information processing systems*, 37:56766–56801, 2024.

Yuxuan Lei, Tianfu Wang, Jianxun Lian, Zhengyu Hu, Defu Lian, and Xing Xie. Humanllm: Towards personalized understanding and simulation of human nature. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 621–632, 2026.

Cong Lu, Philip Ball, Yee Whye Teh, and Jack Parker-Holder. Synthetic experience replay. *Advances in Neural Information Processing Systems*, 36:46323–46344, 2023.

Richard D McKelvey and Thomas R Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

John J Nay and Yevgeniy Vorobeychik. Predicting human cooperation. *PloS one*, 11(5):e0155656, 2016.

Tibor Neugebauer and Reinhard Selten. Individual behavior of first-price auctions: The importance of information feedback in computerized experimental markets. *Games and Economic Behavior*, 54(1):183–204, 2006.

Yaw Nyarko and Andrew Schotter. An experimental study of belief learning using elicited beliefs. *Econometrica*, 70(3):971–1005, 2002.

OpenAI. Gpt-5.6 luna model. OpenAI API Documentation, 2026. URL `https://developers.openai.com/api/docs/models/gpt-5.6-luna`. Accessed: September 20, 2026.

Joon Sung Park, Carolyn Q Zou, Jonne Kamphorst, Niles Egan, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Percy Liang, Robb Willer, et al. Llm agents grounded in self-reports enable general-purpose simulation of individuals. *arXiv preprint arXiv:2411.10109*, 2024.

Silviu Pitis, Elliot Creager, and Animesh Garg. Counterfactual data augmentation using locally factored dynamics. *Advances in Neural Information Processing Systems*, 33:3976–3990, 2020.

Silviu Pitis, Elliot Creager, Ajay Mandlekar, and Animesh Garg. Mocoda: Model-based counterfactual data augmentation. *Advances in Neural Information Processing Systems*, 35:18143–18156, 2022.

Yong Ren, Jun Zhu, Jialian Li, and Yucen Luo. Conditional generative moment-matching networks. *Advances in Neural Information Processing Systems*, 29, 2016.

Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Matej Balog, M Pawan Kumar, Emilien Dupont, Francisco JR Ruiz, Jordan S Ellenberg, Pengming Wang, Omar Fawzi, et al. Mathematical discoveries from program search with large language models. *Nature*, 625(7995):468–475, 2024.

Alireza Salemi, Sheshera Mysore, Michael Bendersky, and Hamed Zamani. Lamp: When large language models meet personalization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 7370–7392, 2024.

Jiaming Song, Hongyu Ren, Dorsa Sadigh, and Stefano Ermon. Multi-agent generative adversarial imitation learning. *Advances in neural information processing systems*, 31, 2018.

Karthik Sreedhar and Lydia B Chilton. Simulating strategic reasoning: Comparing the ability of single llms and multi-agent systems to replicate human behavior. In *HICSS*, pp. 1–10, 2025.

XiaoPeng Yu, Wanpeng Zhang, and Zongqing Lu. Llm-based explicit models of opponents for multi-agent games. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 892–911, 2025.

Lei Yuan, Yuqi Bian, Lihe Li, Ziqian Zhang, Cong Guan, and Yang Yu. Efficient multi-agent offline coordination via diffusion-based trajectory stitching. In *International Conference on Learning Representations*, volume 2025, pp. 25018–25045, 2025.

Jian-Qiao Zhu, Joshua C Peterson, Benjamin Enke, and Thomas L Griffiths. Capturing the complexity of human strategic decision-making with machine learning. *Nature Human Behaviour*, 9 (10):2114–2120, 2025.

Table 3: Final program-bank sizes and realized proposal and verification statistics for each dataset and authoring LLM. Verification rejection rates use only programs that reached reference or generated-suite checks.

