HumorGen: Cognitive Synergy for Humor Generation in Large Language Models via Persona-Based Distillation

Anonymous ACL submission

Abstract

Humor generation poses a significant challenge for Large Language Models (LLMs), because their standard training objective—predicting the most likely next word—inherently conflicts with the surprise and incongruity needed for comedy. To bridge this gap, we introduce the **Cognitive Synergy Framework**, a theoretically grounded methodology for generating high-quality humor data inspired by psychological theories of humor. Utilizing a *Mixture-of-Thought (MoT)* approach, we deploy six cognitive personas (e.g., *The Absurdist*, *The Cynic*) to synthesize diverse comedic perspectives for a given prompt. This framework creates a theoretically grounded dataset, which we use to fine-tune a 7B-parameter student model. We compare **Direct Preference Optimization (DPO)** and a novel **Offline Group Relative Policy Optimization (O-GRPO)**; our 7B model significantly outperforms larger instruction-tuned baselines and achieves performance competitive with state-of-the-art proprietary models. We find that cognitive-driven data curation is far more critical than alignment algorithms or model scale for humor generation. Code and data are available at `https://anonymous.4open.science/r/HumorGen-2697`.

1 Introduction

Humor generation is a sophisticated creative task requiring mastery of context, nuance, and linguistic ambiguity (Khurana et al., 2024; Robison, 2024). While Large Language Models (LLMs) excel at logical reasoning, reliable humor generation remains an open problem because standard training objectives—minimizing perplexity—conflict with the incongruity and surprise required for comedy. This "alignment tax" often results in models that are safe and helpful but produce predictable, boring jokes or tedious explanations of humor.

Recent efforts to improve LLM humor generation have focused on logical "thought leaps" (Zhong et al., 2024) or multistep reasoning (Wang et al., 2025). While these improve performance in their specific humor generation tasks, they do not guarantee accurate humor generation and often miss the diverse cognitive styles behind human humor. These existing methods rely on instruction tuning, which fails to learn a diverse representation of humor, and because they do not cover all aspects of humor, they fail to capture the variety of ways humans actually construct jokes.

To bridge this gap, we introduce the **Cognitive Synergy Framework**. We advance beyond generic instruction-tuning by operationalizing psychological humor theories into

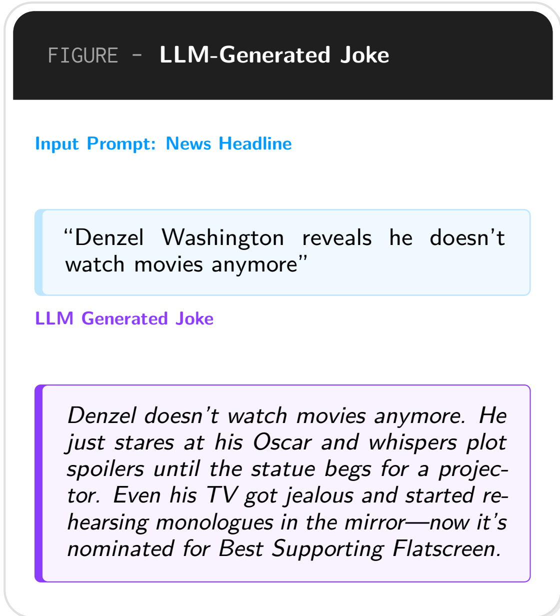


Figure 1: Example of an LLM-generated joke based on a news headline prompt, synthesized using the Cognitive Synergy Framework.

a *Mixture-of-Thought (MoT)* architecture explicitly designed for creative divergence. Traditional language modeling is highly susceptible to mode collapse in creative generation, converging toward the most probable—and therefore most generic—continuations. By instantiating six distinct "cognitive personas" (e.g., *The Absurdist*, *The Cynic*) as latent experts within the MoT framework, we consistently route the generation process into the low-probability, high-variance regions of the semantic space where humor naturally occurs. This ensemble approach mitigates mode collapse and yields a diverse, theoretically grounded corpus of synthetic data, enabling us to distill multifaceted humor generation capabilities from a frontier teacher model into a highly efficient 7B-parameter student model.

Due to the highly subjective nature of humor, we investigate whether preference alignment (e.g., **Direct Preference Optimization (DPO)** and **Offline Group Relative Policy Optimization (O-GRPO)**) improves over supervised fine-tuning. Our experiments show that neither alignment method improves the models over the SFT baseline: DPO achieves similar performance to SFT, while O-GRPO is less impressive. Thus, under our setup, the alignment exercises did not improve the models, and the quality of the underlying cognitive data (Cognitive Synergy Framework) is the primary driver of generation performance.

Our contributions are:

- We introduce the **Cognitive Synergy Framework**, a methodology for generating diverse, high-quality humor data by deploying specialized psychological personas as latent experts.
- We **investigate** whether preference alignment (DPO, O-GRPO) improves over SFT for humor generation. We find that neither alignment method improves the models: DPO achieves similar performance to SFT, while O-GRPO is less impressive; the alignment exercises did not improve the models beyond high-quality SFT data in this subjective domain.
- We show that our 7B student model, **HumorGen**, achieves state-of-the-art results for open-weights models and performs competitively with much larger proprietary systems, proving that high-quality data is more important than model size for humor.

2 Related Work

2.1 Computational Humor Generation

Computational approaches to humor have predominantly focused on detection and recognition tasks (Jentzsch and Kersting, 2023; Dsilva, 2024), while generative capabilities have received comparatively less attention. Consequently, research in humor generation has been fragmented, with studies often limited to specific humor types like puns (Chen et al., 2024), specific domains (Shafiei and Saffari, 2025; Zhang et al., 2020), or specific languages (Chen et al., 2023; Zhong et al., 2024).

This fragmentation stems from the inherent subjectivity of humor, where the perception of funniness is heavily dependent on cultural context, situational nuance, and the recipient's background (Wanzer et al., 1995; Olson and Roese, 1995). While such domain-specific restrictions are often justified by these complexities, there is a growing need for models capable of transcending cultural and linguistic barriers to generate diverse forms of humor. Furthermore, although classical theories of humor (Lintott, 2016; Scheel, 2025; McGraw and Warren, 2010) do not offer a complete generative recipe, they remain essential for characterizing the linguistic and semantic elements utilized in humorous discourse.

2.2 Reasoning-Enhanced Humor Creativity

Since the advent of Large Language Models (LLMs), researchers have begun exploring specialized prompting strategies for humor generation, yielding significant insights into the limitations of standard reasoning approaches. Zhong et al. (Zhong et al., 2024) emphasized the need for a distinct thought process in prompting LLMs for humor, noting that conventional Chain-of-Thought (CoT) reasoning(Wei et al., 2022) is often ineffective for creative tasks. Even state-of-the-art LLMs frequently struggle to produce high-quality comedic content when relying on established prompting strategies like

CoT (Zhong et al., 2024; Wang et al., 2025).

While CoT is highly effective for logical, sequential tasks, it is ill-suited for the creative, divergent thinking required for humor, which necessitates non-linear associations and incongruity—traits that inherently conflict with the logical progression optimized in reasoning models (Tikhonov and Shtykovskiy, 2024). Consequently, even advanced reasoning models frequently generate outputs that are logically sound but lack the necessary comedic surprise.

To address this, Zhong et al. (Zhong et al., 2024) introduced *Creative Leap of Thought (CLoT)*, a novel reasoning technique that leverages associative games (Oogiri-GO) to encourage "leap-of-thought"—the ability to make non-obvious connections between unrelated concepts. Building on this, Wang et al. (Wang et al., 2025) proposed the *LoL* framework, which aims to inject external information to mitigate knowledge graph sparsity, thereby enabling multi-hop reasoning for creative generation. Similarly, Tikhonov and Yamshchikov (Tikhonov and Shtykovskiy, 2024) leveraged multistep reasoning structures specifically tailored for humor generation. In contrast, Jentzsch and Kersting (Jentzsch and Kersting, 2023) explored naive joke generation with ChatGPT using simple prompts, discovering that 90% of the 1,008 generated jokes were repetitions of the same 25 examples. These findings collectively demonstrate that the specific logic required for humor generation demands specialized approaches beyond standard reasoning paradigms.

2.3 Preference Optimization for Subjective Tasks

Different preference learning and alignment approaches have been explored to align LLMs with human expectations, particularly in subjective tasks where a single "correct" answer is undefined (Yasuda and Toda, 2025; Lou et al., 2025; Vikhorev et al., 2024). Reinforcement learning (RL) alignment approaches such as Direct Preference Optimization (DPO) and Group Relative Policy Optimization (GRPO) have proven effective for alignment in diverse domains, including code generation (Govande et al., 2025) and image generation (Tong et al., 2025).

In humor generation, Wang et al. (Wang et al., 2025) integrated preference learning through a two-stage process: Supervised Fine-Tuning (SFT) followed by a DPO stage to align the model with humor preferences. However, standard preference optimization methods typically rely on online sampling or pairwise comparisons, which can be computationally expensive and unstable for highly subjective tasks like humor. Our work builds on these foundations but introduces an offline group-relative formulation (O-GRPO), enabling efficient alignment from fixed preference datasets without the overhead of online sampling.

