



**HAL**  
open science

## Cultural evolution in populations of Large Language Models

Jérémy Perez, Corentin Léger, Marcela Ovando-Tellez, Chris Foulon, Joan Dussauld, Pierre-Yves Oudeyer, Clément Moulin-Frier

► **To cite this version:**

Jérémy Perez, Corentin Léger, Marcela Ovando-Tellez, Chris Foulon, Joan Dussauld, et al.. Cultural evolution in populations of Large Language Models. 2024. hal-04715009

**HAL Id: hal-04715009**

**<https://hal.science/hal-04715009v1>**

Preprint submitted on 29 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Cultural evolution in populations of Large Language Models

Jérémy Perez<sup>\*1</sup>, Corentin Léger<sup>1</sup>, Marcela Ovando-Tellez<sup>2</sup>, Chris Foulon<sup>2</sup>, Joan Dussauld<sup>3</sup>, Pierre-Yves Oudeyer<sup>1</sup>, and Clément Moulin-Frier<sup>1</sup>

<sup>1</sup>Flowers Team, INRIA, Bordeaux, France

<sup>2</sup>GIN-IMN, CNRS, Bordeaux, France

<sup>3</sup>Independent Researcher

## Abstract

Research in cultural evolution aims at providing causal explanations for the change of culture over time. Over the past decades, this field has generated an important body of knowledge, using experimental, historical, and computational methods. While computational models have been very successful at generating testable hypotheses about the effects of several factors, such as population structure or transmission biases, some phenomena have so far been more complex to capture using agent-based and formal models. This is in particular the case for the effect of the transformations of social information induced by evolved cognitive mechanisms. We here propose that leveraging the capacity of Large Language Models (LLMs) to mimic human behavior may be fruitful to address this gap. On top of being an useful approximation of human cultural dynamics, multi-agents models featuring generative agents are also important to study for their own sake. Indeed, as artificial agents are bound to participate more and more to the evolution of culture, it is crucial to better understand the dynamics of machine-generated cultural evolution. We here present a framework for simulating cultural evolution in populations of LLMs, allowing the manipulation of variables known to be important in cultural evolution, such as network structure, personality, and the way social information is aggregated and transformed. The software we developed for conducting these simulations is open-source and features an intuitive user-interface, which we hope will help to build bridges between the fields of cultural evolution and generative artificial intelligence.

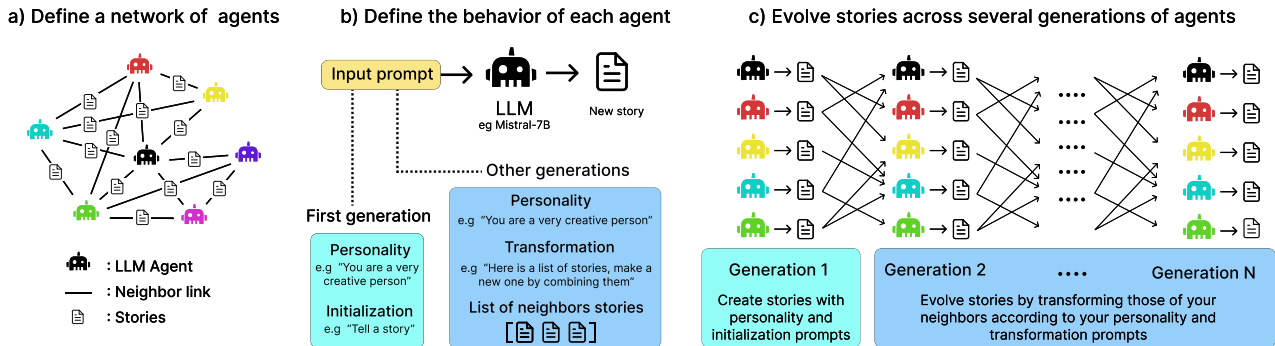
## 1 Introduction

Myths and stories are found in every human culture around the world. Research using phylogenetic methods has revealed that myths evolve in a way analogous to genes, progressively changing as some elements are discarded and new elements added as they are passed through generations [44]. Some of these myths display a surprising stability, being retained with minor modifications though generations. For example, the Baku dream-eater demons, mythological creatures made of various animal parts, have been part of the Japanese folklore since at least the 15th century [31]. Moreover, myths in distinct cultures seem to exhibit shared features, possibly pointing to convergent evolution. This is for example the case of historical myths (i.e., myths

depicting events considered foundational for a given group), which have been found to possess many similar elements, such as narratives featuring the collective challenges faced in the history of the group [41]. This raises questions regarding the factors that determine the cultural success, the stability, and the directions of evolution of myths and stories.

More generally, providing causal explanations to the change of culture over time is the central aim of research in cultural evolution. Contributions to this question are generally seen as falling into two schools of thought: the first one, referred to as the Californian school, adopts the framework of gene-culture co-evolution [5] [7] [40] [18]. This school emphasizes how evolved social learning mechanisms and transmission

<sup>\*</sup>Corresponding author: [jeremy.perez@inria.fr](mailto:jeremy.perez@inria.fr)



**Figure 1:** (a) LLM agents are organized into networks wherein each agent interacts with neighboring agents by exchanging stories. (b) Each agent is assigned a specific personality and either initialization instructions (for the first generation) or transformation instructions (after the first generation), serving as prompts for generating new stories from their neighbors’ narratives. (c) Once the network structure and agent characteristics are defined, we simulate the cultural evolution of texts across generations of agents. The simulation begins by prompting agents to initialize stories, after which we allow the narratives to evolve dynamically through interactions within the agent network

biases respectively contribute to the stability of culture and the direction of its evolution, and assume that the logic of natural selection applies to cultural traits as it does for genes [40]. The second school, referred to as the Parisian school, adopts the framework of Cultural Attraction Theory (CAT) [43][36] [31]. Disagreeing on the central role attributed to selection, this view rather emphasizes how non-random transformations of cultural information during transmission events can explain how cultural traits progressively evolve toward stable forms, which are referred to as attractors. Taking inspiration from population genetics, the Californian school has successfully employed computational models to generate predictions, for instance about the effect of group size [17], network structure [10] [12], or transmission fidelity [27]. As for CAT, although computational models have also been proposed [1] [30] [8], they were mainly aimed at making conceptual points rather than to generate testable predictions, and attraction dynamics have thus been studied mainly through experiments [21] [33] [32] [9][9] [14] and historical data analysis [22] [13].

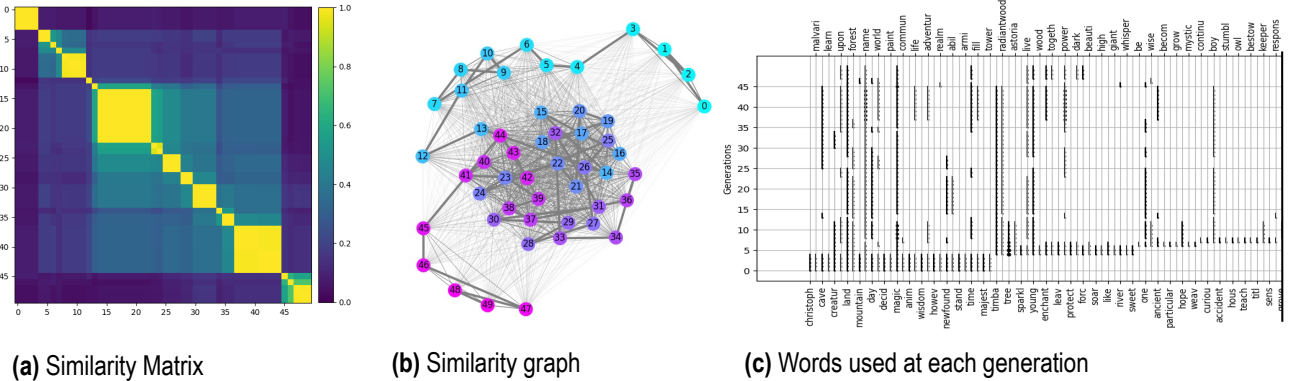
This prevalence of experimental and historical methods is likely attributable to the difficulty in capturing factors of attraction—such as evolved psychological mechanisms—that influence cultural evolution within agent-based or formal models. Here, we propose that this limitation of agent-based models can be overcome by using generative artificial agents. In particular, using Large Language Models (LLMs) to simulate the evolution of linguistic culture appears fruitful, as we can expect those models to transform cultural information in realistic ways.

