A Training-Free Novelty Detection Framework for Large-Scale Multi-Agent Social Creativity Simulations

Luuk Motz and **Tessa Verhoef** and **Rob Saunders**
Leiden Institute of Advanced Computer Science (LIACS), Leiden University
luukmotz@gmail.com t.verhoef@liacs.leidenuniv.nl r.saunders@liacs.leidenuniv.nl

Abstract

Per-agent novelty detection is central to computational models of social creativity, but trained detectors have kept simulations too small to study population-level dynamics such as the emergence of sub-fields or the relationship between per-share reception and cumulative field-level contribution. We present a training-free framework in which agents maintain a personal repository of pretrained ResNet-18 features, with novelty computed via k-nearest-neighbour distances normalised by local standard deviation. The framework scales to significantly larger agent populations while preserving individual aesthetic variance on a single consumer GPU. We validate at 50 to 250 agents, surfacing a rate–volume inversion pattern where the most prolific senders achieve the lowest per-share acceptance rates, with sub-linear per-step wall-clock cost in population size.

Introduction

Creativity is better understood as a social and cultural phenomenon than as a property of individual minds (Csikszentmihalyi 1988), and multi-agent simulation provides a natural tool for studying it as such (Saunders and Bown 2015). In these models, novelty detection is frequently a core mechanism: agents generate artefacts, judge them against their own experience, and decide what to share with the community (Saunders and Gero 2001). However, per-agent novelty detectors based on trained models impose computational costs that have kept simulations too small to study population-level phenomena such as the emergence of subfields or the dynamics of aesthetic divergence.

We propose a training-free framework that removes this bottleneck. Each agent maintains a personal repository of features extracted by a shared pretrained ResNet-18, and novelty is computed via k-nearest-neighbour (kNN) distances normalised by local standard deviation. Because the feature extractor is fixed and the kNN distance computation parallelises on GPU, the method preserves per-agent subjectivity at significantly larger population sizes, running on modest hardware. We validate the framework at populations from 50 to 250 agents, surfacing a population-level pattern where the most prolific senders achieve the lowest per-share acceptance rates, with sub-linear per-step wall-clock cost in population size.

Related Work

Saunders and Gero (2001) introduced a computational model of social creativity based on Csikszentmihalyi's (1988) Domain-Individual-Field-Interaction (DIFI) framework; curiosity-driven agents generate, evaluate, and share novel artefacts, producing emergent social structures such as cliques and stylistic divergence consistent with Martindale's (1990) theory of aesthetic evolution. Linkola, Takala, and Toivonen (2016) extended this early work by adding resource constraints and collective voting over domain entry; Guckelsberger et al. (2016) replaced novelty-seeking with coupled empowerment maximisation to account for both supportive and antagonistic agent relationships. Peeperkorn et al. (2022) replaced hand-crafted feature spaces with Variational Autoencoders, enabling conceptual spaces that adapt through social experience.

Saunders and Bown (2015) situated this approach within the agent-based modelling traditions of Axelrod (1997) and Epstein and Axtell (1996). Within this tradition, Gabora's Meme and Variations (1995) and its successor EVOC (2008) modelled culture as an evolutionary process using simpler but larger-scale agent societies up to 1,024 agents—each comprising an 18-node auto-associative neural network—demonstrating that an intermediate invention-to-imitation ratio maximises cultural fitness and that self-regulation of creativity has measurable population-level effects. Sosa and Gero (2005) modelled 100-agent design communities using Axelrod's cultural diffusion framework, explicitly encoding designers, adopters, and emergent gatekeepers—operationalising Csikszentmihalyi's field as a structurally distinct evaluative class. Sosa, Gero, and Jennings (2009) went on to show that the social worth of ideas grows and decays dynamically through adoption and displacement. Corneli (2016) reframed creative communities through Ostrom's (1990) institutional analysis, treating the shared domain as a creative commons.

Jennings (2010) went on to argue that creative autonomy—the capacity to revise one's own evaluative standards without external instruction—is best achieved through social embedding in a community of creators and critics, and demonstrated that creative search trajectories are consistent with simulated annealing dynamics (2011). Saunders (2012) proposed language games as the mechanism through which agents can autonomously develop shared creative vocabu-

laries and evaluative standards. Building on the language games framework of Steels (1995), Saunders (2011) demonstrated that agents can successfully develop these vocabularies, while Zhang and Saunders (2012) showed that compositional languages with hedonic evaluation promote deeper conceptual exploration. This connects to a broader scientific programme linking language, cultural transmission, and creativity: iterated learning experiments and models have established that languages evolve toward learnability and combinatorial structure (Kirby, Cornish, and Smith 2008; Verhoef, Kirby, and de Boer 2014; 2016), a cumulative cultural process that models of social creativity should aspire to capture. The development of Large Language Models (LLMs) has given rise to LLM-based agentic systems and has sparked a renewed interest in multi-agent social creativity at a new level of agent richness. Lin et al. (2025) survey the growing body of LLM-based multi-agent creative systems; Park et al. (2023) demonstrated that 25 LLM-based agents could produce emergent social behaviours from persona-based prompts; Imasato et al. (2025) applied this to Csikszentmihalyi's systems model, finding that multi-agent settings produce more creative artefacts than isolated agents; and Vallinder and Hughes (2025) showed that LLM-based agent societies can evolve cooperative norms across generations. Agent sophistication and population scale have so far moved in opposite directions: the largest simulations use the simplest agents and forgo core mechanisms such as per-agent novelty detection. The framework described in the next section is designed to break this tradeoff.