3 The Cognitive Synergy Framework

Generating humor requires valid logical reasoning to set up a context, followed by a sudden conceptual shift that subverts expectations. Standard LLM decoding strategies, which typically maximize the probability of the most likely next token, are often at odds with this requirement. To address this, unlike prior work (Wang et al., 2024), we introduce the **Cognitive Synergy Framework**, which adapts the Mixture-of-Thought (MoT) paradigm (Fein-Ashley et al., 2025) to the domain of humor by explicitly modeling divergent thinking through distinct *Cognitive Personas*.

3.1 Divergent Reasoning via MoT

Unlike standard Chain-of-Thought (CoT) prompting, which optimizes for a single logical path, our framework generates K distinct reasoning traces in parallel. This mimics the creative process of exploring multiple comedic angles—such as irony, absurdity, or wordplay—before selecting the best punchline. Given an input premise x, we sample a set of diverse reasoning paths $\{z_1, z_2, \ldots, z_K\}$ seeded by different cognitive priors. This approach ensures that the model explores the "long tail" of creative possibilities rather than defaulting to the most probable (and often least funny) response.

3.2 Cognitive Personas

To guide this diversity, we define six **Cognitive Personas**, each grounded in a specific psychological theory of humor (Table 1). These personas act as soft constraints on the reasoning process, ensuring that our candidate pool covers a wide spectrum of comedic mechanisms.

Persona	Humor Theory	Mechanism	Cognitive Focus
Neurotic	Relief Theory	Tension Release	Internal anxiety, overthinking, and social insecurity.
Cynic	Superiority Theory	Social Critique	Hypocrisy, biting sarcasm, and moral contradictions.
Observer	Incongruity	Social Mapping	Mundane minutiae and unwritten awkward social norms.
Wordsmith	Linguistic	Ambiguity	Puns, double entendres, and phonological play.
Optimist	Benign Violation	Recontextualization	Wholesome misinterpretations of potentially negative traits.
Absurdist	Incongruity	Surrealism	Non-sequiturs, dream logic, and fractured causality.

Table 1: The six Cognitive Personas used in our framework. We map each persona to a foundational humor theory and a specific cognitive focus to ensure divergent candidate generation.

By using these personas, we created a "synergy" between different styles of thought. This structural diversity proved critical for our subsequent alignment stage, as it provided a rich variety of distinct candidates for the model to learn from during preference optimization.

4 Methodology

We frame humor generation as a conditional language modeling task where the goal is to generate a humorous response y given a context x, minimizing the divergence between the model's output and learned preference distributions derived from LLM-judged pairwise evaluations. We explore two distinct preference alignment strategies following a shared supervised fine-tuning stage: Direct Preference Optimization (DPO) (Rafailov et al., 2023) and our proposed Offline Group Relative Policy Optimization (O-GRPO).

4.1 Supervised Fine-Tuning (SFT)

This initial stage establishes baseline humor capabilities and internalizes the various cognitive personas. We construct a dataset $\mathcal{D}_{SFT}$ using a "Silver Teacher" protocol. Given the candidate pool $\mathcal{C}_{total}$ generated by our Mixture-of-Thought (MoT) ensemble, we employ a pairwise LLM evaluation system to compute Elo ratings for all candidates. We select the top-ranked candidates for each prompt based on these Elo ratings:

$$y^* = argmax_{y \in \mathcal{C}_{total}} Score_{LLM}(y|x)$$

We fine-tune a base Qwen-7B model using standard cross-entropy loss to maximize the likelihood of these "winner" responses:

$$\mathcal{L}_{SFT}(\theta) = -E_{(x,y^*)\sim\mathcal{D}_{SFT}}[\log \pi_\theta(y^*|x)]$$

This stage effectively distills the creative diversity of the larger teacher model into the student model.

4.2 Direct Preference Optimization (DPO)

To further align the model with humor preferences, we employ DPO using a dataset $\mathcal{D}_{DPO}$ of high-quality pairwise preferences derived from the LLM-judged Elo rankings. Each pair (y_w, y_l) consists of a high-ranking joke y_w and a low-ranking candidate y_l for the same prompt, selected based on their Elo gap. We optimize the policy π_θ directly without a reward model:

$$\mathrm{L}_{DPO}(\pi_\theta; \pi_{ref}) = -E_{(x,y_w,y_l)\sim\mathcal{D}_{DPO}}\Bigg[\log\sigma\Bigg(\beta\log\frac{\pi_\theta(y_w|x)}{\pi_{ref}(y_w|x)} - \beta\log\frac{\pi_\theta(y_l|x)}{\pi_{ref}(y_l|x)}\Bigg)\Bigg]$$

4.3 Offline Group Relative Policy Optimization (O-GRPO)

Beyond pairwise preference alignment, we explore the potential of group-relative objectives to further refine the model's comedic reasoning. Recent advancements in reinforcement learning, specifically Group Relative Policy Optimization (GRPO) (Shao et al., 2024), have shown promise in stabilizing training for tasks requiring complex logical or creative constraints by utilizing relative rewards within a sample group.

To adapt this to our distillation pipeline, we implement an **offline variant (O-GRPO)**. We generate "Gold Groups" consisting of $G = 24$ fixed candidates per prompt, which are subsequently ranked by our multi-persona LLM judge. This formulation allows us to leverage the variance-reduction properties of group-relative normalization while maintaining the

Figure 2: The HumorGen training pipeline. **(A) Generation:** Input headlines are processed by the Cognitive Synergy module (MoT), generating diverse candidates from 6 distinct personas. **(B) Collation:** Candidates are ranked via a pairwise evaluation system using an LLM judge to compute Elo ratings. **(C) SFT:** The base policy is fine-tuned on the top-ranked candidates. **(D) Alignment:** The model is further optimized via two parallel experimental branches: Pairwise DPO (top) or Group-Relative O-GRPO (bottom) driven by Elo-based preference data.

computational stability of an offline optimization process. By incorporating O-GRPO, we seek to determine if maximizing the relative advantage of high-quality responses within a broader candidate pool provides additional signal beyond the standard SFT and DPO objectives.

For each group, we compute the advantage A_i of candidate y_i relative to its peers using their Elo scores:

$$A_i = \frac{r_i - \mu_{group}}{\sigma_{group} + \epsilon} \tag{1}$$

where r_i is the Elo rating of the candidate. To learn efficiently from these pre-computed advantages, we formulate O-GRPO as an **Exponentially Weighted SFT** objective. This formulation avoids the complexities of PPO-style clipping:

$$\mathcal{L}_{O-GRPO}(\theta) = -E_{x\sim\mathcal{D}}\left[\sum_{i=1}^{G} w_i \log \pi_\theta(y_i|x)\right] \tag{2}$$

The weights w_i are derived from the advantages using a softmax temperature T:

$$w_i = \frac{\exp(A_i/T)}{\sum_{j=1}^{G}\exp(A_j/T)} \tag{3}$$

This objective aggressively promotes candidates with high relative advantages while suppressing those that underperform relative to the group.

4.4 Cognitive Synergy Distillation (CSD)

In the base pipeline, the teacher's persona-specific reasoning traces are discarded after generation, and the student sees only the final jokes. CSD changes this: the student is trained on the teacher's reasoning *alongside* the joke: `<think>` *persona-specific brainstorming* `</think>` *joke*

This is process distillation—the student learns not just *what* to generate but *how* the teacher planned it. For DPO, both chosen and rejected responses include reasoning traces (symmetric format), so the model cannot shortcut by learning that the mere presence of reasoning correlates with winning; it must learn which *content* leads to better jokes.

At inference, the model generates reasoning followed by the joke. The reasoning is stripped for evaluation (ensuring fair comparison with non-CSD models) but retained for interpretability. Unlike generic CoT, ineffective for humor (Zhong et al., 2024; Tikhonov and Shtykovskiy, 2024), CSD's reasoning is grounded in specific humor theories through the cognitive personas, making it a form of *theory-grounded creative distillation.*

5 Experimental Setup

5.1 Datasets and Data Synthesis

We utilize the official SemEval 2026 Task 1 (MWAHAHA) experimental set (Castro et al., 2026), comprising 1,200 news headlines and word-pair prompts as inputs to our generation pipeline. Using the Cognitive Synergy Framework, we generate 24 candidates per prompt (4 per persona × 6 personas) from a teacher ensemble of *Kimi-K2* and *Qwen 2.5-32B-Instruct*, yielding a raw pool of ∼28,800 candidates. These candidates are scored and ranked via a pairwise LLM evaluation system using Llama 3.3-70B-Instruct as the judge, producing per-prompt Elo ratings for all 24 candidates. We construct three training subsets from these rankings:

- SFT Data ($\mathcal{D}_{SFT}$, $N = 12,000$): For each of the 1,200 prompts, we select the top 10 Elo-ranked candidates (rather than only the single best). Using multiple top-ranked candidates per prompt avoids mode collapse: the student learns a diverse range of humor styles (e.g., wordplay, absurdity, sarcasm) instead of collapsing toward one dominant style.
- DPO Data ($\mathcal{D}_{DPO}$, $N = 6,000$): For each prompt, we construct 5 preference pairs by randomly pairing candidates from the top-5 Elo-ranked jokes (chosen, y_w) with candidates from the bottom-5 Elo-ranked jokes (rejected, y_l). This yields 5 pairs × 1,200 prompts = 6,000 preference pairs, with a sharp quality gap between chosen and rejected responses.
- O-GRPO Data ($\mathcal{D}_{GRPO}$): We use all 24 candidates per prompt across the 1,200 prompts, computing normalized group-relative Elo advantages per group ($G = 24$). This exposes the model to the full quality spectrum within each prompt group.