On top of allowing to generate hypotheses about the causes of cultural change in human societies, studying the evolution of culture in populations of generative agents is also crucial to understand the dynamics

of AI-generated cultural evolution. Indeed, although technology has long influenced cultural evolution, for example with the invention of the printing press, some authors have argued that the recent advances in generative algorithms are likely to result in an unprecedented shift in cultural evolution. The fact that algorithms are now participating in the creation of cultural traits sets us at the very beginning of an era of “machine culture”, defined as “culture mediated or generated by machines” [6]. Therefore, studying the dynamics of machine-generated culture becomes highly important. Applying paradigms from Cultural Attraction theory to study LLMs has already begun to generate precious insights about the bias exhibited by LLMs when transmitting linguistic content [2].

Overall, it appears that there is much to gain by using LLMs as agents to simulate the evolution of culture. So far, the fields of cultural evolution and generative artificial intelligence have remained quite disconnected, and facilitating exchanges between these fields seems promising for generating important findings. Some parallel works have investigated the use of Large Language Models as agents in multi-agent simulations, although these studies were not specifically designed to model cultural evolution [38][47].

Here, we propose an open-source software for simulating the transmission and evolution of linguistic culture in populations of Large Language models. This software features an intuitive interface that allows to manipulate many variables of interest, such as the structure of the social network, the ways in which different sources of information are aggregated, or the personalities of the different agents. We also introduce several visualizations and measures that are useful for tracking the results of the simulations. We hope that this tool will pave the way for more interaction between the fields of cultural evolution and generative artificial



**Figure 2:** Visualization of the evolution of the texts generated along a chain of 50 agents (i.e. 50 generations of one agent per generation). **(a)** The similarity matrix represents the semantic similarity between all stories generated. The color of the cell at row  $i$ , column  $j$  corresponds to the similarity of the stories  $i$  and  $j$ , which here corresponds to the stories generated at generation  $i$  and generation  $j$ . **(b)** The similarity graph is another way of visualizing the similarity between all generated stories. Each node corresponds to one story, and the distance between nodes is proportional to the semantic distance between corresponding stories. Stories generated at successive generations are linked by a wider edge and are assigned similar colors. **(c)** We also visualize the words used at each generation. Words are along the x-axis, and generation along the y-axis. A dot at position  $(x,y)$  means that the word  $x$  was used in a story at generation  $y$ . A line means that the corresponding word was used by two successive generations. We only display the first half of the figure here. The complete figure can be found in the Fig. 15 of Appendix.

intelligence. This software is open source and accessible at this link.

personality or have different personalities.

## 2 Methods

## 2.2 Analysis

### 2.1 Description of the model

#### 2.2.1 Similarity

The model presented here simulates the cultural evolution of linguistic content in a population of large language models (LLMs). Each agent can be seen as an independent instance of a LLM. The agents are arranged according to a specified social network structure (Fig. 1.a). Network structures that can currently be generated with the software are detailed in the Appendix A. At the first generation, all agents are prompted with an Initialization Prompt. This prompt describes what kind of content the agents should generate. For example, one could choose “Tell me a story” as the initialization prompt. All agents then output an answer by passing the initialization prompt to their respective instance of the LLM. The agents then transmit stories to their neighbors (according to the specified network structure): each agent receives a new prompt, which is the concatenation of a Transformation Prompt and the list of stories produced by its neighbors at the previous generation (Fig. 1.b). An example of a Transformation Prompt could be “Here are some stories. Make up a new one by combining two of them”. Agents can additionally be provided with a personality, which is always added at the top of their prompt. For example, a personality prompt could be “You are very imaginative”. Agents may either all have the same

The main metric we use to analyze our results is the similarity between texts. To compute this metric, we first use TfidfVectorizers from scikit-learn [39] to convert texts into meaningful numerical representations. We then compute the cosine similarity [16] between all the texts generated, resulting in a similarity matrix of size  $(N_{agents} * N_{generations}) \times (N_{agents} * N_{generations})$ , where the color of the cell  $(i,j)$  represents the similarity between story  $i$  and story  $j$ .

From this similarity matrix, we extract more interpretable measures. The first one is the within-generation similarity, which captures how similar are the texts generated at a given generation with each other. The second one is the successive similarity, which represents the average similarity between the texts generated at a given generation and the texts generated at the previous generation. The last one is the similarity with the first generation, which is the average similarity between the texts generated at a given generation and the texts generated at the first generation.

#### 2.2.2 Visualization

We also provide two visualization techniques that offer qualitative insights into the generated data.

**Word chains** We extract the key words from each text and represent their evolution through generations. To extract keywords from texts, we tokenize the text into words, remove common stopwords and non-alphanumeric tokens, calculate the frequency distribution of the remaining words, and select the top keywords based on their frequency. This allows to visualize which words are the most frequent, the most stable, or the most often reinvented.

**Similarity network** We also represent the similarity between generations using a graph network, where each node represents a generation of texts. The positioning of nodes is determined by a layout algorithm provided by the NetworkX library [15], arranging them based on their similarities and interconnections. Generations with the highest similarity are positioned closer together, and successive generations, represented by similar colors, are linked by thicker edges. This approach provides an intuitive depiction of the evolutionary dynamics of the generated content.

### 3 Preliminary Results

The results presented in this paper are preliminary, and are mainly meant to illustrate how one can manipulate with the different variables to see their effects on the measures described above. However, they do already offer some interesting insights.

#### 3.1 Transmission chain

We first illustrate the dynamics of the model using a linear transmission chain of 50 agents (i.e one agent per generation for 50 generations). The initialization and transmission prompts used are provided in the Section A of Appendix. The agents were not assigned any personality for this experiment (personality prompt is empty). Fig. 2.a shows the similarity matrix for this chain. We can notice that stories seem to evolve in a punctuated manner: there is an alternation of phases where stories are transmitted with no modifications, and phases where stories are modified. These dynamics recall previous observations from experiments and modelling work in cultural evolution [24] [46] [35] [34], where cultural information was found to evolve in a sequence of bursts and stasis. The transition matrix also suggests the existence of hierarchically structured clusters of stories that are not identical but share some degree of similarity. Fig. 2.b shows the graph of similarity of this chain, which allows to visualize how stories progressively evolve through a semantic space. Looking at this graph suggests that stories got "trapped" in a specific part of the semantic space for a while (from generation 14 to 41), possibly indicating the presence

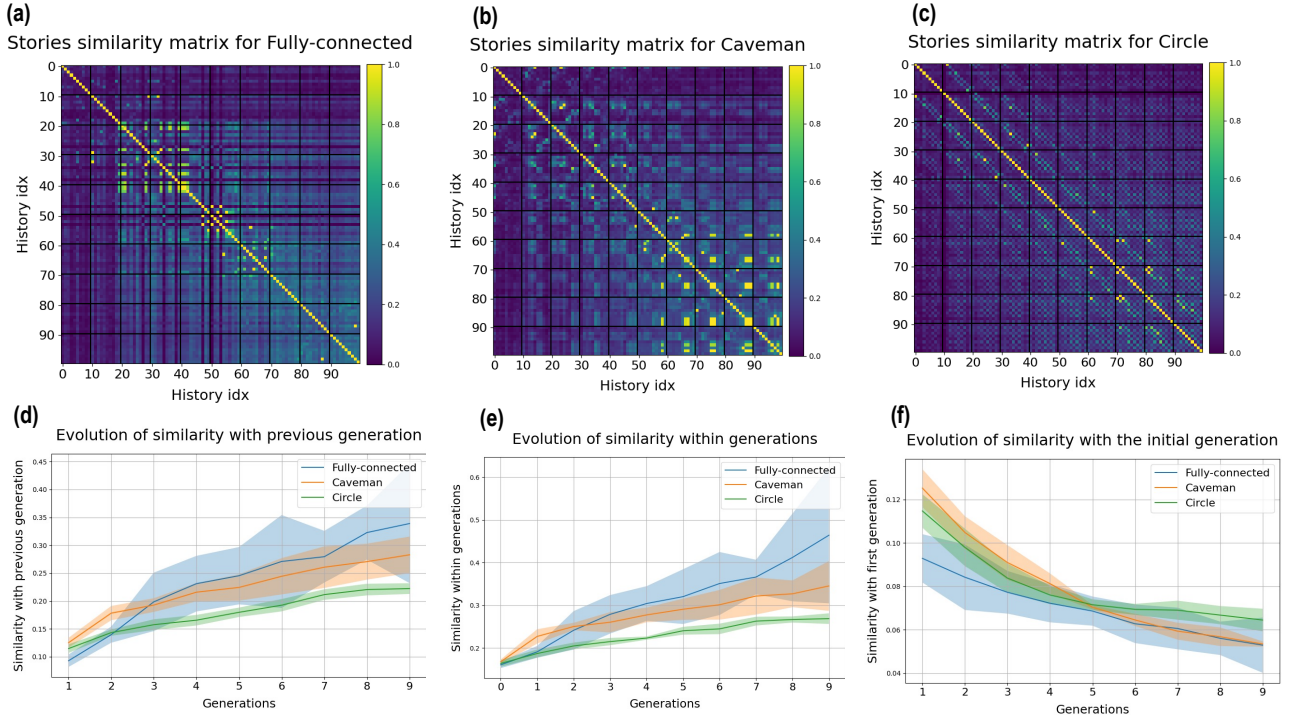
of an attractor. Fig. 2.c shows the chain-of-words representation for this chain. This visualization allows to notice that some words are very stable, being kept almost throughout the chain, such as the word "magic", while some words only last for the first few generations before being lost, such as the word "learn".