Method

Our approach replaces per-agent feature extraction and training with a shared pretrained ResNet-18 and a batched kNN process for novelty detection. Bergman, Cohen, and Hoshen (2020) showed that this approach can be effective for anomaly detection. We extend it here to address the individuality-scale tradeoff in social creativity simulations. Shared feature extraction amortises across all agents to a single batched forward pass per step, and per-agent novelty estimation reduces to a kNN lookup against a size-limited per-agent repository. All experiments were conducted on an NVIDIA RTX 3090 GPU and AMD Ryzen 5 5600X CPU.

Artefact generation is based on the process described for *The Digital Clockwork Muse* (Saunders and Gero 2001); agents generate 32×32 RGB images by evaluating quaternion-based expression trees at every pixel location (Witbrock and Reilly 1999). New expressions are evolved by subtree crossover and mutation; parent trees are chosen with equal probability from the agent's current expression or a personal repository of previously evaluated trees; a random subtree is selected for crossover, followed by a 5% point mutation. See Figure 4 for examples and expressions.

Feature Extraction Agents extract features using a ResNet-18 pretrained on ImageNet (He et al. 2016); to prevent overly aggressive spatial downsampling of the 32×32 images, the original 7×7 stride-2 first convolution is replaced by a 3×3 stride-1 layer initialised by bilinearly downsampling the pretrained kernel. 128-dimensional feature vectors are extracted from Layer 2 after global average pooling and ℓ_2 normalisation.

Extractor	AUROC	Noise
Pretrained L_2	0.827 ± 0.004	0.498
Pretrained L_3	0.821 ± 0.004	0.496
Pretrained L_4	0.809 ± 0.004	0.501
Random-init L_2	0.769 ± 0.005	0.500
Raw pixels	0.674 ± 0.004	0.498
Pretrained L_2 PCA-16	0.828 ± 0.003	0.495

Table 1: AUROC for kin vs. unrelated pair classification (mean $\pm$ std, 50 runs). The Noise column reports AUROC on uniform random image pairs.

The choice of Layer 2 was motivated by the observation that early CNN layers capture local texture, whereas deeper layers prioritise global semantics (Gatys, Ecker, and Bethge 2015; Geirhos et al. 2018). We observed that Layers 3 and 4 tend to map visually distinct artefacts onto coarse category-level representations, compressing the variance that novelty detection depends on, while Layer 2 provides a more appropriate representation of stylistic variance. We have validated this choice by classifying pairs of generated artefacts as kin (single-operator subtree substitution) or unrelated (independent random trees) via cosine distance in feature space, scoring by AUROC over 50 runs with 5,000 pairs per condition. Table 1 reports the result for six separately tested extractors and confirms that pretrained Layer 2 achieves the highest AUROC (0.827 ± 0.004), with Layers 3 and 4 trailing at 0.821 and 0.809 despite requiring two to six times more PCA components than Layer 2 to reach the 90% variance threshold. A random-initialisation baseline lags by 0.058, isolating the contribution of ImageNet pretraining, and a uniform-noise control returns AUROC near 0.5 across all extractors. A PCA-16 projection of Layer 2 matches the full-width AUROC (0.828 ± 0.003), suggesting that the per-agent embedding could be compressed eightfold without loss; however, we retain the full 128-dimensional representation for the experiments reported below.

Novelty Measurement Each agent i maintains a personal repository M_i implemented as a circular buffer of fixed capacity (1,000 entries) to model a recency bias and keep memory requirements bounded as simulations scale. Feature vectors for all generated and evaluated artefacts are added to the personal repository. When agent i encounters a feature vector f, it retrieves the k nearest neighbours in M_i to compute a novelty score:

$$\nu_i(f) = \frac{\bar{d}_k(f, M_i)}{\sigma_k(f, M_i)}$$

where $\bar{d}_k$ and σ_k are the mean and standard deviation of the k retrieved cosine distances on ℓ_2-normalised features, and $k = 15$ is fixed across all agents and steps. The denominator rescales each agent's signal by the local dispersion of its retrieved neighbourhood, yielding a unitless score that is comparable across agents regardless of memory size or

feature-space region. Novelty scores are normalised to $[0, 1]$ via a population-wide rolling percentile transform (1st and 99th percentiles over most recent 10,000 novelty scores), yielding $\tilde{\nu}_i(f)$, which is passed to a per-agent hedonic function, i.e., a Wundt curve (Berlyne 1971), to produce an interest score H_i:

$$H_i(\tilde{\nu}) = R_i(\tilde{\nu}) - \alpha P_i(\tilde{\nu})$$

where R_i and P_i are cumulative Gaussians encoding reward and punishment for novelty, and $\alpha = 1.2$ is used in the following simulations to weight the punishment term.

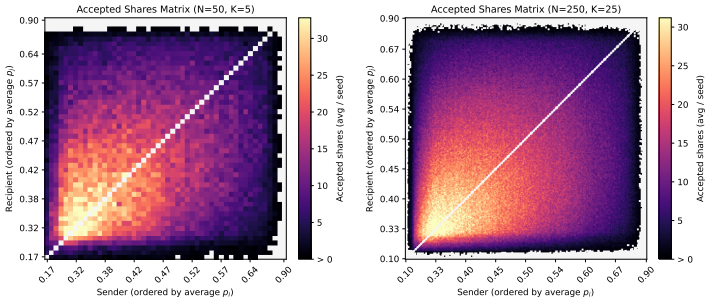
Each agent's Wundt curve, based on (Saunders and Gero 2001), is centred on a preferred novelty level $p_i \sim \mathcal{N}(0.5, 0.15)$, which distributes aesthetic preferences continuously across the population instead of assigning discrete roles. Its reward and punishment means are derived from p_i as $\mu_R = \max(0.1, p_i - 0.2)$ and $\mu_P = \min(0.9, p_i + 0.2)$, with fixed standard deviations $\sigma_R = \sigma_P = 0.15$; the ± 0.2 offset separates the two activations and the $\max$ / $\min$ clipping keeps both midpoints within the novelty range.