Dataset	Authoring LLM	$\lvert\mathcal{B}\rvert$	Attempts	Rejected / checked
BiRG-finite	DeepSeek-V4-Flash	29	38	$23/36$ (63.9%)
	GPT-5.6 Luna	39	44	$6/38$ (15.8%)
	Claude Haiku 4.5	21	25	$8/23$ (34.8%)
ooExp26c	DeepSeek-V4-Flash	46	48	$31/48$ (64.6%)
	GPT-5.6 Luna	22	38	$8/28$ (28.6%)
	Claude Haiku 4.5	5	21	$13/16$ (81.3%)

A Appendix

A.1 Generation Settings

We impose a shared strict-past query-time setting for all generators, including the direct LLM baselines: each query exposes only the target participant's observed history and environment information available before the queried decision or period; the queried response and all later records are withheld. For RAGDA, this setting determines the admissible historical pool, within which participant-specific retrieval programs rank candidate cases.

All methods use the same participant matching and rollout protocol. We retain opponent pairings observed in the real data whenever they are available, and rematch only opponent relations for which no corresponding real-data match exists, drawing replacements from the eligible opponent pool. In BiRG-finite, participants are then rolled out against the retained or newly rematched opponent; the augmented data combine generated histories with generated continuation outcomes, with belief–action pairs included only in supergames 5–7, where beliefs were elicited. In ooExp26c, a generated period record is inserted into the participant's history for the subsequent period, whose response remains the real data's label; each generated period contains bids and beliefs at all six private-value nodes. Thus, the augmented CMMD windows pair a substituted history $\tilde{H}_t$ with the subsequently observed response Y_{t+1}^{real}. These windows are compared with real windows for distributional history–response compatibility; Y_{t+1}^{real} is not treated as a causal response to $\tilde{H}_t$.

A.2 Retrieval Program Search and Selection

For each dataset and authoring LLM, program discovery was scheduled for six rounds with two proposal attempts per semantic family per round and a cap of 48 attempts per run. In each completed search round, the candidate program combination and, when available, the incumbent were each refit with 20 Optuna trials. Once the program bank was fixed, the final refit used 300 Optuna trials per BiRG-finite fold or ooExp26c run. The training tie pool used the absolute retrieval-loss tolerance $\tau_{\text{tie}} = 10^{-12}$ in Equation 12; validation selected among configurations within this tolerance of the minimum training loss.

Table 3 reports the realized counts. Bank size is the number of candidate programs available at final refit, including initial and parameter-expanded programs, whereas the selected retriever uses a subset. Proposal attempts are the logged search slots, including a terminal budget-exhausted slot when present. Verification rejection rate counts failures at the reference or generated-suite checks divided by all programs reaching those checks, including accepted and duplicate programs in the denominator. Specification failures, proposals rejected before these checks, and budget-exhausted slots do not enter that rate. Only the DeepSeek-V4-Flash run on ooExp26c completed all six planned rounds; the other runs stopped at their LLM-request or token budgets.

A.3 Metric Settings

CMMD ratios. We report action and belief CMMD ratios. For BiRG-finite, both components condition on the preceding L bilateral-action records, the focal participant's own absolute-belief history, response-round time, and treatment or payoff context. CMMD-A uses the ordered bilateral-action sequence starting at the response round, whereas CMMD-B uses the ordered own absolute-belief sequence as its response. For ooExp26c, both components use the shared previous-period condition consisting of the focal participant's bid curve and expectation curve, bilateral settlement

pairs, opponent-visibility masks, and period fraction; the response is the next bid curve for CMMD-A and the next absolute-expectation curve for CMMD-B. Here L is the history length and h is the future-response horizon. The main tables use $(L, h) = (1, 1)$ for both datasets, so their CMMD columns measure one-step history–response compatibility. For the repeated-game BiRG-finite data, we additionally evaluate $L, h \in \{1, 2, 3\}$ to study multi-round history–response compatibility. Both ratios use $\lambda_i = n_{\mathrm{real},i}^{-1/4}$, where $n_{\mathrm{real},i}$ is the number of eligible real windows for subject i in the corresponding cell; the same value is used for the numerator and real–real denominator.