#### 3.2 Effect of network structure

We first manipulated the network structure by running 5 simulations of 10 agents for 10 generations, for each network structure (Fig .3). We compared three network structures: a fully-connected network, a circle network, and a caveman network with 2 cliques (see Fig .S1 of the Appendix). We only show one of the five similarity matrices for each network structure, and provide all matrices in Appendix C.2. Qualitatively, the similarity matrices reveals distinct patterns for each network structure: while similarity appears to increase homogeneously for the fully-connected network, it does so in a clustered manner for the caveman and circle networks. Looking at the evolution of the similarity measures, we observe that for all structures, stories get more and more similar to stories of the same generation and of the previous generation, indicating a progressive homogenisation of the population. We also observe that stories get more and more dissimilar with stories from the first generation. For all measures, these dynamics seem to happen at a faster rate for the fully-connected structure than for the caveman network, itself faster than the circle network. These results replicate findings from previous work in network science and cultural evolution, where models and experiments reveal that more efficient networks (that is, networks with smaller average path length) lead to quicker diffusion of information and thus lower diversity [11] [12] [37] [26].

#### 3.3 Effect of transformation prompt

We then manipulated the transformation prompt by running 5 simulations of 10 agents for 10 iterations (Fig. 4). We compared the effect of four different transformation prompts referred to as "Combine2", "MinorChanges", "Repeat" and "MaximizeDifference". The exact content of these prompts is provided in Appendix A. These dynamics reveal that clear differences can be observed, and those differences are aligned with what would be expected. For example, the Repeat prompt leads to high within- and between- generation similarity, while these metrics are low for the MaximizeDifference prompt. Apart for the MaximizeDifference condition, all conditions show a gradual increase in the similarity with previous generation and within-generation, indicating a progressive homogenisation of the cultural content. Lastly, for all conditions but the Repeat condition, similarity with the initial generation appears to decrease over time. This is in particular in-



**Figure 3: Effect of network structure :** Cultural dynamics of a population of 10 agents over 10 generations for three different types of network structures. (a), (b), (c) Similarity matrices for a fully-connected network, a caveman network with 2 cliques, and a circle network. The index of a story is defined as the agent’s index (here between 0 and 10) +  $N_{agents}$  \* generation-index. For example, the story 3 belongs to the first generation, and story 13 to the second generation. The color of the cell at row  $i$  and column  $j$  represents the semantic similarity between stories  $i$  and  $j$ . Black lines mark the separation between generations. (d) Evolution of the average similarity between stories produced at a generation and stories at the generation just before. (e) Evolution of the average similarity between each pair of stories produced at a given generation. (f) Evolution of the average similarity between stories produced a given generation and stories produced at the first generation. Lines represent averages over 5 simulations, and the filled areas represent the standard deviations.

teresting for the MaximizeDifference condition, as the second generation was already instructed to maximize the difference with the first generation.

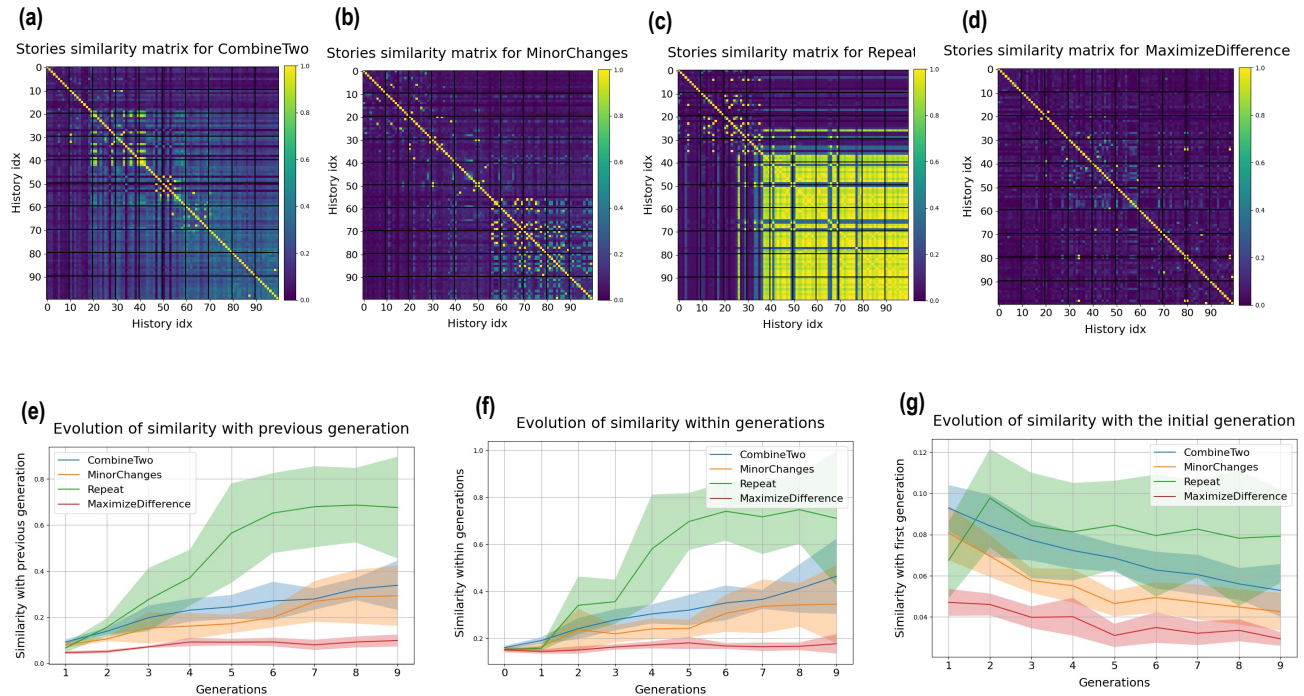
### 3.4 Effect of different personalities

We then manipulated the personality assigned to agents by running 5 simulations of 10 agents for 10 generations (Fig. 5). We compared 4 different conditions. In the first one, agents were not assigned any personality. In the second one, all agents were assigned the personality "Creative". In the third one, all agents were assigned the personality "NotCreative". In the fourth one, half of the agents were assigned the personality "Creative" and the other half was assigned the personality "NotCreative". Details of the personality prompts can be found in Appendix A. Although performing more simulations would be necessary to draw conclusions, it does seem that different personalities differentially impact cultural dynamics. In particular, our results suggest that agents in the Creative condition generate more variation than agents in the NotCreative

condition: indeed, the NotCreative appears to exhibit higher similarity with the first generation, between successive generations, and within generations compared to the Creative condition. Also interesting is the fact the the MixedPopulation condition seems to be more similar to the Creative condition than to the NotCreative condition, even though it contains Creative and NotCreative agents in equal proportions. This suggests that the cultural dynamics of an heterogeneous population may not reflect the proportions of the different personalities it contains.