Social Interaction Policy Interest scores translate into social action through three global thresholds, namely a self-approval threshold τ_C (80th percentile), a domain threshold τ_D (80th percentile), and a boredom threshold τ_B (10th percentile), recomputed every step and applied population-wide. τ_C and τ_D are rolling percentiles over recent self-evaluations and recipient-side share evaluations respectively. τ_B is computed over the current distribution of per-agent average interest. After self-evaluating a generated artefact, agent i shares it with K other agents sampled randomly from the population, conditional on $H_i > \tau_C$. The K recipients each evaluate the artefact independently through their own kNN repository and Wundt curve. For each recipient agent j, if $H_j > \tau_D$, the artefact enters the shared domain and a directed edge $(i \to j)$ is recorded in the accepted-share graph; if H_j exceeds the recipient's interest in its own expression, agent j adopts the artefact as its current expression, independent of τ_D. When an agent's average interest falls below τ_B, it adopts a randomly sampled domain artefact, provided the agent's interest in that artefact exceeds its interest in its current expression. Before each step, agents re-evaluate their current expression against their updated repository, so that interest in unchanging material decays as similar artefacts accumulate in memory. The first 15 iterations are excluded from analysis while per-agent repositories and the rolling window populate.

Experimental Validation

We validate the framework across a $3 \times 3 \times 5$ grid of conditions: $N \in \{50, 100, 250\}$ agents, $K \in \{1, 5, 25\}$ shares per step, across five runs per (N, K) condition, with 2,500 iterations per run. Figures anchor on the $(N{=}50, K{=}5)$ and $(N{=}250, K{=}25)$ conditions, where both parameters are quintupled to test scaling. Enlarged figures and $N = 100$ results appear in the Appendix.

Figure 1 renders the accepted-share multigraph as a sender $\times$ recipient matrix where each cell counts run-averaged accepted shares from sender i to recipient j, with



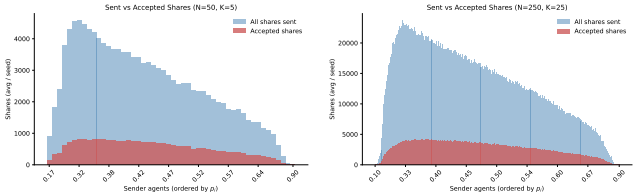
(a) $N = 50$, $K = 5$. (b) $N = 250$, $K = 25$.

Figure 1: Accepted-share interaction matrices. Each cell counts run-averaged accepted shares from sender i (x-axis) to recipient j (y-axis), with both axes ordered by preferred novelty; a position on either axis refers to the agent occupying that rank in a given run. White cells indicate pairs with zero accepted shares; black cells indicate pairs with at least one accepted interaction averaged out over all runs.

both axes ordered by preferred novelty; axis tick labels report the mean p_i at each rank across runs. Interaction density concentrates in the low-novelty preference (p_i) region, peaking near $p_i \approx 0.33$ for both senders and recipients. Scaling N and K by a factor of five reproduces this distribution at higher resolution, confirming it as a stable structural behaviour across population size and sharing rate.

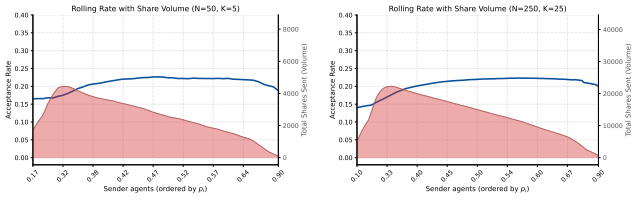
Figure 2 shows that share volume concentrates in the low-p_i region, with low-p_i senders emitting several times more shares than high-p_i senders and the total-sent distribution decaying monotonically toward the tail. The generator recombines the agent's current expression with one sampled from its personal memory, so the mean produced-novelty sits near 0.3–0.35 across runs, below the population-mean preference of 0.5. Agents whose Wundt peak aligns with this regime clear τ_C more often and compound their sending advantage over the run; high-p_i agents face the inverse. Low-p_i agents therefore dominate the accepted-share total through volume alone, and this behaviour replicates across all (N, K) runs.

Figure 3 shows that the rate–volume inversion is stable across both conditions. Low-p_i senders ship artefacts across a broad range that many recipients reject, depressing ac-



(a) $N = 50$, $K = 5$. (b) $N = 250$, $K = 25$.

Figure 2: Per-sender share volume ordered by preferred novelty p_i, run-averaged. Each bar pair reports total shares sent and the accepted subset; the accepted subset reflects the domain-threshold τ_D filter applied on the recipient side.



(a) $N = 50$, $K = 5$. (b) $N = 250$, $K = 25$.

Figure 3: Per-sender acceptance rate (rolling curve, left axis) and total shares sent (right axis) against preferred novelty p_i, run-averaged. The rate curve is inverted against volume: lowest where sending is densest, rising toward high-p_i where sending is sparsest.

ceptance rate; high-p_i senders clear τ_C only when generation lands near their preferred novelty, so the artefacts they do ship encounter less rejection. Per-share, high-p_i senders are thus more successful, but the multiplicative volume gap from Figure 2 dominates and absolute accepted-share counts remain concentrated in the low-p_i region.

Table 2 shows sub-linear scaling in N (log-log slopes 0.61–0.76), reflecting that the shared ResNet-18 and GPU-batched kNN amortise fixed overhead across the population. Per-step cost does not drift with repository growth for $K \geq 5$, as repositories are capped at 1,000 entries per agent.