λ-MAE and Fisher-CI IoU. For the main comparisons, we fit a logit quantal-response parameter separately to each participant's human and generated belief–action observations and average the absolute parameter difference and 95% Fisher-CI IoU over matched participants. BiRG-finite pools belief-elicited observations from supergames 5–7 within each participant and fold, then minimizes the Bernoulli negative log likelihood plus $\lambda^2/(2 \cdot 2^2)$ over $[-10, 10]$. Its interval is $\hat{\lambda} \pm 1.96/\sqrt{\sum_t p_t(1-p_t)(\Delta U_t)^2 + 1/4}$: the prior precision keeps the interval finite when the likelihood curvature is small. We retain participants with all-identical actions or boundary estimates; participants with fewer than two nonmissing belief–action pairs have no fit and do not enter the matched-participant average. These intervals are not clipped to the fitting bounds. BiRG-finite metrics are averaged over three participant folds.

For ooExp26c periods 1–6, we hold each elicited opponent-bid belief curve fixed, snap observed bids to the $[-50, 50]$ action grid at unit spacing, and maximize the participant-level softmax likelihood over 121 logarithmically spaced λ values in $[0.01, 10]$. Nonfinite decision slots are skipped; we apply no participant-level outlier trimming and retain boundary estimates. The 95% Fisher interval uses a centered log-likelihood second difference with step 10^{-4}, an information floor of 10^{-10}, and clipping to the grid bounds. For a boundary estimate, the curvature evaluation point is moved inside the bounds by one step; if interval computation fails, the full $[0.01, 10]$ range is used. Thus near-deterministic choices yield a finite, potentially wide interval rather than an infinite standard error. CI IoU is averaged over matched participants with a 10^{-12} denominator floor for zero-width unions.

Synthetic substitution utility. We replace real training observations with generated records at probability $p = 0.5$, holding the learner, training-set size, and training budget fixed. We compare held-out prediction against matched real-only training; no records are added to the training set. The action task predicts current actions from historical actions and settlement outcomes; the belief task predicts current beliefs from historical beliefs and settlement outcomes.

A.4 Metric Calibration Tables

The CMMD values in Table 4 are the numerical values underlying the curves in Figure 3(a,b). The corresponding model-relative belief–action consistency values are reported in Table 5, which provides the numerical values for Figure 3(c).

Table 4: CMMD calibration under controlled conditional mismatch using the one-step window $(L, h) = (1, 1)$. The entries correspond to the curves in Figure 3(a,b).

Data	Component	Ratio at target mismatch fraction				
		0	0.25	0.5	0.75	1.0
BiRG	action	1.033	1.116	1.230	1.304	1.386
BiRG	belief	1.027	1.163	1.307	1.446	1.603
ooExp	action	1.002	1.003	1.008	1.012	1.012
ooExp	belief	1.002	1.004	1.008	1.012	1.012