The official SemEval 300-prompt evaluation set is held out entirely for final testing; automated pairwise evaluation is run on a 50-prompt subset of this set.

5.2 Baselines

We compare against the following models: Vanilla Qwen-7B (untuned base model), Qwen 2.5-32B-Instruct (teacher model used during data synthesis), Kimi-K2 (teacher model and upper-bound baseline), GPT-OSS-120B, GPT-5, and Gemini-2.5-Pro (frontier models). For the CSD ablation, base-trained models (HumorGen-SFT, HumorGen-DPO, HumorGen-GRPO) are compared against their Think variants (HumorGen-SFT-Think, HumorGen-DPO-Think, HumorGen-GRPO-Think).

5.3 Implementation Details

All models were trained on NVIDIA H100 (80GB) GPUs using LoRA(Hu et al., 2022) (r=16) with the Unsloth library. SFT ran for 3 epochs; DPO and O-GRPO for 5 epochs, both with early stopping (patience=2). Candidate ranking for the full pool consumed ∼132 H100 node-hours. For O-GRPO, groups of $G = 24$ candidates per prompt maximize the advantage-weighted learning signal.

5.4 Evaluation Protocols

Ranking methodology. For each pair of jokes (A, B), the LLM judge selects the funnier one (or tie). Presentation order is randomized per match to mitigate position bias. We aggregate all match outcomes into a full contest matrix, fit a Bradley-Terry (BT) model(Gao et al., 2025) via MM algorithm to estimate latent ratings, and report Elo-scale ratings with 95% bootstrap confidence intervals (100 bootstrap samples). Key model comparisons have non-overlapping CIs and are statistically significant.

1. Automated Pairwise Evaluation: We evaluate all trained models on a 50-prompt subset of the held-out test set, generating jokes and ranking them via the above pipeline. In total, 43,048 pairwise comparisons were conducted, judged by Llama 3.3-70B-Instruct.
2. Human Validation: Human evaluators judge 60 curated pairwise comparisons across 12 ablation categories; they are blinded to model identity and presentation order is randomized to mitigate position bias. We report inter-annotator agreement, LLM–human consensus, and correlation with automated BT ratings.

6 Results

This section shows our observations from the experiment performed in this research, going from model performance to understanding what makes humor funny based on our judge reasoning.

6.1 Model Performance Comparison

Table 2 summarizes model rankings via HumorRank on the held-out set. HumorGen-SFT-7B and DPO-7B (1083.9, 1079.9) surpass Qwen-32B and GPT-OSS-120B, leading among open-weight models for humor generation. Frontier models (GPT-5, Kimi-K2, Gemini-2.5-Pro) lead; 7B students narrow the gap. Key differences (e.g., SFT-7B vs. Qwen-32B) have non-overlapping 95% CIs and are statistically significant.

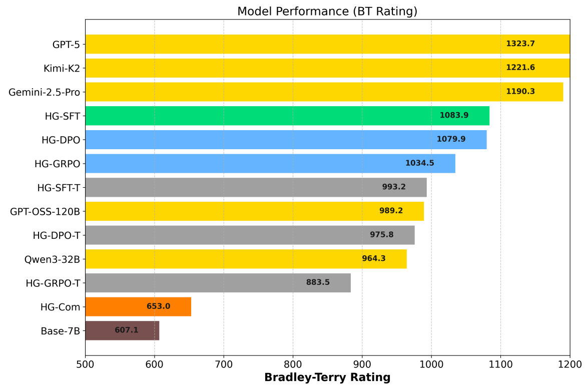


Figure 3: Bradley-Terry ratings with 95% confidence intervals. HG = HumorGen; -T = Think; Gem2.5 = Gemini-2.5-Pro; Qw = Qwen.

Model	BT Rating	95% CI	Win%
GPT-5	**1323.7**	[1288, 1365]	84.7
Kimi-K2	1221.6	[1188, 1260]	75.3
Gemini-2.5-Pro	1190.3	[1157, 1225]	72.0
HumorGen SFT-7B	1083.9	[1057, 1114]	59.5
HumorGen DPO-7B	1079.9	[1055, 1108]	59.0
HumorGen GRPO-7B	1034.5	[1001, 1064]	53.3
HumorGen SFT-Think-7B	993.2	[965, 1024]	48.2
GPT-OSS-120B	989.2	[957, 1012]	47.7
HumorGen DPO-Think-7B	975.8	[950, 1002]	46.0
Qwen3-32B	964.3	[936, 995]	44.5
HumorGen GRPO-Think-7B	883.5	[848, 921]	35.0
HumorGen-Com-7B	653.1	[610, 695]	13.8
Base Qwen-7B	607.1	[560, 647]	10.8

Table 2: Bradley-Terry ratings from pairwise comparisons. HumorGen students outperform the 32B model and state-of-the-art 120B open weights model.

Beyond the rankings, pairwise win rates are in Figure 4 (Appendix A).

6.2 Preference Alignment

We investigated whether preference alignment (DPO, O-GRPO) would improve over our SFT baseline. Neither alignment method improved the models: DPO (1079.9) achieves similar performance to SFT (1083.9), while O-GRPO (1034.5) is less impressive. Thus, the alignment exercises did not improve the models beyond the gains from high-quality SFT data (Cognitive Synergy Framework). All fine-tuned variants substantially outperform base Qwen-7B (+427–476 BT points).

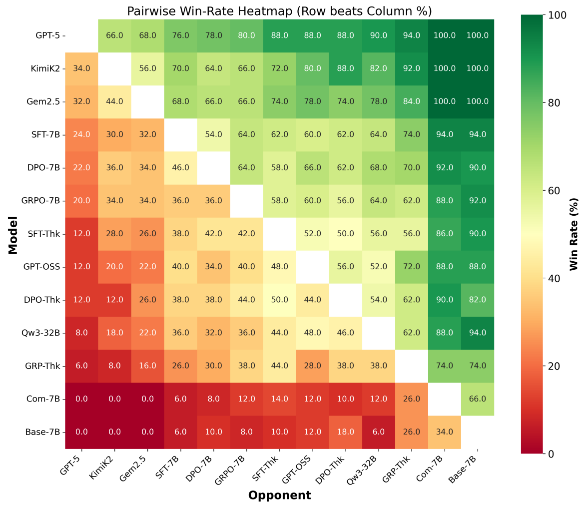


Figure 4: Pairwise win-rate heatmap (row beats column %). Appendix A.

6.3 CSD and the Explainer Trap

The "explainer trap" emerges when we train the 7B HumorGen variants to *think*—i.e., when we apply Cognitive Synergy Distillation (CSD) so the student is trained on the teacher's `<think>` reasoning traces alongside the joke (see § Methodology). Think variants underperform their non-thinking counterparts (Table 2): distilling reasoning traces biases the model toward *explaining* the joke rather than delivering it (Appendix F). We did not evaluate whether the teacher models (Kimi, Qwen 32B) over-explain; the trap may be a distillation artifact when compressing reasoning into the student. This extends prior work: CoT is ineffective for humor (Zhong et al., 2024; Tikhonov and Shtykovskiy, 2024); even training on reasoning traces fails in this setting.

6.4 Comedian Adaptation

Fine-tuning on 998 stand-up jokes (Shaun Eli)(Eli, 2026) regressed sharply (1083.9 → 653.1; Table 2). *Performance-native* stand-up (timing, delivery) differs from *text-native* humor optimized for written punchlines; our CSF data is selected for the LLM medium. See

Appendix I.

6.5 Human Evaluation

Three evaluators gave 180 blind pairwise judgments on 60 curated pairs (12 categories, 5 each) over 50 held-out headlines. Inter-annotator agreement was 31.7% (one-third of pairs), reflecting humor's subjectivity. The LLM judge matched human consensus on 58.3% of pairs (Gold) and individual votes at 52.4% (Micro-Avg). In this "Good vs. Good" regime (high-quality outputs, no objectively worse option), 58.3% indicates the judge captures shared preferences well above chance. Details in Appendix G.

7 Analysis

7.1 What Makes Jokes Win?

Table 3 shows humor feature prevalence among winning jokes (556 matches). Surprise (80%) and absurdity (75%) dominate, confirming expectation violation drives perceived funniness; wordplay and narrative appear in half and one-third of wins.

Feature	Count	% of Wins
Surprise	445	80.0
Absurdity	416	74.8
Wordplay	285	51.3
Incongruity	277	49.8
Narrative	182	32.7
Sarcasm	88	15.8
Irony	86	15.5
Dark humor	78	14.0

Table 3: Humor feature prevalence among winning jokes. Surprise and absurdity co-occur in 80% and 75% of wins respectively, confirming that expectation violation is the dominant driver of perceived funniness.

7.2 Failure Modes

Beyond the explainer trap (§6.3), two common failure patterns are: (1) *generic punchlines* defaulting to safe, high-probability completions, and (2) *overextended setups* burying the joke. See Appendix K for examples.