Discussion

Figure 3 reveals a rate–volume inversion: high-preference senders achieve per-share acceptance rates several times those of low-preference senders and yet contribute a smaller fraction of accepted shares overall because volume dominates the total (Figure 2). Per share, high-p_i agents are the more effective participants, limited by throughput rather than by reception. The generator combines parents drawn from the agent's current expression and personal memory, biasing each agent's generative trajectory toward its own history when reception is sparse; received artefacts enter the recipient's own repository regardless of acceptance, and adoption (share-driven or boredom-driven) additionally substitutes external material into the generative pool. Social input therefore shapes per-agent evaluation continuously through reception and intermittently through adoption, rather than acting only on the output side. This inversion—in which the agents whose artefacts are most valued by other agents are also the least prolific contributors to the shared domain—connects to longstanding questions in computational modelling of social creativity about the relationship between individual creative quality and field-level influence (Jennings 2010; Sosa, Gero, and Jennings 2009), and is only reliably detectable at the population sizes this framework enables.

	N=50	N=100	N=250	slope
K=1	0.32	0.50	0.95	0.68
K=5	0.33	0.56	0.90	0.61
K=25	0.41	0.74	1.40	0.76

Table 2: Mean wall-clock seconds per simulation step across the grid, with fitted log-log power-law slopes in N. Cost scales sub-linearly in N for all K.

Previous simulations of agents with novelty detection, capped at around 20 agents, could not reliably separate signal from noise on such patterns. The tension between agent sophistication and population size has until now precluded empirical investigation of dynamics of this kind. The shared pretrained ResNet and batched kNN novelty detection resolve this tension, preserving per-agent subjectivity at populations of hundreds while keeping each simulation on the order of a second on a single GPU (Table 2).

The same scale opens variations that prior population ceilings made difficult to attempt. Alternative social policies, including preference-weighted partner sampling, τ_C informed by acceptance feedback, and per-agent throughput caps, each predict different redistributions of the accepted-share total across the preference range, and larger populations make these differences detectable. The generator is independently substitutable since any feature-extractable medium would do, and a broader effective produced-novelty range would shift the population dynamics observed here. The PCA-16 row of Table 1 confirms that the per-agent embedding compresses eightfold without loss, leaving headroom for larger extractors or more complex generators at the same memory budget. Perhaps the largest open axis is agent learning: preferences could drift over time, repositories could decay selectively, and generators could adapt to their own evaluative history. This framework places each of these extensions within reach at population scales where their effects are visible.

Conclusion

Per-agent novelty detection has been the principal bottleneck on population size in computational social creativity. The training-free framework presented here, using a shared pretrained ResNet-18 with per-agent kNN repositories, removes that bottleneck, supporting populations of hundreds of agents on a single consumer GPU while preserving the individual aesthetic variance that makes social creativity modelling meaningful. The per-share inversion of Figure 3 illustrates the kind of population-level finding that only becomes visible at this scale, and points toward a broader research agenda: understanding how individual creative quality, social influence, and field-level dynamics interact in large creative communities. Further feature-vector compression offers a path to larger simulations, and agent learning represents the natural next step toward models that can capture the cumulative cultural processes the field has long aspired to study.

Author Contributions

Motz conceived the study, developed the methodology, conducted the experiments, and wrote the original draft. Verhoef and Saunders provided supervision, reviewed the manuscript, and contributed to the final editing and revision.

Acknowledgments We thank the anonymous reviewers for their constructive feedback.

References

Axelrod, R. 1997. The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution* 41(2):203–226.

Bergman, L.; Cohen, N.; and Hoshen, Y. 2020. Deep nearest neighbor anomaly detection. *arXiv preprint arXiv:2002.10445*.

Berlyne, D. E. 1971. *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.

Corneli, J. 2016. An institutional approach to computational social creativity. In *Proc. 7th Int. Conf. Computational Creativity (ICCC 2016)*, 131–138.

Csikszentmihalyi, M. 1988. Society, culture, and person: A systems view of creativity. In Sternberg, R. J., ed., *The Nature of Creativity: Contemporary Psychological Perspectives*. Cambridge University Press. 325–339.

Epstein, J. M., and Axtell, R. L. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press.

Gabora, L. 1995. Meme and variations: A computer model of cultural evolution. In Nadel, L., and Stein, D. L., eds., *1993 Lectures in Complex Systems*, 471–486.

Gabora, L. 2008. EVOC: A computer model of the evolution of culture. In Sloutsky, V.; Love, B.; and McRae, K., eds., *Proc. 30th Annual Meeting of the Cognitive Science Society*. Sheridan Publishing.

Gatys, L. A.; Ecker, A. S.; and Bethge, M. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.

Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F. A.; and Brendel, W. 2018. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*.

Guckelsberger, C.; Salge, C.; Saunders, R.; and Colton, S. 2016. Supportive and antagonistic behaviour in distributed computational creativity via coupled empowerment maximisation. In *Proc. 7th Int. Conf. Computational Creativity (ICCC 2016)*, 9–16.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Imasato, N.; Miyazawa, K.; Nagai, T.; and Horii, T. 2025. Creative agents: Simulating the systems model of creativity with generative agents. *IEEE Access* 13:157712–157729.

Jennings, K. E. 2010. Developing creativity: Artificial barriers in Artificial Intelligence. *Minds Mach.* 20(4):489–501.

Jennings, K. E. 2011. A computational perspective on human exploratory creativity: Theory and methods. In *Proc. 2nd Int. Conf. Computational Creativity (ICCC 2011)*.