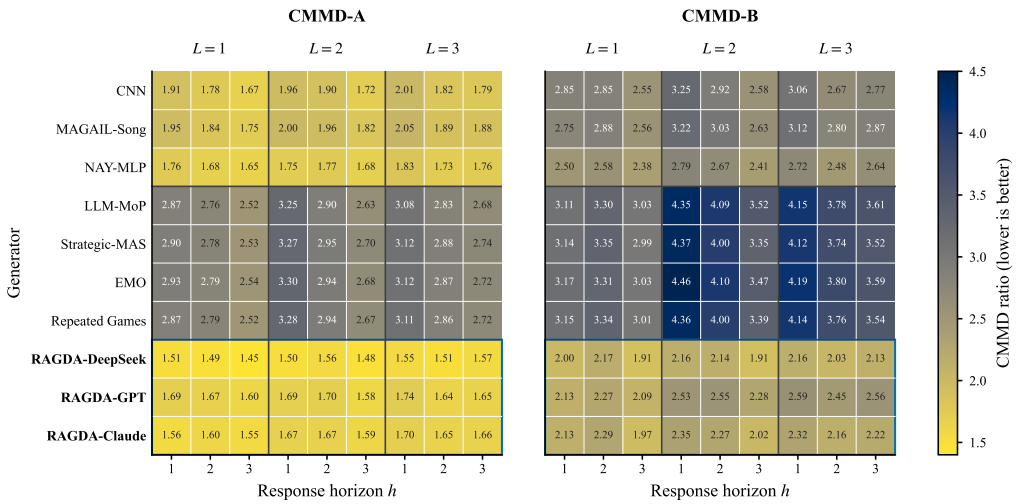


Figure 4: **BiRG-finite CMMD sensitivity across history lengths and response horizons.** The two panels report CMMD-A for action responses and CMMD-B for belief responses across all ten generators. Columns are grouped by history length $L \in \{1, 2, 3\}$, and each group varies the response horizon $h \in \{1, 2, 3\}$. Lower ratios indicate closer distributional history–response compatibility with human participants.

Table 5: Calibration of λ-MAE and Fisher-CI IoU under subject-conditioned action mismatch, with the belief panel held fixed. Unlike the main BiRG-finite comparison, this diagnostic fits each participant–supergame block and combines estimates using likelihood-Fisher weights. The entries correspond to the curves in Figure 3(c).

Data and metric	Mismatch fraction ρ			
	0	0.25	0.5	1.0
BiRG λ-MAE	0.000	0.403	0.699	1.106
BiRG Fisher-CI IoU	1.000	0.681	0.496	0.341
ooExp λ-MAE	0.000	0.093	0.119	0.184
ooExp Fisher-CI IoU	1.000	0.546	0.434	0.259

A.5 CMMD across History Lengths and Response Horizons

Figure 4 compares all ten generators across nine (L, h) combinations. The three RAGDA configurations consistently showed lower CMMD-A and CMMD-B ratios than the baselines, indicating closer agreement with human history–response relationships across different history lengths and response horizons.

A.6 Selected Retrieval Items and Weights

This table lists the retrieval items selected by each authoring LLM for BiRG-finite and ooExp26c, their comparison logic, and their final normalized coefficients. Candidate items are authored by the LLM; the frozen search then selects items and fits coefficients using training data with validation tie selection. Family names follow Figure 2. Each $\bar{\alpha}_k$ is the normalized product of the effective family coefficient and the item's within-family coefficient. BiRG entries are ordered as F0/F1/F2; "–" denotes an item absent from that fold. Coefficients are rounded to three decimals. Under the BiRG data contract, own outcome is aliased to own commitment; unlisted families are inactive, and block position is unavailable in ooExp26c.

Table 6: Selected retrieval items authored by GPT-5.6 Luna. BiRG-finite weights are ordered as F0/F1/F2; "–" denotes that an item was not selected in that fold.

Family	Selected item	Logic	Weight $\bar{\alpha}_k$
		BiRG-finite (F0/F1/F2)	
Local temporal position	*Neutral phase distance*	Normalized phases with a neutral penalty for unavailable indices.	.270/–/–
	Ordinal bucket	Compares early, middle, and late round buckets.	–/.096/.000
	Ordinal phase	Compares normalized within-supergame positions.	–/.004/.046
Block position	*Logistic block gap*	Bounded transform of normalized block-position separation.	.374/–/–
	Missing-neutral block gap	Symmetric missing-value handling for block positions.	–/.112/.227
Own commitment	*Coverage-aware alignment*	Aligned recent commitments retain missing-value disagreement.	.053/–/–
	Recency-aligned distance	Aligned recent commitments preserve temporal order.	.052/–/–
	Recency profile	Compares recency-ranked commitment levels.	–/.105/.120
	Recent medoid gap	Compares representative recent commitment levels.	–/.104/.252
Counterpart observables	*Opponent CDF + count*	Opponent-action distributions with unequal-history correction.	.251/–/–
	Opponent level histogram	Binned opponent-action levels plus a recency discrepancy.	–/.580/.355
		ooExp26c	
Local temporal position	*Common-horizon rank*	Compares period positions on a common ordinal horizon.	.284
	Triangular temporal profile	Midpoint-centered profile of normalized periods.	.241
Own commitment	*Ordinal commitment pattern*	Ordinal commitment pattern across value nodes.	.271
	Median commitment gap	Compares robust median commitment levels.	.071
Own outcome	*Missing-history penalty*	Penalizes unequal realized-outcome histories.	.076
	Period transport	Transports unmatched period profiles.	.057

Table 7: Selected retrieval items authored by DeepSeek-V4-Flash. BiRG-finite weights are ordered as F0/F1/F2.