## 4 Discussion

We introduced a framework for modelling cultural evolution in population of LLMs agents, along with several metrics and visualizations that allows to extract qualitative and quantitative insights about the resulting cultural dynamics. We presented how this framework allows to manipulate several variables known to impact cultural evolution, namely the network structure, the



**Figure 4: Effect of transmission prompts :** Cultural dynamics of a population of 10 agents over 10 generations for four different transmission prompts. (a), (b), (c), (d) Similarity matrices for the "CombineTwo", "MinorChanges", "Repeat" and "MaximizeDifference" transformation prompts. (e) Evolution of the average similarity between stories produced at a generation and stories at the generation just before. (f) Evolution of the average similarity between each pair of stories produced at a given generation. (g) Evolution of the average similarity between stories produced a given generation and stories produced at the first generation. Lines represent averages over 5 simulations, and the filled areas represent the standard deviations.

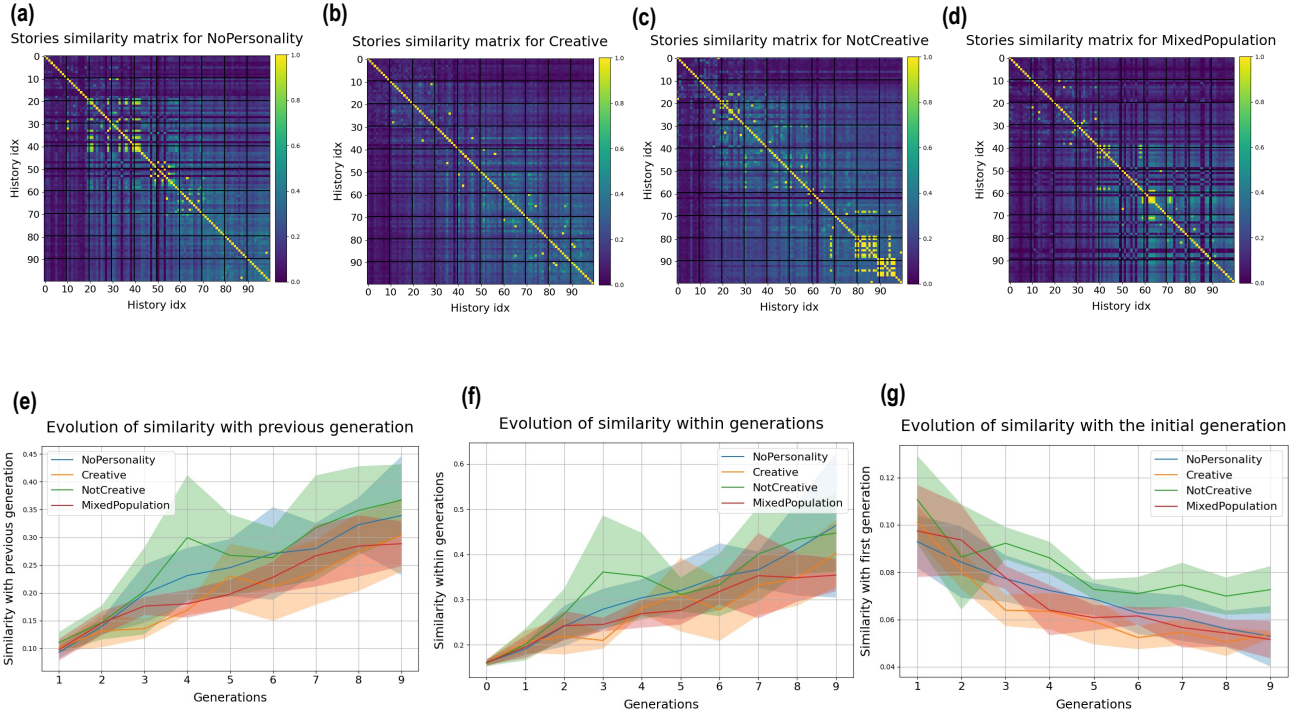
way of transforming previous social information, and agents' personalities. Although the results presented in this paper are still preliminary, they already reveal several insights. First, the fact our simulations replicate several findings from empirical and theoretical work in cultural evolution confirms that LLMs-based multi-agent models are an adequate tool for generating hypotheses about the dynamics of human culture. It also suggests that results from studies of human-generated culture may apply to machine-generated culture. Moreover, we found that manipulating personality and social information transformation appears to have significant impact on the observed dynamics. This suggests that using LLMs to study cultural evolution is promising, as these variables have so far been difficult to model using traditional modelling tools. It also indicates that the dynamics of machine-generated culture will likely be influenced by the specific way in which generative agents are instructed to use pre-existing social information, and which personality traits to emulate.

These results were preliminary and should mainly be seen as proof of concept. Future work will involve performing more systematic and rigorous analyses of the effect of the different variables on cultural evolution. It would also be interesting to compare how dif-

ferent groups, all starting with the same story, evolve over time.

Although Large Language models are a useful proxy of human behavior, some precautions must be taken when generalizing results of such simulation to human behavior. First, because of the biases in their training set, LLMs are mainly representative of western culture [3]. As such, they may not be able to capture the specific dynamics of other cultures. A potential way to alleviate that is to use models that were trained on more representative datasets, such as the BLOOM model whose training data includes 46 different languages [49]. More generally, precisely evaluating to what extent current models miss some aspects of non-western cultures is a major direction for future work on machine-generated cultural evolution. A second limitation to keep in mind is that in some situations, LLMs may struggle to role-play and impersonate specific characters [25]. Better understanding to what extent they take into account their assigned personality when creating cultural content is a crucial step in order to derive hypotheses about human culture using such models.

Another important difference between this model and human cultural evolution is that the latter is grounded in a physical environment. As such, the suc-



**Figure 5: Effect of personalities :** Cultural dynamics of a population of 10 agents over 10 generations for four different personalities. (a), (b), (c), (d) Similarity matrices for "NoPersonality", "Creative", "NotCreative" and Mixed population of "Creative" and "NotCreative". (e) Evolution of the average similarity between stories produced at a generation and stories at the generation just before. (f) Evolution of the average similarity between each pair of stories produced at a given generation. (g) Evolution of the average similarity between stories produced a given generation and stories produced at the first generation. Lines represent averages over 5 simulations, and the filled areas represent the standard deviations

cess of cultural traits not only depend on the cognitive mechanisms of the agents transmitting it, but also on their relationship with the environment of the agents [45]. For example, beliefs in moralizing gods have been found to be predicted with high accuracy by the ecological and historical context [4]. Extending this model to include interaction with a physical environment therefore seems like a very promising direction.

Finally, the current model assumes that personalities are fixed and stable through time. However in humans, there are feedback loops between the cultural traits expressed in a group and the personalities, values and preferences of its member [42] [19]. Including such interactions would be fruitful to model phenomena such as opinion dynamics and polarization.

More generally, although we here focused on the dynamics of collective creation, this framework is suited to explore other questions related to collective behavior, including opinion dynamics, collective innovation and the evolution of language.

Overall, we have sketched how using generative agents to simulate cultural evolution may be promising, both for generating hypotheses about human cultural evolution and for better understanding the dynamics of machine-generated culture. We hope that making this tool easy to use (see usage details in Appendix B) and accessible to the scientific community will foster more exchanges between the fields of cultural evolution and generative artificial intelligence.

## Acknowledgements

This research was partially funded by the French National Research Agency (ANR, project ECOURL, Grant ANR-20-CE23-0006). This work benefitted from access to the Jean Zay (Idris) supercomputer associated with the Genci grant A0151011996. This research originated as a project from the Hackathon Hack1Robo. We also thank Maxime Derex for participating in helpful discussions.