Kirby, S.; Cornish, H.; and Smith, K. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc. National Academy of Sciences* 105(31):10681–10686.

Lin, Y.-C.; Chen, K.-C.; Li, Z.-Y.; Wu, T.-H.; Wu, T.-H.; Chen, K.-Y.; Lee, H.-y.; and Chen, Y.-N. 2025. Creativity in LLM-based multi-agent systems: A survey. In *Proc. 2025 Conf. Empirical Methods in Natural Language Processing (EMNLP 2025)*, 27584–27607. Suzhou, China: Association for Computational Linguistics.

Linkola, S.; Takala, T.; and Toivonen, H. 2016. Novelty-seeking multi-agent systems. In *Proc. 7th Int. Conf. Computational Creativity (ICCC 2016)*, 1–8.

Martindale, C. 1990. *The Clockwork Muse: The Predictability of Artistic Change*. New York: Basic Books.

Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Political Economy of Institutions and Decisions. Cambridge University Press.

Park, J. S.; O'Brien, J. C.; Cai, C. J.; Morris, M. R.; Liang, P.; and Bernstein, M. S. 2023. Generative agents: Interactive simulacra of human behavior. In *Proc. 36th Annual ACM Symposium on User Interface Software and Technology (UIST 2023)*, 1–22. ACM.

Peeperkorn, M.; Saunders, R.; Bown, O.; and Jordanous, A. 2022. Mechanising conceptual spaces using variational autoencoders. In *Proc. 13th Int. Conf. Computational Creativity (ICCC 2022)*, 287–290.

Saunders, R., and Bown, O. 2015. Computational social creativity. *Artificial Life* 21(3):366–378.

Saunders, R., and Gero, J. S. 2001. The digital clockwork muse: A computational model of aesthetic evolution. In *Proc. AISB'01 Symposium on AI and Creativity in Arts and Science*, 12–21. York, UK: SSAISB.

Saunders, R. 2011. Artificial creative systems and the evolution of language. In *Proc. 2nd Int. Conf. Computational Creativity (ICCC 2011)*, 36–41.

Saunders, R. 2012. Towards autonomous creative systems: A computational approach. *Cognitive Computation* 4(3):216–225.

Sosa, R., and Gero, J. S. 2005. A computational study of creativity in design: The role of society. *AI EDAM: Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 19(4):229–244.

Sosa, R.; Gero, J. S.; and Jennings, K. E. 2009. Growing and destroying the worth of ideas. In *Proc. 7th ACM Conf. Creativity and Cognition*, 295–304. Berkeley, CA: ACM.

Steels, L. 1995. A self-organizing spatial vocabulary. *Artificial Life* 2(3):319–332.

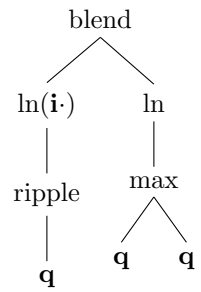
Vallinder, A., and Hughes, E. 2025. Cultural evolution of cooperation among LLM agents. In *Proc. 24th Int. Conf. Autonomous Agents and Multiagent Systems (AAMAS 2025)*.

Verhoef, T.; Kirby, S.; and de Boer, B. 2014. Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *J. Phon.* 43:57–68.

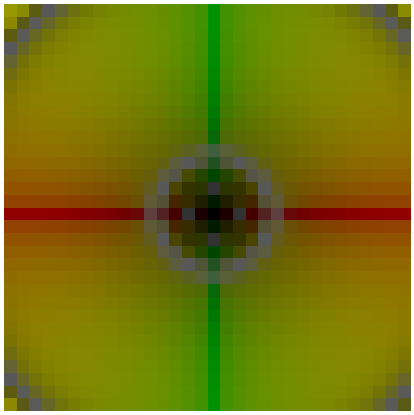
Verhoef, T.; Kirby, S.; and de Boer, B. 2016. Iconicity and the emergence of combinatorial structure in language. *Cognitive Science* 40(8):1922–1940.

Witbrock, M., and Reilly, S. N. 1999. Evolving genetic art. In Bentley, P. J., ed., *Evolutionary Design by Computers*. San Francisco: Morgan Kaufmann. 251–259.

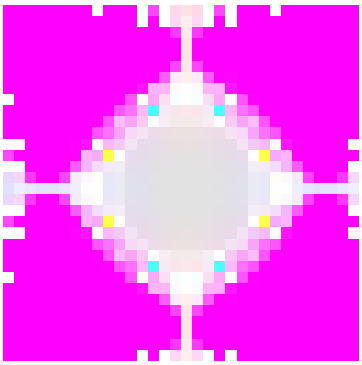
Zhang, A., and Saunders, R. 2012. Towards the evolution of a language for creative design. In *Proc. 2012 IEEE Congress on Evolutionary Computation (CEC)*, 2778–2783.



(a) Evolved AST (Genotype)

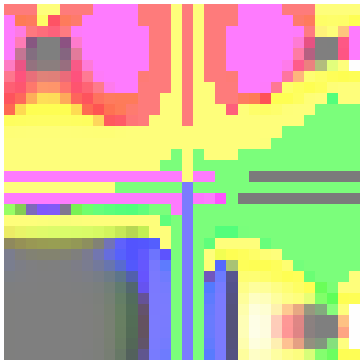


(b) Resulting Artefact (Phenotype)