Family	Selected item	Logic	Weight $\bar{\alpha}_k$
		BiRG-finite (F0/F1/F2)	
Local temporal position	*Cyclic within-block phase*	Compares round positions using a cyclic phase distance.	.162/.333/.083
	Median-centered ordinal deviation	Compares round ranks relative to the horizon median.	.094/.077/.002
Block position	*Normalized block-rank gap*	Compares normalized separation between supergame ranks.	.388/.323/.181
Counterpart observables	*Opponent action-distribution shift*	Compares empirical opponent-action levels with a short-history penalty.	.248/.102/.481
	Most-recent opponent-action shift	Compares the most recent observed opponent actions with a short-history penalty.	.107/.166/.252
		ooExp26c	
Local temporal position	*Normalized ordinal period-rank gap*	Compares normalized ordinal period positions.	1.000

Table 8: Selected retrieval items authored by Claude Haiku 4.5. BiRG-finite weights are ordered as F0/F1/F2.

Family	Selected item	Logic	Weight $\bar{\alpha}_k$
		BiRG-finite (F0/F1/F2)	
Local temporal position	*Horizon-scaled ordinal gap*	Compares round positions after scaling by the horizon.	.153/.042/.015
	Relative-phase alignment	Compares relative positions within the supergame horizon.	.167/.019/.022
Block position	*Cyclic block position*	Compares supergame positions using cyclic distance.	.081/.022/.149
	Missing-robust linear block-position distance	Compares linear block positions with explicit missing-value handling.	.108/.011/.145
Own commitment	*Own-action activity ratio*	Compares recent commitment activity with overall activity.	.007/.177/.271
Counterpart observables	*Cumulative opponent-path drift*	Compares cumulative drift of the observed opponent-action path.	.483/.728/.397
		ooExp26c	
Local temporal position	*Period quantile-rank gap*	Compares periods by their quantile ranks.	1.000

A.7 BiRG-finite Rematch Case Studies

Figure 5 illustrates three qualitative BiRG-finite rematch cases. Each row pairs multiple complete real SGs (left) with a rematch prefix (right). The SGs show recurring action–belief patterns, and the final rematch cell is the next response generated by each method. The cases ask whether the action fits the interaction state and whether belief is coherent with that action. The three rows cover persistent defection, stable cooperation, and betrayal during cooperation.