## References

- [1] Alberto Acerbi, Mathieu Charbonneau, Helena Miton, and Thom Scott-Phillips. Culture without copying or selection. *Evolutionary Human Sciences*, 3:e50, January 2021. Publisher: Cambridge University Press.
- [2] Alberto Acerbi and Joseph M. Stubbersfield. Large language models show human-like content biases in transmission chain experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 120(44):e2313790120, October 2023.
- [3] Mohammad Atari, Mona J Xue, Peter S Park, Damián E Blasi, and Joseph Henrich. Which humans?, Sep 2023.
- [4] Carlos A. Botero, Beth Gardner, Kathryn R. Kirby, Joseph Bulbulia, Michael C. Gavin, and Russell D. Gray. The ecology of religious beliefs. *Proceedings of the National Academy of Sciences*, 111(47):16784–16789, 2014.
- [5] Robert Boyd and Peter J. Richerson. *Culture and the Evolutionary Process*. University of Chicago Press, June 1988. Google-Books-ID: MBg4oBsCKU8C.
- [6] Levin Brinkmann, Fabian Baumann, Jean-François Bonnefon, Maxime Derex, Thomas F. Müller, Anne-Marie Nussberger, Agnieszka Czaplicka, Alberto Acerbi, Thomas L. Griffiths, Joseph Henrich, Joel Z. Leibo, Richard McElreath, Pierre-Yves Oudeyer, Jonathan Stray, and Iyad Rahwan. Machine culture. *Nature Human Behaviour*, 7(11):1855–1868, November 2023. Number: 11 Publisher: Nature Publishing Group.
- [7] Luigi Luca Cavalli-Sforza and Marcus W. Feldman. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press, May 1981. Google-Books-ID: FDfBDwAAQBAJ.
- [8] Nicolas Claidière, Thomas C. Scott-Phillips, and Dan Sperber. How Darwinian is cultural evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1642):20130368, May 2014. Publisher: Royal Society.
- [9] Nicolas Claidière, Kenny Smith, Simon Kirby, and Joel Fagot. Cultural evolution of systematically structured behaviour in a non-human primate. *Proceedings. Biological sciences / The Royal Society*, 281, December 2014.
- [10] Nicole Creanza, Oren Kolodny, and Marcus W. Feldman. Greater than the sum of its parts? Modelling population contact and interaction of cultural repertoires. *Journal of The Royal Society Interface*, 14(130):20170171, May 2017. Publisher: Royal Society.
- [11] Maxime Derex and Robert Boyd. Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, 113(11):2982–2987, March 2016. Publisher: Proceedings of the National Academy of Sciences.
- [12] Maxime Derex, Charles Perreault, and Robert Boyd. Divide and conquer: intermediate levels of population fragmentation maximize cultural accumulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1743):20170062, February 2018. Publisher: Royal Society.
- [13] Edgar Dubourg and Nicolas Baumard. Why imaginary worlds? The psychological foundations and cultural evolution of fictions with imaginary worlds. *Behavioral and Brain Sciences*, 45:e276, January 2022.
- [14] Olga Feher, Haibin Wang, Sigal Saar, Partha P. Mitra, and Ofer Tchernichovski. De novo establishment of wild-type song culture in the zebra finch. *Nature*, 459(7246):564–568, May 2009.
- [15] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.
- [16] Jiawei Han, Micheline Kamber, and Jian Pei. 2 - Getting to Know Your Data. In Jiawei Han, Micheline Kamber, and Jian Pei, editors, *Data Mining (Third Edition)*, The Morgan Kaufmann Series in Data Management Systems, pages 39–82. Morgan Kaufmann, Boston, January 2012.
- [17] Joseph Henrich. Demography and Cultural Evolution: How Adaptive Cultural Processes Can Produce Maladaptive Losses—The Tasmanian Case. *American Antiquity*, 69(2):197–214, April 2004.
- [18] Joseph Henrich. The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter. In *The Secret of Our Success*. Princeton University Press, October 2015.
- [19] E. Tory Higgins. Culture and Personality: Variability across Universal Motives as the Missing Link. *Social and Personality Psychology Compass*, 2(2):608–634, 2008. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-9004.2007.00075.x>.

- [20] Dan R. Johnson, James C. Kaufman, Brendan S. Baker, John D. Patterson, Baptiste Barbot, Adam E. Green, Janet van Hell, Evan Kennedy, Grace F. Sullivan, Christa L. Taylor, Thomas Ward, and Roger E. Beaty. Divergent semantic integration (DSI): Extracting creativity from narratives with distributional semantic modeling. *Behavior Research Methods*, 55(7):3726–3759, October 2023.
- [21] Michael L. Kalish, Thomas L. Griffiths, and Stephan Lewandowsky. Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, 14(2):288–294, 2007. Place: US Publisher: Psychonomic Society.
- [22] Piers Kelly, James Winters, Helena Miton, and Olivier Morin. The predictable evolution of letter shapes: An emergent script of West Africa recapitulates historical change in writing systems. February 2024. Publisher: OSF.
- [23] Yoed N Kenett. What can quantitative measures of semantic distance tell us about creativity? *Current Opinion in Behavioral Sciences*, 27:11–16, June 2019.
- [24] Oren Kolodny, Nicole Creanza, and Marcus W. Feldman. Evolution in leaps: The punctuated accumulation and loss of cultural innovations. *Proceedings of the National Academy of Sciences*, 112(49):E6762–E6769, December 2015. Publisher: Proceedings of the National Academy of Sciences.
- [25] Grgur Kovač, Rémy Portelas, Masataka Sawayama, Peter Ford Dominey, and Pierre-Yves Oudeyer. Stick to your role! stability of personal values expressed in large language models, 2024.
- [26] David Lazer and Allan Friedman. The Network Structure of Exploration and Exploitation. *Administrative Science Quarterly*, 52(4):667–694, December 2007. Publisher: SAGE Publications Inc.
- [27] Hannah M. Lewis and Kevin N. Laland. Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599):2171–2180, August 2012.
- [28] Wing Lian, Bleys Goodson, Guan Wang, Eugene Pentland, Austin Cook, Chanvichet Vong, and "Teknum". Mistralorca: Mistral-7b model instruct-tuned on filtered openorca1 gpt-4 dataset. <https://huggingface.co/Open-Orca/Mistral-7B-OpenOrca>, 2023.
- [29] Steven Loria. textblob documentation. *Release 0.15*, 2, 2018.
- [30] Alex Mesoudi. Cultural selection and biased transformation: two dynamics of cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376, May 2021.
- [31] Helena Miton. Cultural Attraction. February 2024. Publisher: OSF.
- [32] Helena Miton, Nicolas Claidière, and Hugo Mercier. Universal cognitive mechanisms explain the cultural success of bloodletting. *Evolution and Human Behavior*, 36(4):303–312, 2015. Place: Netherlands Publisher: Elsevier Science.
- [33] Helena Miton, Thomas Wolf, Cordula Vesper, Günther Knoblich, and Dan Sperber. Motor constraints influence cultural evolution of rhythm. *Proceedings of the Royal Society B: Biological Sciences*, 287(1937):20202001, October 2020. Publisher: Royal Society.
- [34] Elena Miu, Ned Gulley, Kevin N. Laland, and Luke Rendell. Innovation and cumulative culture through tweaks and leaps in online programming contests. *Nature Communications*, 9(1):2321, June 2018. Number: 1 Publisher: Nature Publishing Group.
- [35] Joel Mokyr. Punctuated Equilibria and Technological Progress. *The American Economic Review*, 80(2):350–354, 1990. Publisher: American Economic Association.
- [36] Olivier Morin. *How Traditions Live and Die*. Oxford University Press, 2016. Google-Books-ID: kSukC-gAAQBAJ.
- [37] Eleni Nisioti, Mateo Mahaut, Pierre-Yves Oudeyer, Ida Momennejad, and Clément Moulin-Frier. Social Network Structure Shapes Innovation: Experience-sharing in RL with SAPIENS, November 2022. arXiv:2206.05060 [cs].
- [38] Joon Sung Park, Joseph C. O’Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative Agents: Interactive Simulacra of Human Behavior, August 2023. arXiv:2304.03442 [cs].
- [39] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

- [40] Peter J. Richerson and Robert Boyd. *Not By Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, June 2008.
- [41] Amine Sijilmassi, Lou Safra, and Nicolas Baumard. ‘Our Roots Run Deep’: Historical Myths as Culturally Evolved Technologies for Coalitional Recruitment. *Behavioral and Brain Sciences*, pages 1–44, January 2024. Publisher: Cambridge University Press.
- [42] Paul E. Smaldino, Aaron Lukaszewski, Christopher von Rueden, and Michael Gurven. Niche diversity can explain cross-cultural differences in personality structure. *Nature Human Behaviour*, 3(12):1276–1283, December 2019. Number: 12 Publisher: Nature Publishing Group.
- [43] Dan Sperber. *Explaining Culture: A Naturalistic Approach*. Oxford: Basil Blackwell, 1996.
- [44] Jamshid J. Tehrani and Julien d’Huy. Phylogenetics Meets Folklore: Bioinformatics Approaches to the Study of International Folktales. In Ralph Kenna, Máirín MacCarron, and Pádraig MacCarron, editors, *Maths Meets Myths: Quantitative Approaches to Ancient Narratives*, Understanding Complex Systems, pages 91–114. Springer International Publishing, Cham, 2017.
- [45] Bill Thompson and Thomas L. Griffiths. Human biases limit cumulative innovation. *Proceedings of the Royal Society B: Biological Sciences*, 288(1946):20202752, March 2021. Publisher: Royal Society.
- [46] Sergi Valverde and Ricard V. Solé. Punctuated equilibrium in the large-scale evolution of programming languages†. *Journal of The Royal Society Interface*, 12(107):20150249, June 2015. Publisher: Royal Society.
- [47] Alexander Sasha Vezhnevets, John P. Agapiou, Avia Aharon, Ron Ziv, Jayd Matyas, Edgar A. Duéñez-Guzmán, William A. Cunningham, Simon Osindero, Danny Karmon, and Joel Z. Leibo. Generative agent-based modeling with actions grounded in physical, social, or digital space using Concordia, December 2023. arXiv:2312.03664 [cs] version: 2.
- [48] Duncan J. Watts. Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology*, 105(2):493–527, September 1999.
- [49] BigScience Workshop. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model, June 2023. arXiv:2211.05100 [cs].