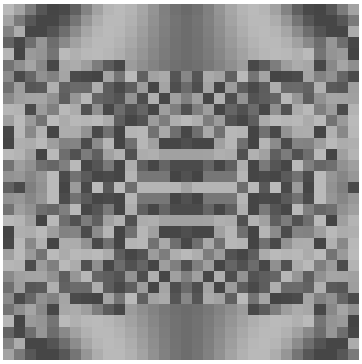
(c) Artefact 1

$$\cosh(\ln(\mathbf{i}\sin(\mathbf{i}\sin(\cos^3(\mathbf{q})))))$$



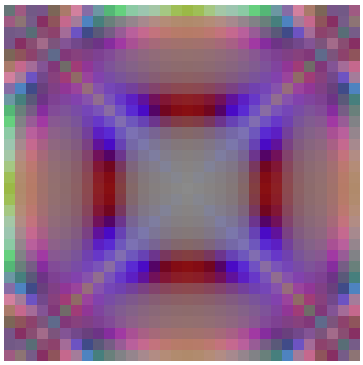
(d) Artefact 2

$$\text{spiral}\left(\left(\text{swirl}(\ln(\mathbf{i}\sin(\mathbf{iq}))) \cdot \text{spiral}(e^{\mathbf{i}\overline{\mathbf{q}}})\right)^3\right)$$



(e) Artefact 3

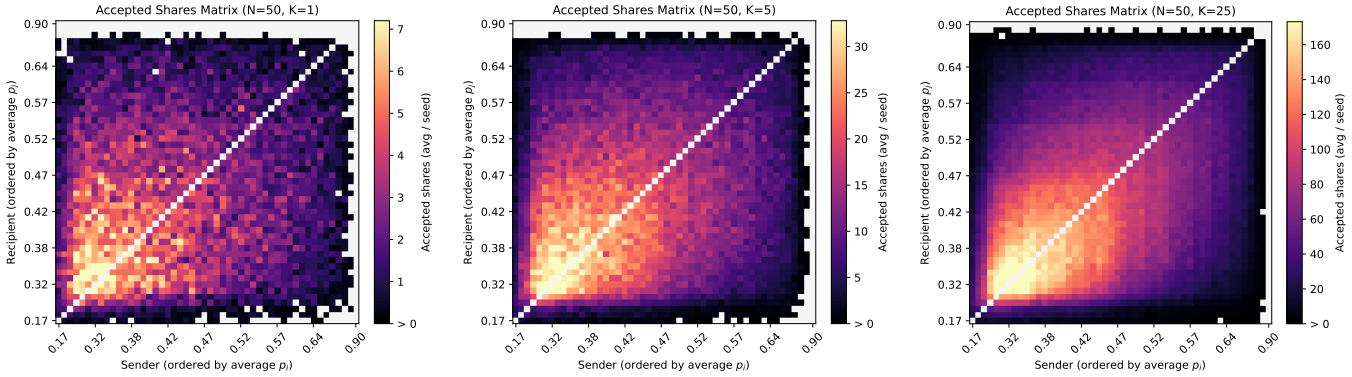
$$\text{ripple}\left(\lfloor\cosh^3(\text{blend}(\mathbf{q},\mathbf{q}))\rfloor - |\sqrt{\mathbf{q}}|^3\right)$$



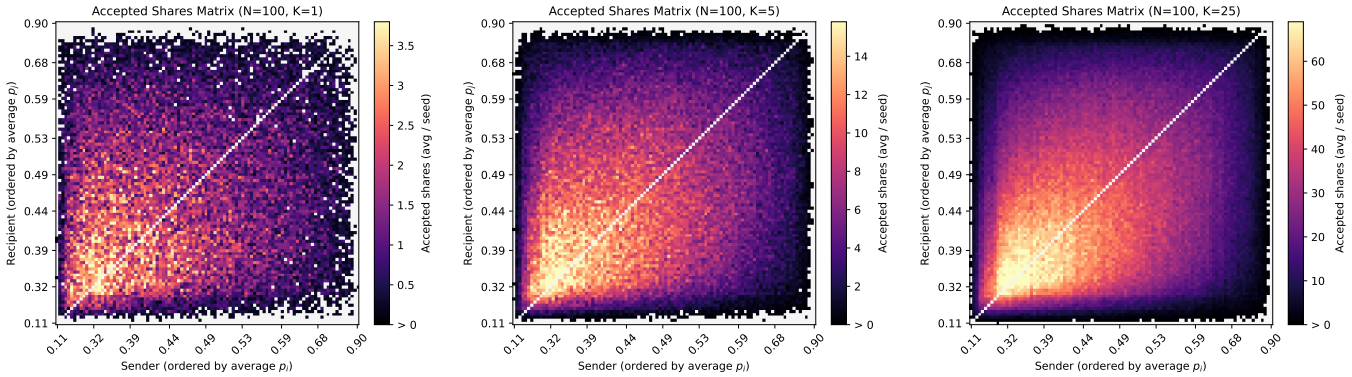
(f) Artefact 4

$$\tan\left(\sqrt{\sinh\left(\cosh(\overline{\text{rot}_{45}(\mathbf{q})})\right)}\right)$$