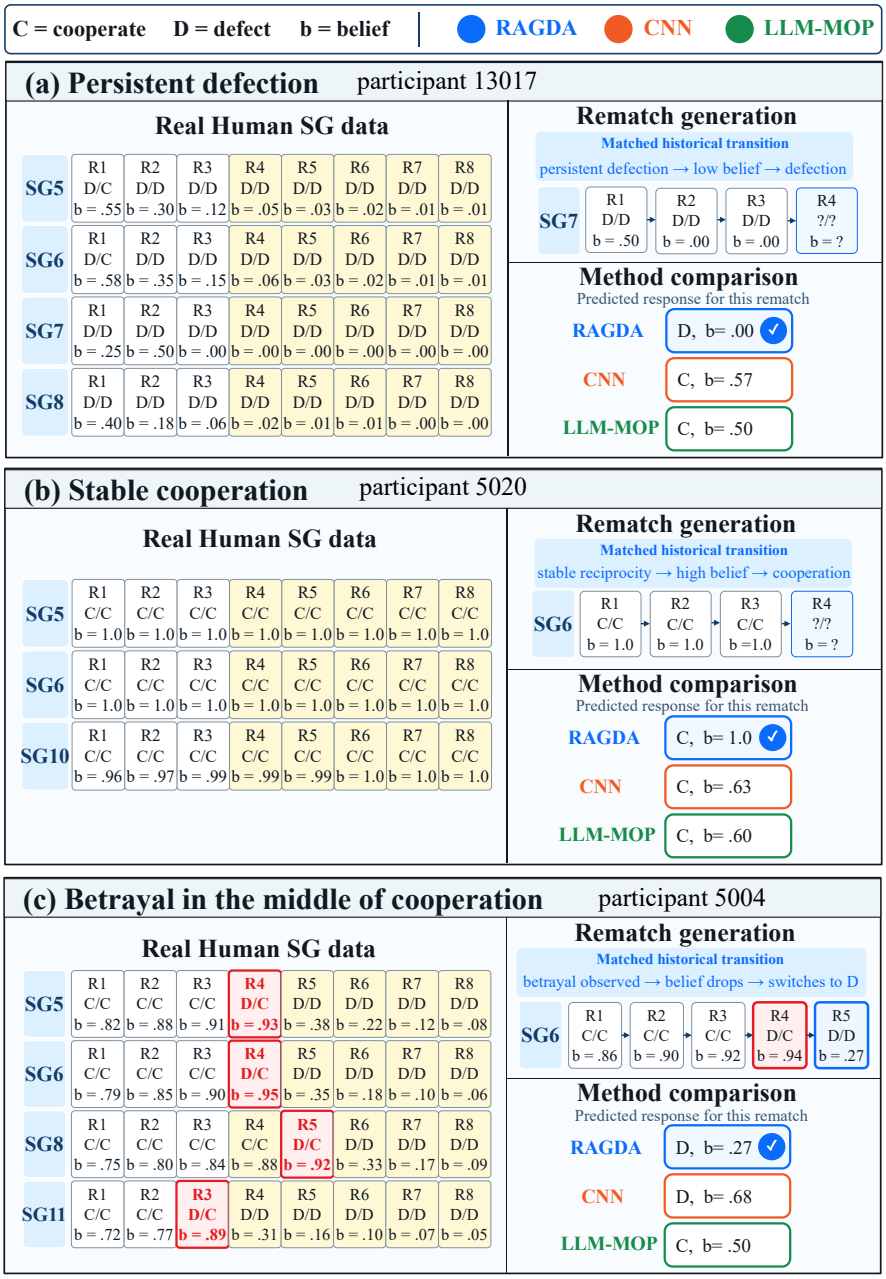


Figure 5: **BiRG-finite rematch cases.** Each row combines multiple complete real human SGs with a rematch prefix whose final response is generated by the compared methods. The cases show persistent defection with low belief (a), stable cooperation with high belief (b), and an opponent betrayal followed by an action switch and belief drop (c). C and D denote cooperation and defection, b denotes belief, and X/Y denotes opponent/participant actions. The examples illustrate local action plausibility and belief–action alignment with observed cases.

Across the three cases, RAGDA captures the appropriate response action and associated belief, matching their joint pattern in the observed cases. In (a), it reproduces persistent defection with low belief, whereas both baselines switch to cooperation. In (b), it preserves stable cooperation with high belief, while the baselines assign lower beliefs. In (c), it captures both the belief drop and switch to defection after betrayal; CNN captures the action switch but retains high belief, while LLM-MOP misses the joint transition. Thus, RAGDA consistently matches the action–belief response indicated by the real histories.