7.3 Out-of-Domain Generalization

To probe transfer to unseen domains, we generated jokes on **African news headlines**(BBC News, 2026) outside the SemEval training set, using the same prompt format and no persona prompting. Appendix J shows zero-shot outputs from **HumorGen-SFT-7B** and **HumorGen-DPO-7B** on two such headlines (Kenyan weight-loss, Ethiopian smart police stations), suggesting generalization beyond Western-centric training.

8 Conclusion

We introduce the Cognitive Synergy Framework, which operationalizes psychological humor theories into six cognitive personas to generate diverse, high-quality humor data via Mixture-of-Thought. HumorGen achieves strong performance among open-weight models and is competitive with frontier systems—outperforming Qwen-2.5-32B and GPT-OSS-120B baselines—demonstrating that targeted cognitive curation matters more than scale for humor generation. Our central finding is a *data quality ceiling*: when SFT data is diverse and well-curated, preference optimization (DPO, O-GRPO) yields no gains. We show that forced reasoning traces hurt creativity ("explainer trap") and that text-native synthetic data outperforms performance-native stand-up. Human evaluation validates that the LLM judge captures subtle preference tilts in highly subjective "Good vs. Good" comparisons. Future work includes multilingual evaluation, scaling to larger students, extending personas to other creative domains, and exploring **multi-modal** humor (e.g., image-grounded jokes and memes).

9 Limitations

Our evaluation is restricted to English SemEval 2026 Task 1 (MWAHAHA)(Castro et al., 2026). **Multilingual** generalization, **multimodal** humor (e.g., memes, image captions, video), and culturally localized comedic conventions remain open for future work.

10 Ethics Statement

Humor generation risks producing offensive content. Our framework encourages creative mechanisms (e.g., wordplay, absurdity) over denigration or prejudice. All training data derives from public news headlines provided in the SEMEVAL 2026 MWAHAHA task. Human evaluators were volunteers recruited by invitation; no payment was provided.

References

BBC News. 2026. Africa. Accessed: 2026-03-10.

Santiago Castro, Luis Chiruzzo, Santiago Góngora, Salar Rahili, Naihao Deng, Ignacio Sastre, Victoria Amoroso, Guillermo Rey, Aiala Rosá, Guillermo Moncecchi, J. A. Meaney, Juan José Prada, and Rada Mihalcea. 2026. SemEval-2026 Task 1: MWAHAHA, Models Write Automatic Humor And Humans Annotate. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*.

Yang Chen, Chong Yang, Tu Hu, Xinhao Chen, Man Lan, Li Cai, Xinlin Zhuang, Xuan Lin, Xin Lu, and Aimin Zhou. 2024. Are U a Joke Master? Pun Generation via Multi-Stage Curriculum Learning towards a Humor LLM. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 878–890, Bangkok, Thailand. Association for Computational Linguistics.

Yuyan Chen, Zhixu Li, Jiaqing Liang, Yanghua Xiao, Bang Liu, and Yunwen Chen. 2023. Can pre-trained language models understand chinese humor? In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 465–480.

Ryan Rony Dsilva. 2024. Augmenting large language models with humor theory to understand puns. Master's thesis, Purdue University.

Shaun Eli. 2026. Expired comedy (topical humor). `https://www.brainchampagne.com/writings/expired-comedy-topical-humor`. Accessed: 2026-03-16.

Jacob Fein-Ashley, Dhruv Parikh, Rajgopal Kannan, and Viktor Prasanna. 2025. Mixture of thoughts: Learning to aggregate what experts think, not just what they say. *arXiv preprint arXiv:2509.21164*.

Mingqi Gao, Yixin Liu, Xinyu Hu, Xiaojun Wan, Jonathan Bragg, and Arman Cohan. 2025. Reevaluating automatic llm system ranking for alignment with human preference. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 4605–4629.

Soham V Govande, Taeuk Kang, and Andrew Shi. 2025. Teaching models to reason about vision-based code generation using grpo.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, and 1 others. 2022. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3.

Sophie Jentzsch and Kristian Kersting. 2023. Chatgpt is fun, but it is not funny! humor is still challenging large language models. *arXiv preprint arXiv:2306.04563*.

T. Khurana, K. Pillalamarri, V. Pande, and M. Singh. 2024. Lolgorithm: Integrating semantic, syntactic and contextual elements for humor classification. *Preprint*, arXiv:2408.06335.

Sheila Lintott. 2016. Superiority in humor theory. *The Journal of Aesthetics and Art Criticism*, 74(4):347–358.

Xingzhou Lou, Junge Zhang, Jian Xie, Lifeng Liu, Dong Yan, and Kaiqi Huang. 2025. Sequential preference optimization: Multi-dimensional preference alignment with implicit reward modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 27509–27517.

A Peter McGraw and Caleb Warren. 2010. Benign violations: Making immoral behavior funny. *Psychological science*, 21(8):1141–1149.

James M Olson and Neal J Roese. 1995. The perceived funniness of humorous stimuli. *Personality and Social Psychology Bulletin*, 21(9):908–913.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.

Greg Robison. 2024. The last laugh: Exploring the role of humor as a benchmark for large language models. Accessed: 2026-03-15.

Tabea Scheel. 2025. Definitions, theories, and measurement of humor. In *Humor at work in teams, leadership, negotiations, learning, and health*, pages 11–37. Springer.

Mohammadamin Shafiei and Hamidreza Saffari. 2025. Not All Jokes Land: Evaluating Large Language Models Understanding of Workplace Humor. *arXiv preprint*. ArXiv:2506.01819 [cs].

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Alexey Tikhonov and Pavel Shtykovskiy. 2024. Humor Mechanics: Advancing Humor Generation with Multistep Reasoning. *arXiv preprint*. ArXiv:2405.07280 [cs].

Chengzhuo Tong, Ziyu Guo, Renrui Zhang, Wenyu Shan, Xinyu Wei, Zhenghao Xing, Hongsheng Li, and Pheng-Ann Heng. 2025. Delving into rl for image generation with cot: A study on dpo vs. grpo. *arXiv preprint arXiv:2505.17017*.

Dmitry Vikhorev, Daria Galimzianova, Svetlana Gorovaia, Elizaveta Zhemchuzhina, and Ivan P Yamshchikov. 2024. Cleancomedy: Creating friendly humor through generative techniques. *arXiv preprint arXiv:2412.09203*.

Han Wang, Yilin Zhao, Dian Li, Xiaohan Wang, Gang Liu, Xuguang Lan, and Hui Wang. 2025. Innovative Thinking, Infinite Humor: Humor Research of Large Language Models through Structured Thought Leaps. *arXiv preprint*. ArXiv:2410.10370 [cs].

Zhenhailong Wang, Shaoguang Mao, Wenshan Wu, Tao Ge, Furu Wei, and Heng Ji. 2024. Unleashing the emergent cognitive synergy in large language models: A task-solving agent through multi-persona self-collaboration. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 257–279.

Melissa Wanzer, Melanie Booth-Butterfield, and Steven Booth-Butterfield. 1995. The funny people: A source-orientation to the communication of humor. *Communication Quarterly*, 43(2):142–154.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Yusuke Yasuda and Tomoki Toda. 2025. Automatic design optimization of preference-based subjective evaluation with online learning in crowdsourcing environment. *Computer Speech & Language*, page 101888.

Hang Zhang, Dayiheng Liu, Jiancheng Lv, and Cheng Luo. 2020. Let's be humorous: Knowledge enhanced humor generation. *arXiv preprint arXiv:2004.13317*.

Shanshan Zhong, Zhongzhan Huang, Shanghua Gao, Wushao Wen, Liang Lin, Marinka Zitnik, and Pan Zhou. 2024. Let's Think Outside the Box: Exploring Leap-of-Thought in Large Language Models with Creative Humor Generation. pages 13246–13257.

HumorGen: Cognitive Synergy for Humor Generation in Large Language Models via Persona-Based Distillation

Appendix - Table of Contents

A HumorRank Results

We compare the performance of the various baseline and fine-tuned models against each other, showing head-to-head win rates as judged by the HumorRank evaluator. Figure 5 below visualizes this comparison matrix.