## Appendix A Simulation details

**LLM used** The results presented in this paper were generated using MistralOrca [28], which is a version of openAI’s Mistral-7B Model, instruct-tuned on Filtered OpenOrcaV1 GPT-4 Dataset. We used a quantized GGUFv2 version, quantized using the method "GGML-TYPE-Q4-K".

**Prompts used for the experiments** For the chain of 50 agents (Section 3.1), the initialization prompt used was "Imagine that you are telling a story to your kid. What would that story be? Just output the story, nothing else." The transformation prompt used was "Here is one or more stories you were told as a kid. It is now your turn to tell a story at your kid. Tell that story. Write only one story. Do not output anything else."

For all other simulations, the initialization prompt was "Tell me a story". Except when manipulating the transformation prompt, the transmission prompt for all simulations was CombineTwo. Here is the complete list of transformation prompts used in our experiments :

- **CombineTwo (default):**

"You will receive stories. Pick the two stories you prefer, and create a story that is combination of these two stories. Just output your story, don't write anything else."

- **MinorChanges:**

"You will receive a list of one or more stories. Create a new story by making some minor changes to one of those stories. Just output one story, do not output anything else."

- **Repeat:**

"You will receive stories. Select only one of these stories, and repeat it. Just output the story, don't write anything else."

- **MaximizeDifference:**

"You will receive stories. Create a story that is as different as possible from the stories you received. Just output your story, nothing else."

Except when manipulating the personality, the personality prompt was an empty string for all simulations. When manipulating the personalities, the prompts were:

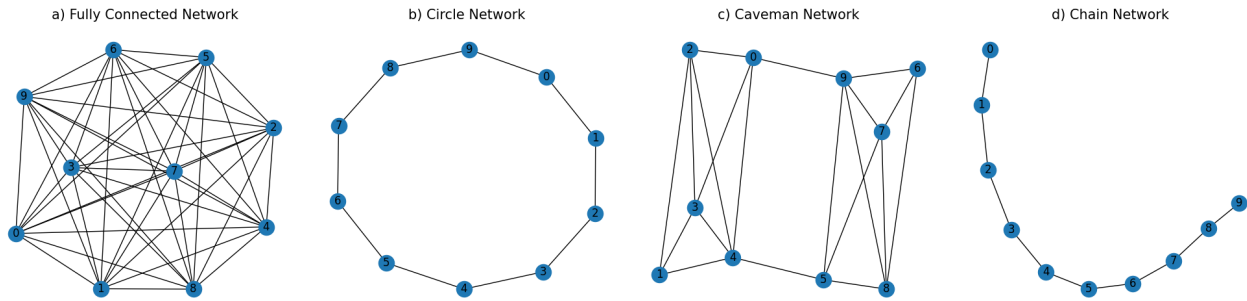
- **Creative:**

"For what follows, pretend that you are a very creative person."

- **NotCreative:**

"For what follows, pretend that you are not a very creative person."

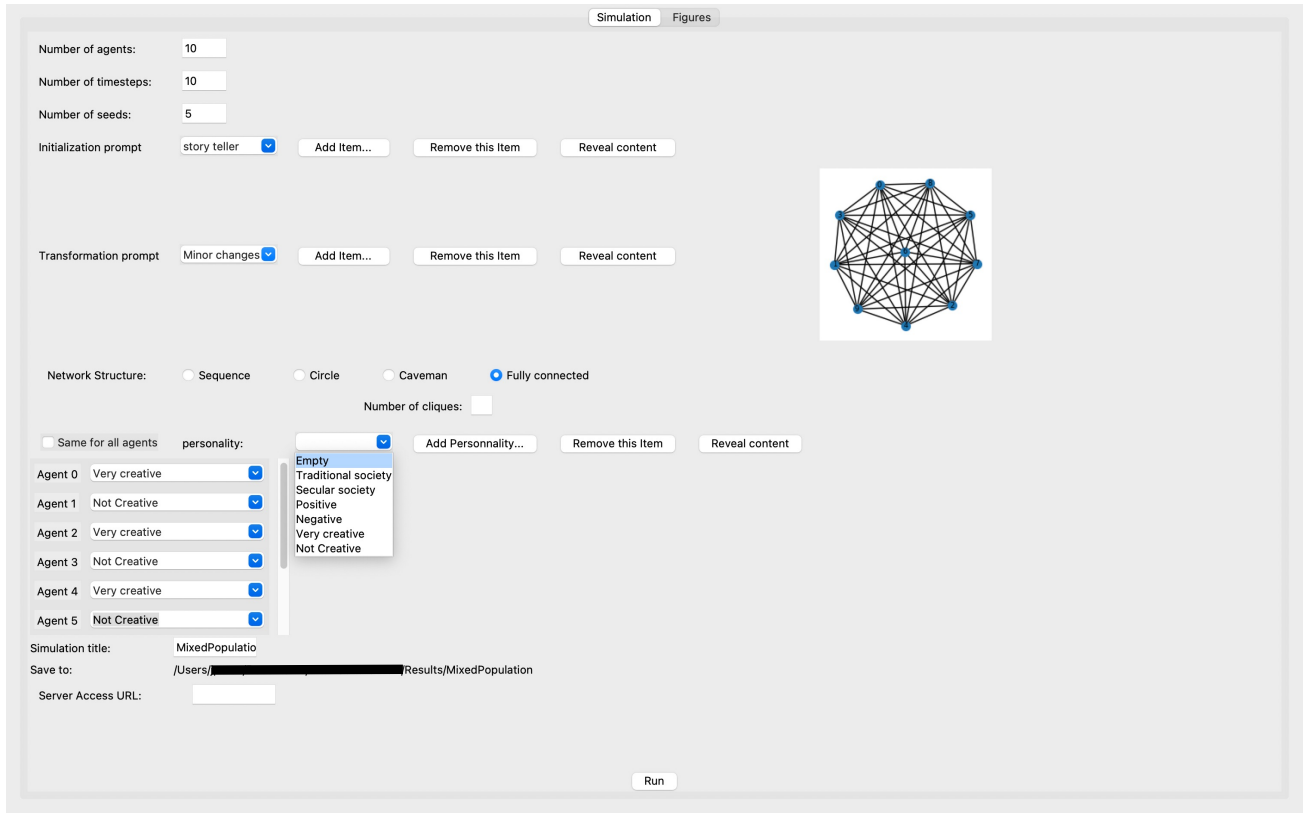
**Network structure** Unless the network structure was the manipulated variable, the default structure was a fully-connected network (Supplementary Fig .1.a). We also used a chain network (Supplementary Fig .1.d) in Section 3.1, as well as circle (Supplementary Fig .1.b) and caveman (Supplementary Fig .1.d) networks in Section 3.2.



**Supplementary Figure 1:** Different types of networks used in our experiments with 10 agents. In a fully connected network (a), every agent is directly connected to every other agent, enabling efficient dissemination of information. In a circular network (b), agents are connected in a circular fashion, thereby forming a closed loop where stories flow sequentially around the circle. The caveman network (c) consists of agents organized into cliques, with our example showcasing two cliques of 5 agents each. In the chain network (d), agents are arranged linearly, and content generation happens sequentially, meaning that agents only generate a story after receiving the story of the previous agent in the chain.