A.8 ooExp26c History-Conditioned Case Study

Unlike BiRG-finite, ooExp26c consists of ordered one-shot auctions, so the relevant transition is from a participant's earlier period history to a later period response. Figure 6 follows one focal participant ("sid=658"). The upper block shows three observed transitions, while the lower block holds the real $P4$ response fixed and replaces only the preceding $P3$ history with each method's generated bid and belief curves. Across the real transitions, action and belief rise at all six private-value bins. RAGDA produces a $P3$ record compatible with this joint upward direction; CNN is partially aligned, whereas LLM-MOP is mostly opposite. Feedback is the visible-opponent-bid mean within each bin; generated-history feedback is recomputed under the common rematch rule, and the CNN/LLM-MOP entries are derived because their original records lack this field. The panel is a qualitative diagnostic of local history–response compatibility, complementary to the aggregate metrics and substitution-utility results.

Observed real transitions

Real transition P2→P3

Private-value bin	v<5	5≤v<15	15≤v<25	25≤v<35	35≤v<45	v≥45
Opponent feedback	1.65	—	34.33	—	—	32.14
Δ action	↑ +10	↑ +10	↑ +10	↑ +10	↑ +10	↑ +10
Δ belief	↑ +15	↑ +20	↑ +20	↑ +20	↑ +20	↑ +20

Real transition P3→P4

Private-value bin	v<5	5≤v<15	15≤v<25	25≤v<35	35≤v<45	v≥45
Opponent feedback	30.72	17.18*	21.10	—	—	—
Δ action	↑ +10	↑ +10	↑ +10	↑ +10	↑ +10	↑ +10
Δ belief	↑ +10	↑ +10	↑ +10	↑ +10	↑ +10	↑ +10

Real transition P7→P8

Private-value bin	v<5	5≤v<15	15≤v<25	25≤v<35	35≤v<45	v≥45
Opponent feedback	6.86	11.15*	24.33	—	—	—
Δ action	↑ +10	↑ +10	↑ +10	↑ +10	↑ +10	↑ +10
Δ belief	↑ +15	↑ +15	↑ +15	↑ +15	↑ +15	↑ +15

Observed pattern: action and belief rise at all six nodes in every transition.

Generated P3 history → the same real P4 response

Rematch opponent: sid=703. Real P4 action = belief = [−10, 0, 10, 20, 30, 40].

RAGDA All 6 action and belief changes are upward.

Private-value bin	v<5	5≤v<15	15≤v<25	25≤v<35	35≤v<45	v≥45
Opponent feedback	-5.30	2.48*	17.14	—	35.57	—
Δ action	↑ +20	↑ +20	↑ +20	↑ +20	↑ +20	↑ +20
Δ belief	↑ +25	↑ +30	↑ +30	↑ +30	↑ +30	↑ +30

CNN Belief rises throughout; action rises at 5 of 6 nodes.

Private-value bin	v<5	5≤v<15	15≤v<25	25≤v<35	35≤v<45	v≥45
Opponent feedback	3.81	8.25*	17.46	—	37.60	—
Δ action	↓ -1.25	↑ +1.85	↑ +3.85	↑ +3.96	↑ +3.37	↑ +2.76
Δ belief	↑ +5.16	↑ +8.33	↑ +10.22	↑ +10.75	↑ +10.75	↑ +9.69

LLM-MOP Action mostly falls; belief falls at 5 nodes and is flat at 1.

Private-value bin	v<5	5≤v<15	15≤v<25	25≤v<35	35≤v<45	v≥45
Opponent feedback	—	—	—	—	35.14	—
Δ action	↓ -10	↓ -10	↓ -10	↓ -1	— 0	↑ +10
Δ belief	↓ -10	↓ -10	↓ -10	↓ -10	↓ -10	— 0

Figure 6: **ooExp26c history-conditioned case study.** The upper block shows three observed transitions for the same focal participant ("sid=658") at six private-value nodes. The lower block compares each generated $P3$ history with the fixed real $P4$ response. Feedback entries are visible-opponent-bid means by the focal participant's private-value bin; an asterisk marks two observations and "–" denotes no observation. RAGDA produces a $P3$ record compatible with the upward action–belief direction, CNN is partially aligned, and LLM-MOP is mostly opposite. Feedback for CNN and LLM-MOP is derived under the common rematch rule rather than emitted by their original rollouts. Δ is later/target minus earlier/generated.