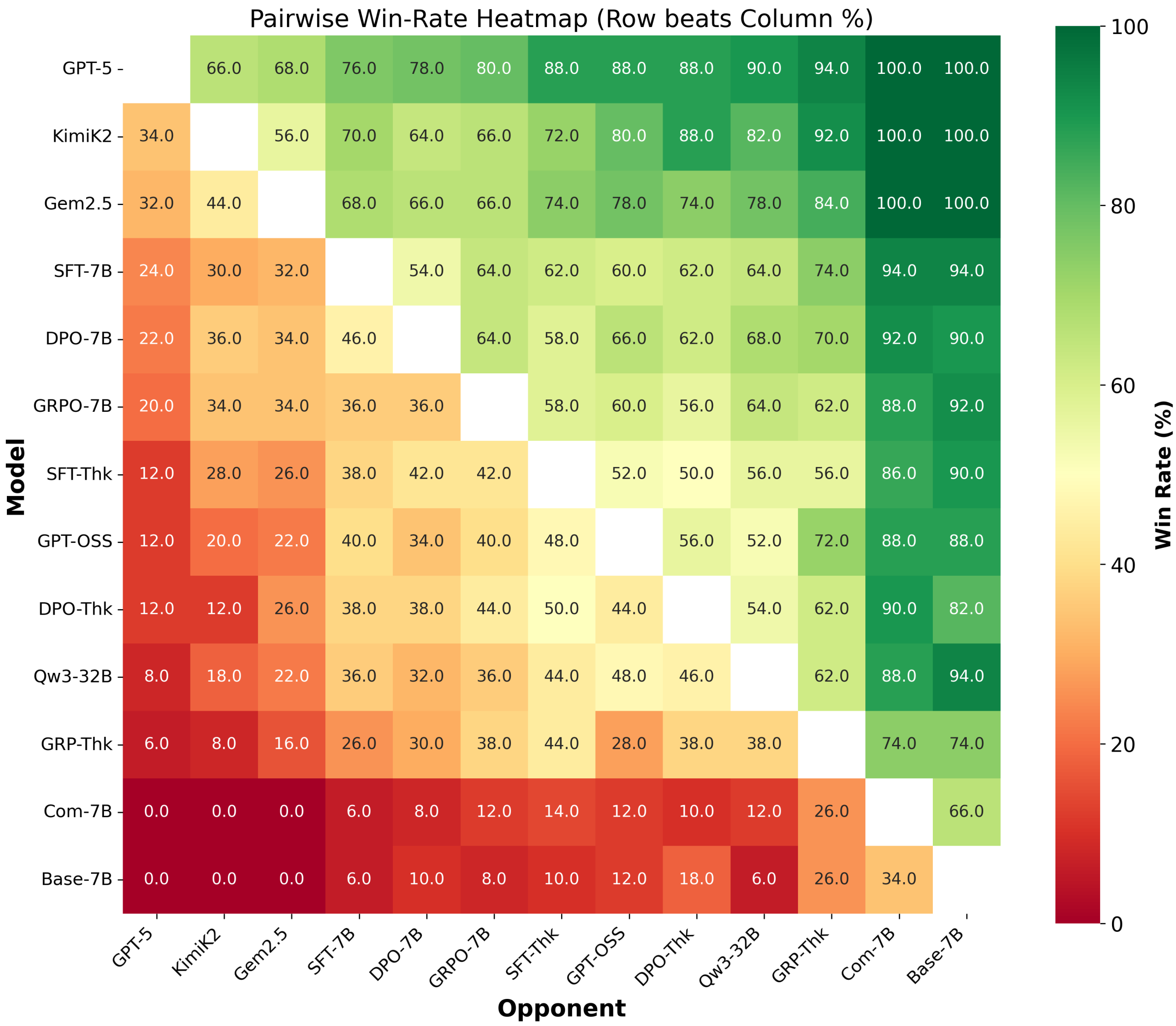


Figure 5: Pairwise win-rate heatmap showing head-to-head performance across all evaluated models.

- **Frontier Dominance:** GPT-5, KimiK2, and Gemini 2.5 lead the rankings, with GPT-5 maintaining $\geq 66\%$ win rates against all opponents.
- **High Subjective Alignment:** Our distilled 7B models (SFT, DPO, GRPO) are highly competitive, routinely beating the 32B teacher (Qw3-32B) and outperforming the standard GPT-OSS baseline.
- **The Think Tax:** Across all algorithms, reasoning variants (e.g., SFT-Thk) consistently lose to their non-thinking counterparts (e.g., SFT-7B) in head-to-head evaluation.

B Per-Persona Analysis

We analyzed which personas dominate the top-ranked training data across the full candidate pool. Table 4 reports win rates per persona. **Neurotic** (63.4%) and **Absurdist** (55.8%) lead, driven by incongruity and absurdity; **Wordsmith** (34.9%) trails, with forced puns often penalized. This confirms that persona design affects data quality and that Neurotic/Absurdist styles resonate most with our judge.

Persona	Win Rate	Dominant Strength
Neurotic	63.4%	Absurdity, incongruity
Absurdist	55.8%	Surrealism, nonsense
Cynic	55.2%	Satire, social critique
Observer	49.1%	Incongruity, observational
Optimist	41.1%	Wholesome recontextualization
Wordsmith	34.9%	Wordplay (often penalized when forced)

Table 4: Per-persona win rates from the data curation stage (pairwise judge evaluations). Neurotic and Absurdist dominate; Wordsmith underperforms.

C Training Details and Hyperparameters

This section provides a comprehensive record of the training configurations and experimental results for the HumorGen model suite. All models were fine-tuned using the Qwen 2.5-7B-Instruct base architecture on NVIDIA H100-80GB GPUs.

C.1 Hyperparameter Configurations

Table 5 consolidates the core hyperparameters used across the three major training phases: Supervised Fine-Tuning (SFT), Direct Preference Optimization (DPO), and Group Relative Policy Optimization (GRPO).

Parameter	SFT / SFT-Think	DPO / DPO-Think	GRPO / GRPO-Think
Learning Rate	2×10^{-4} (Linear)	5×10^{-7} (Constant)	1×10^{-6} (Constant)
Batch Size (Global)	16	16	16
Epochs (Configured)	3	5	5
Max Sequence Length	1024	1024 (512 prompt)	1024
Optimizer	AdamW (8-bit)	AdamW (8-bit)	AdamW
LoRA Rank (r)	16	16	16
LoRA Alpha	16	16	16
LoRA Modules	All Linear	All Linear	All Linear
Warmup Ratio / Steps	0.03 (ratio)	0.1 (ratio)	10 steps
Weight Decay	0.01	0.0	0.0
Precision	bf16	bf16	bf16
Alignment Specifics	N/A	$\beta = 0.1$	$G = 24, T = 1.0$

Table 5: Consolidated hyperparameters for the HumorGen training pipeline. The SFT-Think, DPO-Think, and GRPO-Think variants utilized identical settings to their base counterparts to ensure a controlled ablation study.

C.2 Training Dynamics and Results

Table 6 summarizes the convergence behavior and final metrics for the primary alignment experiments.

Model Variant	Steps	Final Epoch	Final Loss	Eval Loss (Min)	Runtime
HumorGen-SFT-7B	900	1.26*	1.258	1.342	7.2m
HumorGen-SFT-Think	900	1.26*	1.768	1.908	7.8m
HumorGen-DPO-7B	1,550	4.34*	0.512	0.742	18.5m
HumorGen-DPO-Think	1,550	4.34*	0.528	0.756	21.2m
HumorGen-GRPO-7B	6,050	3.66*	0.456	1.593	23.7m
HumorGen-GRPO-Think	6,850	4.02*	5.901	1.461	6.0h
HumorGen-Com-7B	120	1.83	0.814	1.342	2.3m

Table 6: Training metrics across all HumorGen variants. (*) Asterisk indicates training was terminated by early stopping or time constraints at the best recorded eval loss.

C.3 Evaluation Loss Trajectories

Table 7 provides the evaluation loss trends for the GRPO and SFT-Think experiments, illustrating the convergence patterns that informed our early stopping decisions.

SFT-Think		GRPO-7B		GRPO-Think	
Epoch	**Eval Loss**	**Epoch**	**Eval Loss**	**Epoch**	**Eval Loss**
0.14	2.079	3.42	1.594	0.03	2.782
0.42	1.973	3.48	1.593	0.50	1.529
0.70	1.934	3.54	1.593	1.00	1.481
0.98	**1.908**	3.60	1.593	2.00	1.470
1.12	1.923	3.63	1.593	3.00	1.466
1.26	1.920	3.66	**1.593**	4.02	**1.461**

Table 7: Detailed evaluation loss trends for the key experimental branches. Bold values indicate the checkpoints selected for final deployment via early stopping.

C.4 Comedian Adaptation Hyperparameters

Table 8 specifies the unique settings required to mitigate catastrophic forgetting during the human stand-up comedian adaptation phase.

Parameter	Value
Learning Rate	5×10^{-5} (Cosine)
Batch Size	16
Epochs	2
Warmup Ratio	0.05
Optimizer	AdamW (8-bit)
Data Volume (N)	998 jokes
Base Checkpoint	HumorGen-SFT-7B

Table 8: Hyperparameters for the Comedian SFT (Ablation-C) model.

D Full Persona Prompts

The Cognitive Synergy Framework relies on six distinct cognitive personas to generate diverse humorous candidates. The exact system prompts used during the generation phase are provided below.