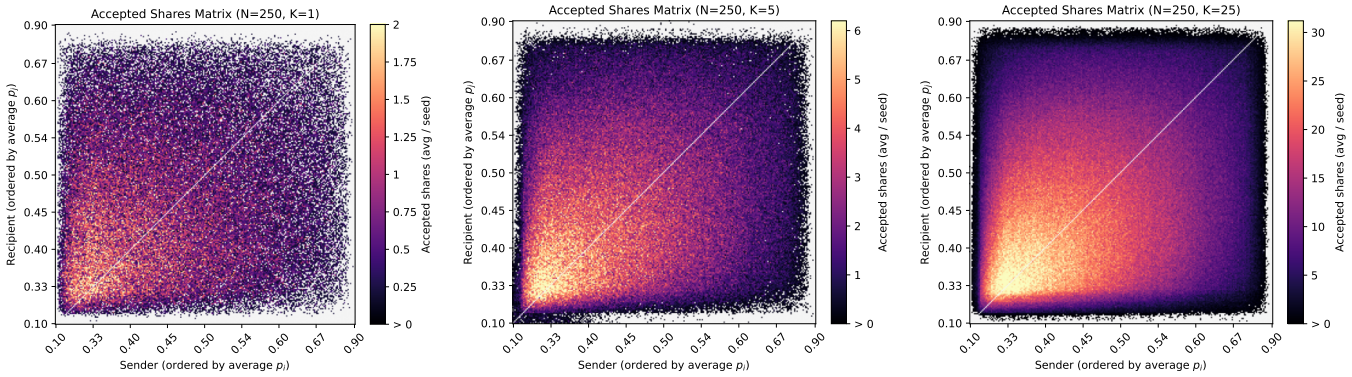
Figure 4: Examples of generated artefacts and their underlying quaternion expressions. The top row pairs an Abstract Syntax Tree (genotype) with its rendered image (phenotype). The four lower artefacts are rendered at 32×32 from independently sampled trees of equal maximum depth. The quaternion $\mathbf{q}$ in the expressions denotes the (x, y) pixel coordinate at which the expression is evaluated. Expressions are direct, unsimplified translations of the evolved ASTs, so they may contain redundancies such as $\text{blend}(\mathbf{q}, \mathbf{q})$ or unreduced nestings. These reflect the mechanical structure of subtree crossover rather than algebraic identities. Operators spiral, swirl, ripple, and rot_{45} are spatial transforms on the quaternion input. blend is a per-channel mean of two quaternions.



(a) $N = 50, K = 1$ (b) $N = 50, K = 5$ (c) $N = 50, K = 25$

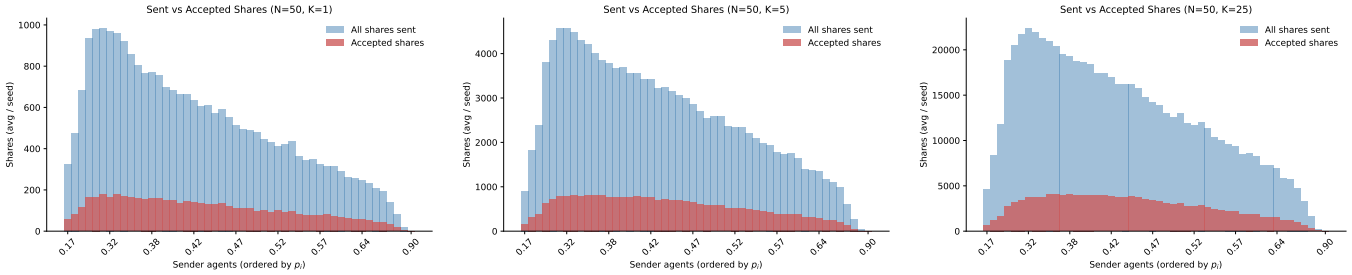


(d) $N = 100, K = 1$ (e) $N = 100, K = 5$ (f) $N = 100, K = 25$



(g) $N = 250, K = 1$ (h) $N = 250, K = 5$ (i) $N = 250, K = 25$

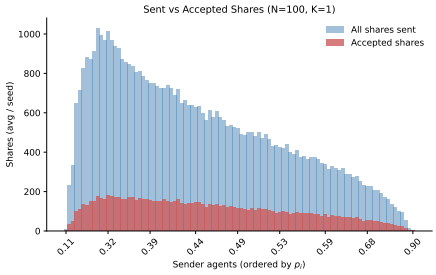
Figure 5: Complete evaluation grid of accepted-share interaction matrices across all experimental configurations. Each cell counts run-averaged accepted shares from sender i (x-axis) to recipient agent j (y-axis), with both axes ordered by preferred novelty (p_i). A white cell represents zero accepted shares across all 5 runs, whereas a black cell represents at least one successful share during one run (therefore the average could be below 1).