## Appendix B User Interface

To allow researchers to easily use our model, we developed an intuitive user interface that allows to manipulate the variables of interest and generate figures (see Supplementary Figure 2). In the Simulation panel, users can set these variables. Starting from the top of the panel, users can select how many agents they wish to simulate (“Number of agents”), how many generations to simulate (“Number of generations”), and how many times to repeat the simulations (“Number of seeds”). Users can also select which initialization prompt and transformation prompt to use (“Initialization prompt” and “Transformation prompt”). Users may add a new prompts by clicking on the “Add prompt...” button. This opens a new window in which they can set the name and content of the prompt. Users can also select which network structure to use for the simulation. When “sequence” is selected, agents are arranged in a chain, and each agents waits to receive the output of the previous agent in the chain before generating its own output. The other options are “Circle”, which creates a cyclic graph, “Caveman”, which creates a connected-caveman network [48] (in which case the user must also select the number of cliques), and a “Fully-Connected”, which creates a fully-connected network. The structure of the chosen graph is displayed on the right side of the window. Finally, users can also select the personalities to assign to the agents. A checkbox allows to indicate whether all agents should have the same personality. If it is the case, users can select the corresponding personality in the drop-down menu. Like for the prompts, users can add new personalities to the list. If agents have different personalities, the users can assign each agent its personality in the corresponding list. On top of these parameters, users should give a name to the simulation, which will be the name of the folder in which results are saved. Users should also specify the URL to send the request to in order to get answers from the LLM. We generated such an URL using a library from oogabooga (<https://github.com/oobabooga/text-generation-webui>), and provide details of this procedure in the README file of our github repository. Once all of these parameters have been set, users can run the simulations by clicking on “Run”. Once the simulation has terminated, figures will be generated and displayed in the Figures tab of the GUI, as well as being stored in the result folder.



Supplementary Figure 2: Screenshot of the Graphical User Interface

## Appendix C Additional metrics and results

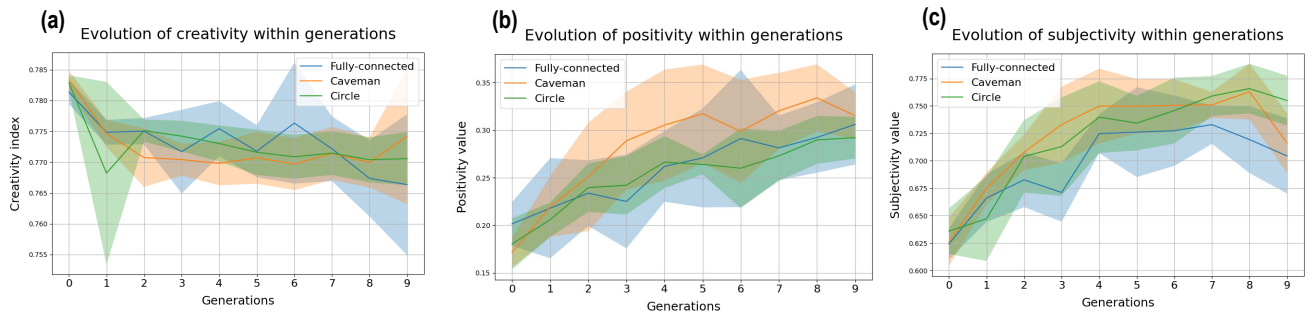
### C.1 Additional metrics

**Creativity** To obtain the creativity of a text, we compute the average semantic distance between all pairs of words in this text. This way of measuring creativity is supported by recent studies [23] [20].

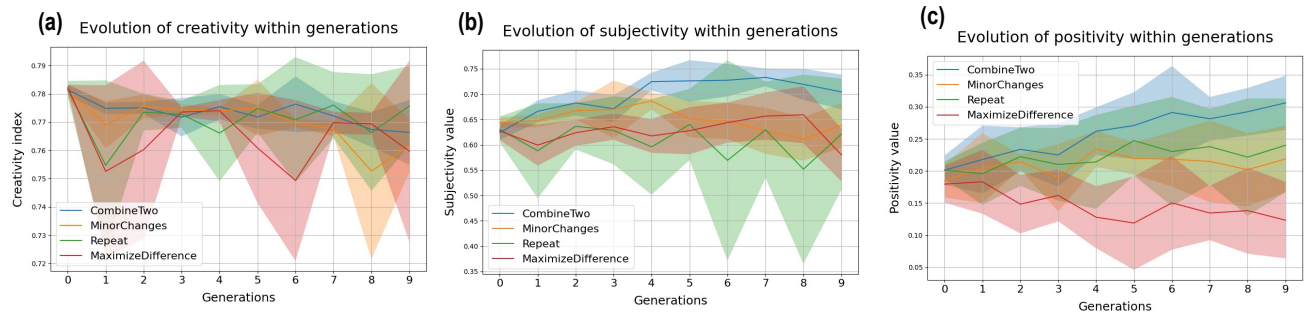
**Positivity and subjectivity** We used TextBlob [29] to perform sentiment analysis on the content generated. In particular, we used their measures of positivity and subjectivity and looked at their evolution through generations.

### C.2 Additional results

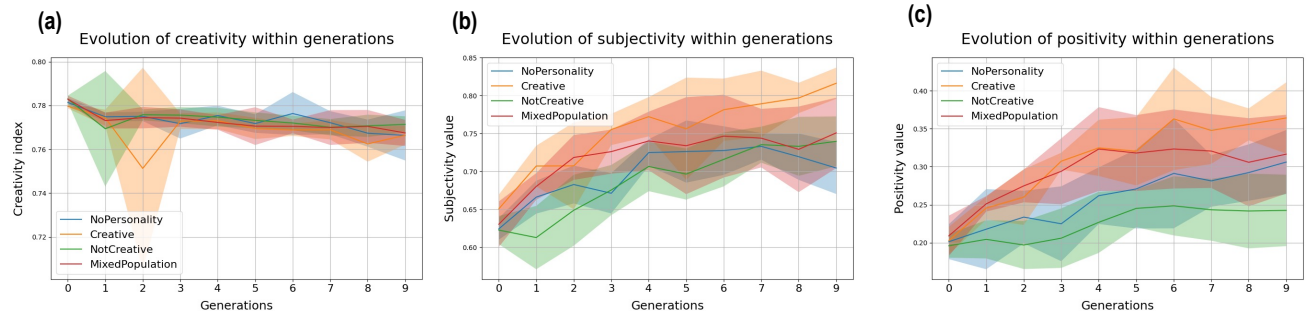
The evolution of those additional metrics through generations is shown in Supplementary Figure 3, Supplementary Figure 4 and Supplementary Figure 5



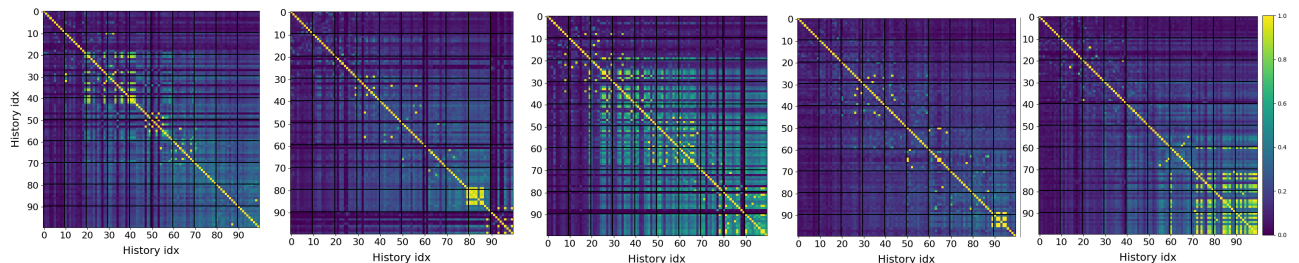
Supplementary Figure 3: Evolution of additional metrics for four different network structures