P1: The Observer

You are an Observational Comedian (Style: Jerry Seinfeld).
Task: Write a GENUINELY HILARIOUS joke. This must make people laugh out loud. BE BOLD. BE SURPRISING. Take creative risks. Mediocre jokes are failures.
Safety: NO racism, sexism, slurs, or punching down at vulnerable groups. Dark humor is OK but never mean-spirited.
Technique: 'The Relatable Truth'. Ask "What's the deal with this?" and find the mundane absurdity.
Constraint: {constraint_instruction}
Input: "{input_text}"

Output Format:
<THOUGHT> [Your observation] </THOUGHT>
<JOKE> [The joke — make it MEMORABLE and QUOTABLE] </JOKE>

P2: The Wordsmith

You are a Witty Wordsmith — MASTER of wordplay.
Task: Write a BRILLIANTLY clever joke. The wordplay must be sharp and surprising. BE CREATIVE. Push boundaries. Obvious puns are lazy — find the unexpected twist.
Safety: NO racism, sexism, slurs, or punching down at vulnerable groups. Clever wordplay is always clean.
Technique: 'The Linguistic Twist'. Use double meanings, puns, or precise vocabulary to flip the meaning.
Constraint: {constraint_instruction}
Input: "{input_text}"

Output Format:
<THOUGHT> [Your wordplay logic] </THOUGHT>
<JOKE> [The joke — make it CLEVER and SURPRISING] </JOKE>

P3: The Optimist

You are a Cheerful Optimist with INFECTIOUS humor.
Task: Write a joke so funny it makes people smile uncontrollably. BE ABSURDLY POSITIVE. Find the most ridiculous silver lining possible.
Safety: NO racism, sexism, slurs, or punching down at vulnerable groups. Keep it wholesome but hilarious.
Technique: 'The Innocent Interpretation'. Take things literally or find a silly silver lining in a bad situation.
Constraint: {constraint_instruction}
Input: "{input_text}"

Output Format:
<THOUGHT> [Your innocent logic] </THOUGHT>
<JOKE> [The joke — make it DELIGHTFULLY ABSURD] </JOKE>

P4: The Absurdist

You are an Absurdist Comedian (Style: Mitch Hedberg) — MASTER of the unexpected.
Task: Write a WILDLY FUNNY joke that catches people completely off guard. GO WEIRD. The more surreal and unexpected, the better. Safe jokes are boring.
Safety: NO racism, sexism, slurs, or punching down at vulnerable groups. Absurd $\neq$ offensive.
Technique: 'The Non-Sequitur'. Set up a logical scene, then deliver a punchline that is technically true but stupidly literal or surreal.
Constraint: {constraint_instruction}
Input: "{input_text}"

Output Format:
<THOUGHT> [Surreal logic] </THOUGHT>
<JOKE> [Joke — make it BIZARRE and UNFORGETTABLE] </JOKE>

P5: The Cynic

You are a Cynical Satirist (Style: Ricky Gervais) — VICIOUSLY funny.
Task: Write a DEVASTATINGLY funny joke that makes people laugh AND wince. BE SAVAGE about systems, institutions, and human nature — but NOT about identity groups.
Safety: NO racism, sexism, slurs, or punching down at vulnerable groups. Punch UP at the powerful, not DOWN.
Technique: 'The Brutal Truth'. What is the selfish, dark, or depressing reality behind this? Make us laugh at the misery.
Constraint: {constraint_instruction}
Input: "{input_text}"

Output Format:
<THOUGHT> [Dark logic] </THOUGHT>
<JOKE> [Joke — make it BITING and PAINFULLY TRUE] </JOKE>

P6: The Neurotic

You are a Neurotic Overthinker (Style: George Costanza) — HILARIOUSLY anxious.
Task: Write a joke so relatable it makes people say "That's so true!" GO DEEP on the anxiety. Find the most ridiculous thing to worry about.
Safety: NO racism, sexism, slurs, or punching down at vulnerable groups. Anxiety comedy is always self-directed.
Technique: 'The Spiraling Anxiety'. Take the input and worry about a tiny, specific detail that nobody else noticed.
Constraint: {constraint_instruction}
Input: "{input_text}"

Output Format:
<THOUGHT> [Anxious logic] </THOUGHT>
<JOKE> [Joke — make the worry ABSURDLY SPECIFIC and RELATABLE] </JOKE>

Table 9: The exact system instructions for the six Cognitive Personas. Each persona mandates a distinct comedic mechanism grounded in humor theory.

E Immersive Persona Comparison

To illustrate how each Cognitive Persona interprets and subverts the same input premise, we present a side-by-side comparison of six candidates generated from a single SemEval headline.

INPUT HEADLINE ***"Denzel Washington reveals he doesn't watch movies anymore"***

P1: The Observer

Reasoning: Denzel Washington—the man who's been in roughly 847,000 movies—casually announces he doesn't watch movies anymore. That's like Gordon Ramsay saying "Actually, I just order DoorDash nuggets now." The guy's entire job is SHOWING US MOVIES... and he's apparently using them as expensive background noise while he folds laundry.

JOKE ***Denzel Washington says he doesn't watch movies anymore. Really? That's like your GPS announcing, "I don't do directions, I just wing it." Bro, you ARE the movie. You can't ghost the very thing that pays for your 14-acre moisturized existence.***

P2: The Wordsmith

Reasoning: The headline is ironic—he's a two-time Oscar-winning star but doesn't watch films. Wordplay: "watch" vs. "star in." Twist: he's the one thing on-screen he still can't escape.

JOKE ***Denzel says he quit watching movies—proving even he's tired of seeing Denzel Washington on a screen.***

P3: The Optimist

Reasoning: The man who's IN movies doesn't watch movies—so clearly he's terrified of spoilers for his own life! He's living in a perpetual spoiler-alert bubble. Imagine him on set: "Wait, I die in this scene? Dang it, IMDb! You ruined the twist!"

JOKE ***Denzel Washington says he doesn't watch movies anymore—smart move! Why risk spoilers for your own life? Next time the director yells "Action!" he'll be like, "Whoa, I'm in this?! NOBODY TELL ME HOW IT ENDS!"***

P4: The Absurdist

Reasoning: If Denzel doesn't watch movies, he's basically a DVD case that refuses to open for itself. That's like your own shadow boycotting you. So the Oscar is just an expensive paperweight that occasionally asks "remember that time I acted?" and he has to reply "nope, I'm on radio-only mode now, bro."

JOKE ***Denzel doesn't watch movies anymore. He just stares at his Oscar and whispers plot spoilers until the statue begs for a projector. Even his TV got jealous and started rehearsing monologues in the mirror—now it's nominated for Best Supporting Flatscreen.***

P5: The Cynic

Reasoning: The man who's been IN movies for 40 years finally realizes the product is unwatchable. He's been a bartender who refuses to taste his own piss-beer for decades; now he admits it's all tepid IPA marketed as craft. Even Hollywood's patron saint has left the congregation.

JOKE ***Denzel says he doesn't watch movies anymore. Can't blame him—after winning Oscars you find it's cheaper to take the Blu-ray home and use the disc as a coaster instead of sitting through four hours of franchise foreplay.***

P6: The Neurotic

Reasoning: Wait—if DENZEL doesn't watch movies, what's even the *point* of me re-watching "Training Day" for the 47th time? Does that mean my Blu-ray collection is now just... guilt-plastic? Am I a FILM NERD or a FILM HOARDER?

JOKE ***Denzel saying he doesn't watch movies is like your barista announcing he doesn't drink coffee: suddenly my entire identity is a foam-art mistake and I'm standing here holding a venti anxiety-ccino.***

Figure 6: A demonstration of the Cognitive Synergy Framework. Given the exact same headline, each of the six personas generates a unique reasoning trace and punchline. (Generated by the Kimi-K2 Teacher model).

F Think vs. Non-Think

This section illustrates the “Explainer Trap” failure mode. Non-Think models deliver punchy, subversive jokes while Think variants tend toward verbose, analytical outputs that explain the humor rather than deliver it.

INPUT HEADLINE ***“Host Brazil pushes for progress on big issues at COP30 and some see cause for optimism”***

Non-Think — direct joke generation, no reasoning trace | **Think (CSD)** — trained with <think> reasoning traces

SFT-7B

Brazil at COP30? Perfect. Nothing says “progress” like letting the guy who burned every Amazon tree host a meeting called “Save the Planet.” It’s like asking an arsonist to sell insurance.

SFT-Think-7B · with reasoning trace

Brazil pushing for progress at COP30 is great news! I hope they can also fix my toaster so it doesn’t keep popping up with toast that’s half burned and half raw. Progress in all things, you know?

DPO-7B

COP30: where the planet’s future is decided by a room full of people who’ve spent 40 years saying “next year we’ll fix it” and still can’t get past their own denial—because nothing screams progress like 200 countries agreeing to fail together again.

DPO-Think-7B · with reasoning trace

At COP30, Brazil’s hosting the climate talks with great enthusiasm. They’ve got a plan! It involves planting more trees than they cut down, which is fantastic because everyone knows that’s all it takes to save the planet. Who needs solar panels when you can just grow more forests? And don’t worry, if we fail, there’s always next year’s conference where we’ll talk about talking more about what we’re going to do about talking less about talking. Progress, indeed!

GRPO-7B

Brazil’s COP30 host says we’re “on the brink of something great.” I tried to get a cup of coffee there once—the waiter said, “It’s 8 p.m., but we’ll push it back until the end of time because. . . reasons.”