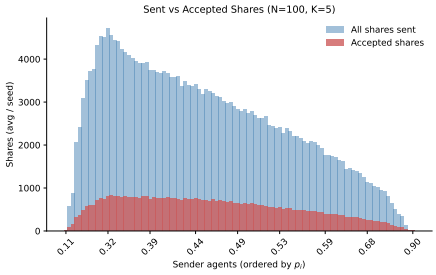
(a) $N = 50, K = 1$

(b) $N = 50, K = 5$

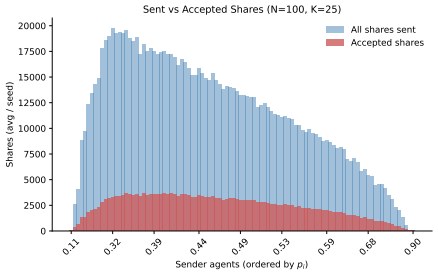
(c) $N = 50, K = 25$



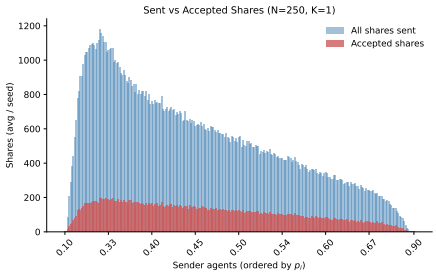
(d) $N = 100, K = 1$



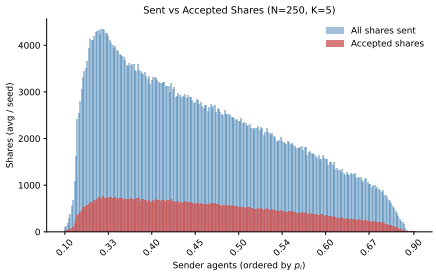
(e) $N = 100, K = 5$



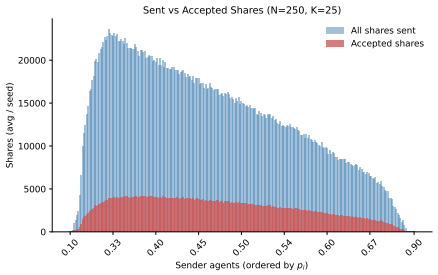
(f) $N = 100, K = 25$



(g) $N = 250, K = 1$

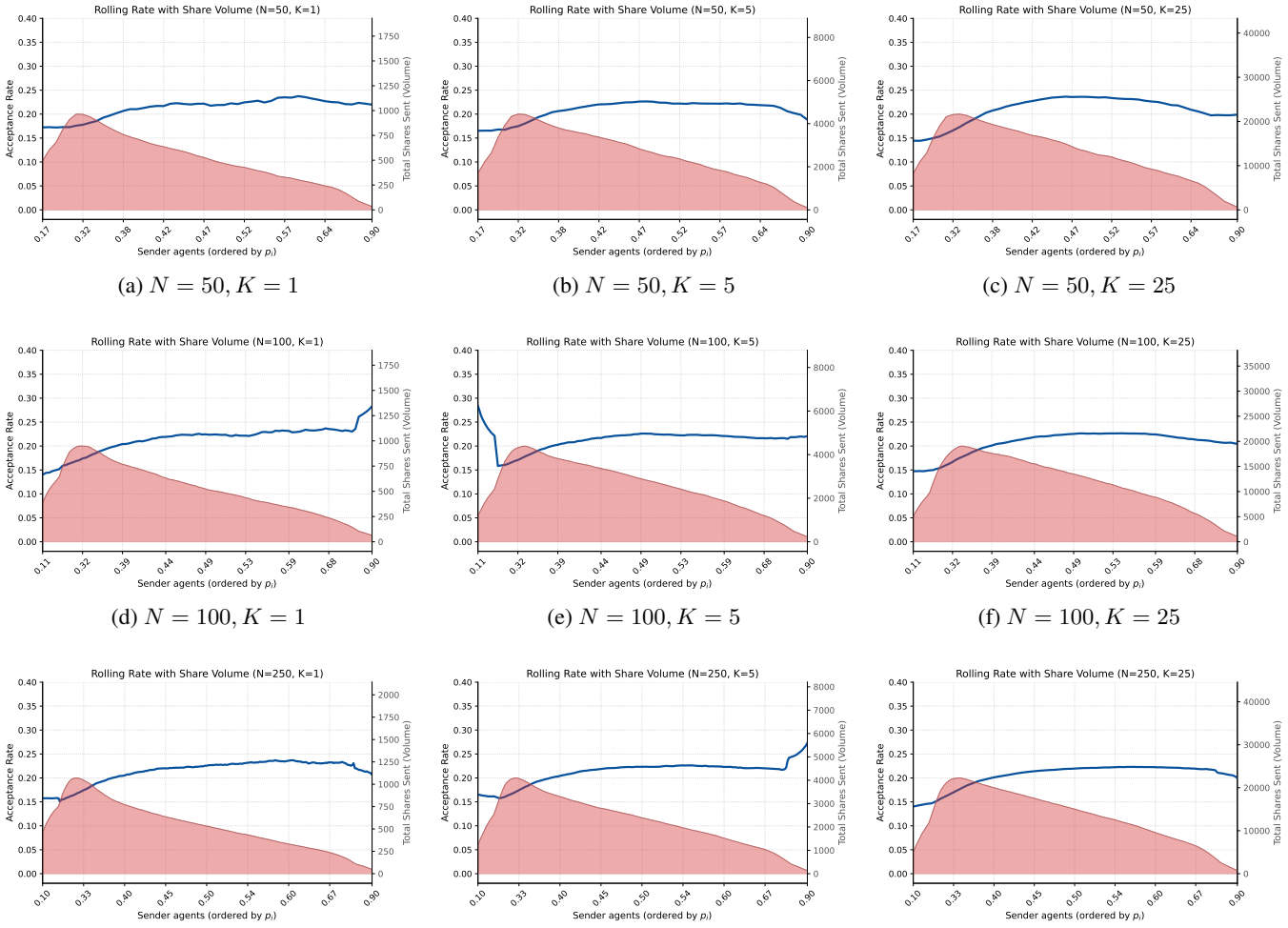


(h) $N = 250, K = 5$



(i) $N = 250, K = 25$

Figure 6: Complete evaluation grid of run-averaged per-sender share volume across all experimental configurations, ordered by individual preferred novelty (p_i). Each subfigure tracks total shared artefacts alongside the accepted subset surviving the recipient-side domain-threshold (τ_D) filter.



(a) $N = 50, K = 1$ (b) $N = 50, K = 5$ (c) $N = 50, K = 25$

(d) $N = 100, K = 1$ (e) $N = 100, K = 5$ (f) $N = 100, K = 25$

(g) $N = 250, K = 1$ (h) $N = 250, K = 5$ (i) $N = 250, K = 25$

Figure 7: Complete evaluation grid of run-averaged per-sender acceptance rates (rolling curves, left axes) overlaying total shares sent (right axes) plotted against individual preferred novelty (p_i). The scaling behaviour underscores the persistence of the rate–volume inversion across all tested population structures and communication capacities. The only anomaly seems to appear at $N = 100$, $K = 5$, where a small group of low-p_i senders achieves a noticeably higher per-share acceptance rate than the rest.