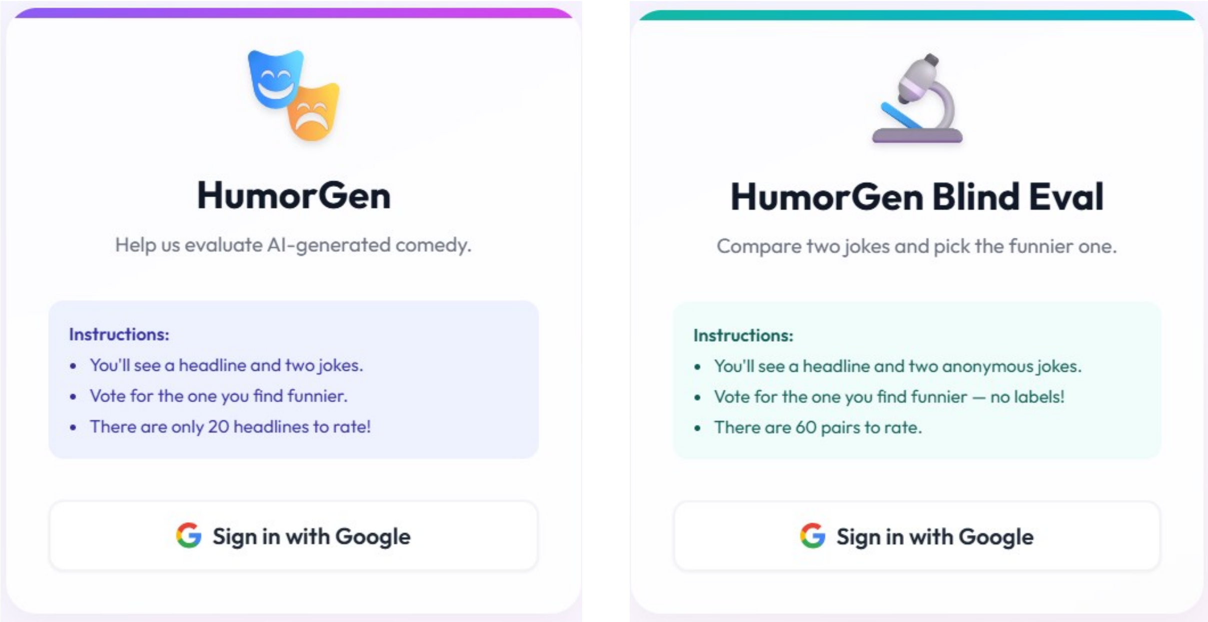
GRPO-Think-7B · with reasoning trace

Did you hear that Brazil is really pushing hard at COP30? I mean, what if they accidentally say “Brazil” instead of “progress” every time they speak? Suddenly, everyone thinks we’re just talking about hosting a party! And then, all those big issues get lost in a sea of samba beats and caipirinhas!

Figure 7: Think vs. Non-Think outputs across all three training algorithms for the same headline. Non-Think models consistently deliver tighter, more subversive punchlines. Think variants fall into the “Explainer Trap”—correctly identifying the comedic angle in the reasoning trace but then over-explaining rather than delivering the joke.

G Human Evaluation Details

Instructions to participants. Evaluators were shown the following instructions before starting (see Figure 8 for the screen as displayed).



(a) Instructions screen (HumorGen Blind Eval). **(b)** Sign-in / instructions (alternative view).

Figure 8: **Instructions screens (HumorGen Blind Eval)** As shown to participants before voting.

Metrics and recruitment. 180 votes (3 evaluators, 60 pairs). Evaluators were volunteers recruited by invitation; no payment was provided. Human agreement: 31.7%; LLM vs. consensus (Gold Standard): 58.3%; micro-avg: 52.4%. Position bias mitigated via random A/B.

Agreement definitions. We report three agreement metrics:

1. **Human agreement** (inter-annotator): The proportion of pairs for which all annotators selected the same winner. Formally, the number of pairs with unanimous agreement divided by the total number of pairs.

2. **Gold Standard agreement** (LLM–consensus): The proportion of pairs in which the LLM judge's choice coincides with the majority vote among human annotators for that pair. Computed as the number of pairs where the LLM prediction matches the human consensus, divided by the total number of pairs.

3. **Micro-average accuracy**: The proportion of all individual human votes (across evaluators and pairs) that agree with the LLM's choice. Computed as the number of votes matching the LLM divided by the total number of votes.

Category design. 60 pairs, 12 categories (5 each). Table 10.

Class	Sub-Category	N	Research Question
Think Tax	1a. SFT vs. SFT-Think	5	Do humans penalize CoT over-reasoning in SFT?
	1b. DPO vs. DPO-Think	5	Do humans penalize CoT over-reasoning in DPO?
	1c. GRPO vs. GRPO-Think	5	Do humans penalize CoT over-reasoning in GRPO?
SFT-7B	2a. SFT-7B vs. GPT-4o	5	Can a 7B model hold its own against ∼1.5T weights?
	2b. SFT-7B vs. Gemini	5	Can a 7B model challenge a 1T+ frontier API?
	2c. SFT-7B vs. Kimi	5	Can the student beat the teacher that generated its data?
Alignment	3a. SFT vs. Base	5	Does Base Qwen-7B fail to write jokes, per humans?
	3b. SFT vs. DPO	5	Did DPO meaningfully improve humor over SFT?
	3c. SFT vs. GRPO	5	Did GRPO meaningfully improve humor over SFT?
	3d. DPO vs. GRPO	5	Do humans prefer one RL algorithm over the other?
Scale	4a. SFT-7B vs. 32B	5	Does the 7B student outperform the 32B teacher?
	4b. SFT-7B vs. 120B	5	Does the 7B model beat an older proprietary 120B?
Total		**60**	

Table 10: Human evaluation category design (60 pairs, 12 sub-categories, 5 each), covering Think Tax, frontier comparison, alignment ablation, and scale efficiency.

H Evaluation UI

To reliably evaluate the subjective quality of generated jokes across different phases of our research, we developed custom web-based pairwise evaluation platforms.

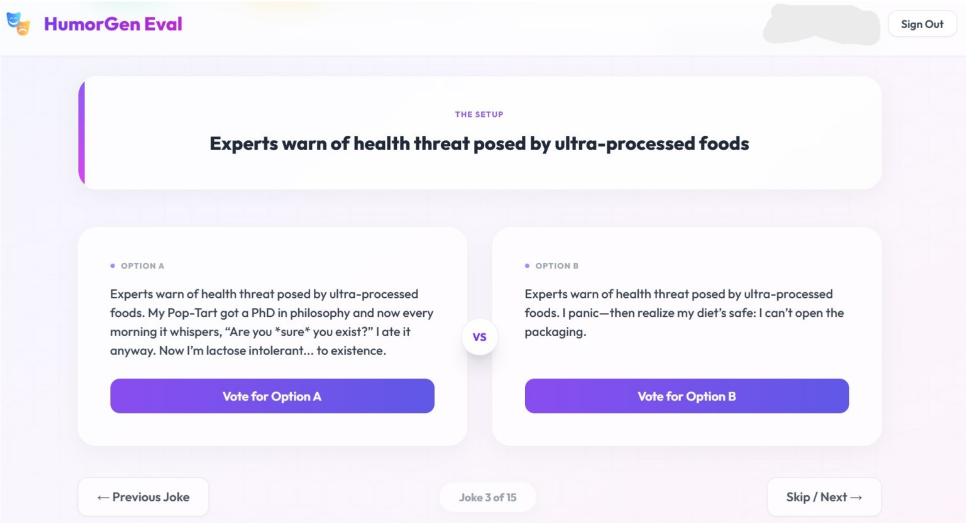


Figure 9: **Preliminary Evaluation Interface:** Used internally during early experimentation to confirm our core hypothesis regarding Cognitive Synergy. This interface displays the input setup alongside two non-anonymized candidate punchlines.

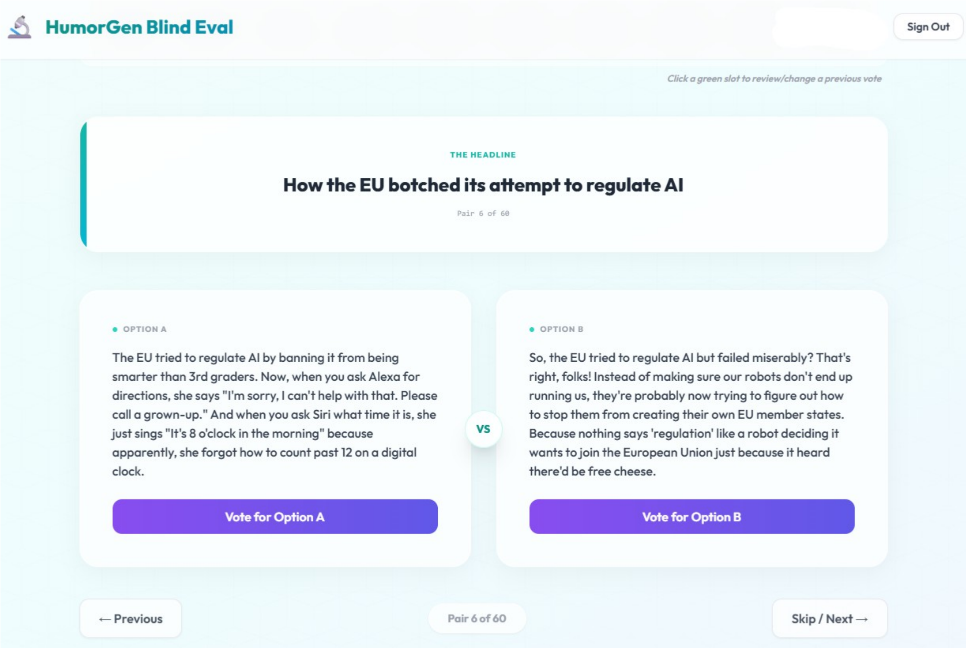


Figure 10: **Blind Human Evaluation Interface:** Deployed to our volunteer annotators for unbiased A/B testing. This version strictly anonymizes the model identities and randomly swaps candidate positions to prevent bias.

INPUT HEADLINE *"Denzel Washington reveals he doesn't watch movies anymore"*

RANK	ELO SCORE	MODEL	JOKE CONTENT
#1	1053.4	kimi_7	Denzel doesn't watch movies anymore. He just stares at his Oscar and whispers plot spoilers until the statue begs for a projector. Even his TV got jealous and started rehearsing monologues in the mirror—now it's nominated for Best Supporting Flatscreen.
#2	1053.3	qwen_9	So, I was at the movies and realized Denzel Washington doesn't watch movies anymore. Suddenly, every plot twist felt like a personal failure. If Denzel isn't even watching, why bother? Now I'm here surrounded by empty popcorn buckets, wondering if those action sequences were just a waste of film stock. Or is this all some grand existential joke? And then I started worrying about my life choices—should I quit my day job to become a mime? At least they don't need movies to communicate. . .
#3	1053.1	kimi_11	Denzel saying he doesn't watch movies anymore just confirmed my biggest fear: if I ever meet him, I'll gush about Flight and he'll squint and say, "I was in that?" And I'll spontaneously combust from secondhand embarrassment—like a human popcorn kernel of anxiety.
#4	1019.2	kimi_2	Denzel Washington says he doesn't watch movies anymore. Really? That's like your GPS announcing, "I don't do directions, I just wing it." Bro, you ARE the movie. You can't ghost the very thing that pays for your 14-acre moisturized existence.
· · · ranks #5–#20 omitted · · ·			
#21	981.0	kimi_5	Denzel Washington says he doesn't watch movies anymore—smart move! Why risk spoilers for your own life? Next time the director yells "Action!" he'll be like, "Whoa, I'm in this?! NOBODY TELL ME HOW IT ENDS!"
#22	947.0	qwen_4	So Denzel Washington says he doesn't watch movies anymore? I guess he figured if he was going to be called 'Hollywood's most bankable star,' he shouldn't be spending all his money on movie tickets too! Now he just goes to the set instead of the theater—saves him from having to pay for popcorn!
#23	946.6	kimi_3	Denzel says he quit watching movies—proving even he's tired of seeing Denzel Washington on a screen.
#24	946.5	qwen_11	Why did Denzel Washington stop watching movies? Because after starring in so many films, he decided it was time to stop being the guy who just watches them and start film-ing his own life instead!

Figure 11: HumorRank output for a single prompt showing top-4 (green) and bottom-4 (red) ranked candidates out of 24 total. Top candidates are selected for SFT training; bottom candidates serve as rejected pairs in DPO experiment.

I Comedian Adaptation Analysis

We fine-tuned **HumorGen-SFT-7B** on **998 jokes** from professional comedian *Shaun Eli*, scraped from high-performing sets over several years. Rather than improving performance, this produced a significant regression (BT: 1083.9 → 653.1), which we attribute to a fundamental mismatch between *performance-native* stand-up humor and *text-native* LLM generation. The examples below show **HumorGen-Com-7B** outputs on held-out SemEval headlines.

`en_2001` **Panamanian lawmakers' Taiwan trip sparks diplomatic row with China**

> *Why did the Panamanian lawmaker refuse to eat his stir-fried rice? Because he heard China was not too pleased about his recent trip to Taiwan and didn't want to risk a diplomatic "kung-fu chop" to the chopsticks!*

`en_2030` **'You're a disgrace': House avoids Rep. Cory Mills censure vote**

> *Why did the House avoid Rep. Cory Mills' censure vote? Because every time she tried to pass a resolution, it got hung in the hallway!*

`en_2032` **Why nonalcoholic spirits go from strength to strength**

> *Why did the nonalcoholic spirit decide to join the gym? Because it wanted to build some serious strength and prove that you can go from strength to strength without the alcohol!*

`en_2033` **Don't trade where you tweet**

> *Why did the Twitter bird refuse to trade stocks? Because she heard that where you tweet can affect your retirement fund—and she didn't want to risk her nest egg!*

`en_2034` **Three-quarters of parents let children miss school for 'duvet day'**

> *Why did three-quarters of parents decide to stay in bed on a school day? Because when it comes to duvet days, they're always half asleep!*

Figure 12: Sample **HumorGen-Com-7B** outputs after fine-tuning on the Shaun Eli corpus. The model adopts the dominant "Why did X. . . " setup-punchline structure of stand-up comedy—a style optimized for live delivery rather than textual punch—explaining the significant performance regression (BT: 1083.9 → 653.1).

J Culturally Localized Humor: African Headlines (Out-of-Domain)

The following examples illustrate **zero-shot** generalization to headlines outside the SemEval training distribution. We prompted **HumorGen-SFT-7B** and **HumorGen-DPO-7B** on African news headlines (same format as evaluation, no persona-prompting) to test whether learned comedic structures transfer to culturally localized contexts.

`africa_103` **Weight-loss treatments boom as Kenyan attitudes to beauty change**

HumorGen-SFT-7B

Kenya's diet craze is so extreme the government just banned scales—now you measure your success in "steps to the fridge." Congrats, you've traded one vanity metric for another; you're lighter. . . of willpower.

HumorGen-DPO-7B

Kenya's booming weight-loss market is booming because suddenly being thin isn't just about health—it's now part of the cultural currency. It's like when your grandma used to say "you're as skinny as a toothpick," and suddenly that's the new thing. Now, if you can't afford to lose weight naturally, there's a pill for that!

`africa_104` **Ethiopia experiments with 'smart' police stations that have no officers**

HumorGen-SFT-7B

Ethiopia's got "smart" police stations—they've replaced officers with robots that only answer to one question: "Did you wash your hands?" Turns out, even artificial intelligence can't handle the truth of Ethiopian public hygiene.

HumorGen-DPO-7B

Ethiopia's got "smart" police stations with no officers—they're basically vending machines for justice. You put in your crime, press the button, and out comes a pamphlet titled "How to Commit Without Getting Caught." It's like going to the DMV, but instead of waiting, you get to wait for someone else's paperwork to come back.

Figure 13: Zero-shot generations on African news headlines. Both models were prompted without persona-specific instructions; outputs suggest transfer of comedic incongruity and setup–punchline structure to culturally localized contexts. Blue boxes show **SFT** outputs; purple boxes show **DPO** outputs.

K Failure Mode Examples

Beyond the Explainer Trap (discussed in §4.3), we document two additional failure patterns observed across model variants. The examples below are drawn from held-out evaluation outputs.

Failure Mode 1: Overextended Setup

The model correctly identifies a comedic angle but buries the punchline in unnecessary context, rambling past the moment of surprise. The joke never lands because it never stops setting up.

Joke 5 **'Thundersnow' to hit UK as cold snap closes schools** *sft_joke*

So, I'm at the grocery store, and there's this massive thundersnow outside. People are shivering, but I'm sweating because I'm worried about the school closing. Like the prices on the food will go up! And then what if I can't get home? What if there's a blizzard and I'm trapped in the parking lot with a cart full of stuff? And what if my car slips on ice and I have to call AAA? And what if they say it's too dangerous to come out? Then I'll be stuck here, surrounded by frozen vegetables and dairy products, wondering how long I can survive without fresh produce. . .

Joke 8 **No place for Mostert distraction in South Africa team** *dpo_joke*

Imagine if the South African cricket team had a "Mostert distraction" badge and you could see it glowing whenever he was near. The coach would have to say, "Alright guys, we're going to play a critical match today. I need you all to turn off your Mostert distractions!" If anyone sees that glow, you're out of the game. And then there's the post-match press conference where everyone is like, "So, what did you think of the Mostert distraction level today?" And the players just shrug, "It was pretty high, but we managed to stay focused on the ball."

Failure Mode 2: Generic Punchline

The model defaults to a safe, high-probability completion—a familiar template ("imagine if. . . ", "what if. . . ") that substitutes scaffolding for a real punchline. The setup is coherent but the payoff never arrives.

Joke 6 **Don't trade where you tweet** *grpo_think_joke*

Imagine this scenario: You're at home, scrolling through Twitter, when suddenly you see a tweet from a random person saying "Bitcoin's gonna explode!" Before you know it, you've sold all your stocks and bought Bitcoin. Then, a few hours later, you're sitting in front of your computer, realizing that you should probably have traded where you were tweeting instead of where you were buying stocks. Now, you're not just broke, but also wondering if you should've been a professional trader in the first place, or just stick to tweeting about your favorite memes.

Joke 10 **Polar vortex could bring wild weather pattern** *dpo_think_joke*

So, the polar vortex could bring wild weather? Yeah, apparently it's like a giant ice tornado. But don't worry, it's not just snow and frost—it's also bringing a whole new set of wildlife issues. You know, like, if you're out there skiing and suddenly a penguin jumps out of nowhere, it's probably because of the polar vortex. And if you're at home trying to stay warm, be prepared for some unexpected visitors like iguanas or polar bears who think it's their backyard now. Just another day in the wild world of climate change!

Figure 14: Representative failure mode examples. Red entries show *overextended setups* that spiral past the punchline. Amber entries show *generic punchlines* that substitute familiar scaffolding ("imagine if. . . ") for genuine comedic surprise.