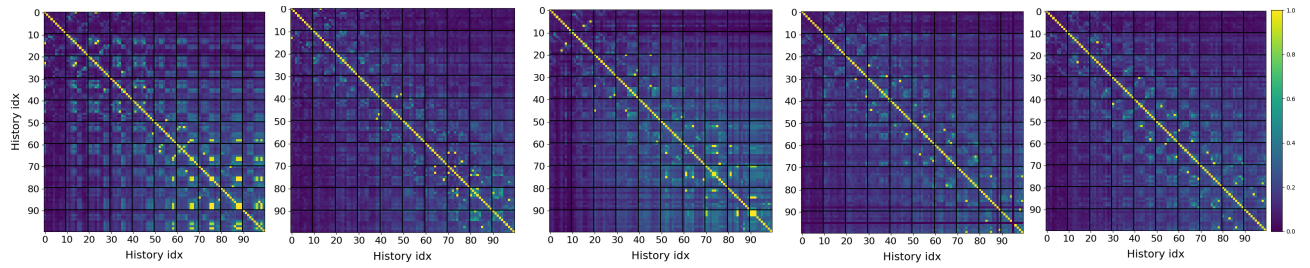
Supplementary Figure 4: Evolution of additional metrics for four different transformation prompts: Combine2, MinorChanges, Repeat, and MaximizeDifference



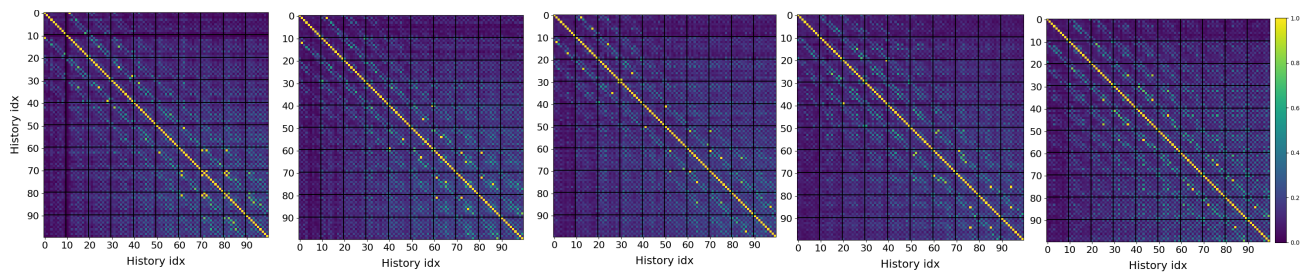
Supplementary Figure 5: Evolution of additional metrics for different personalities: NoPersonality, Creative, NotCreative, Mixed population of Creative and NotCreative



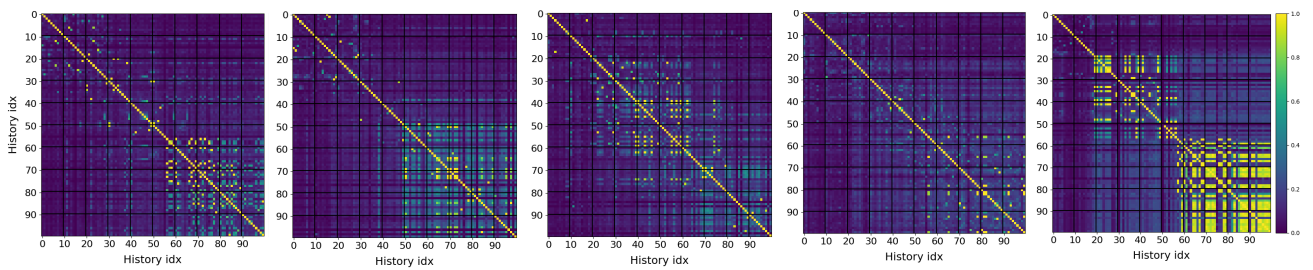
Supplementary Figure 6: All similarity Matrices for the Fully-Connected Network.



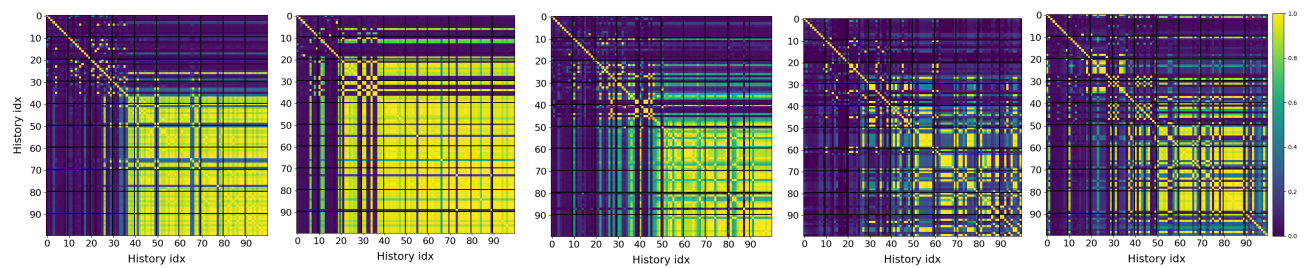
**Supplementary Figure 7:** All similarity Matrices for the Caveman Network



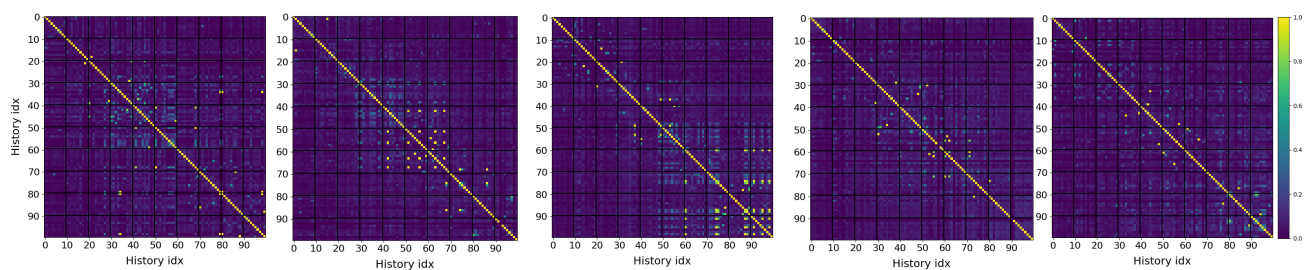
**Supplementary Figure 8:** All similarity Matrices for the Circle Network



**Supplementary Figure 9:** All similarity Matrices for the MinorChanges transformation prompt

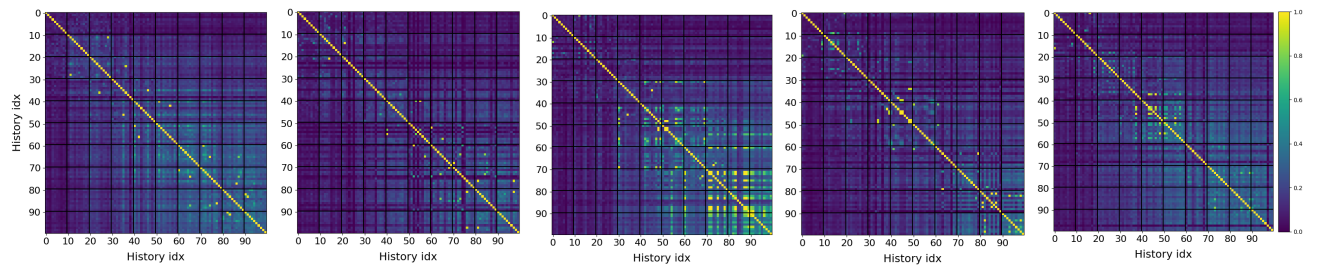


**Supplementary Figure 10:** All similarity Matrices for the Repeat Prompt

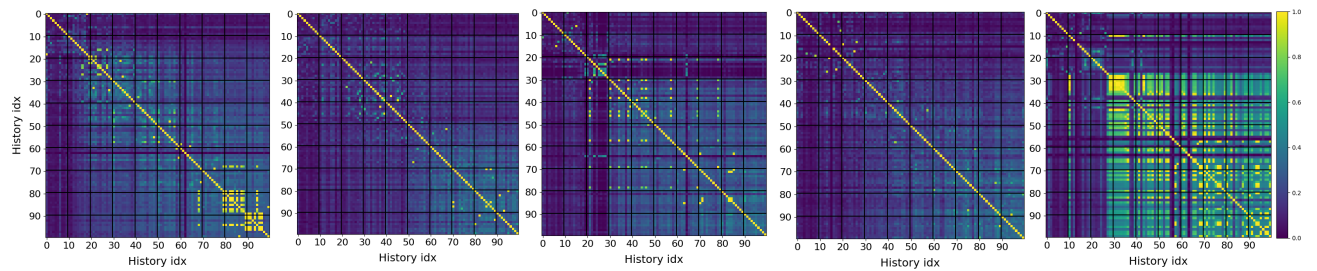


**Supplementary Figure 11:** All similarity Matrices for the MaximizeDifference transformation Prompt

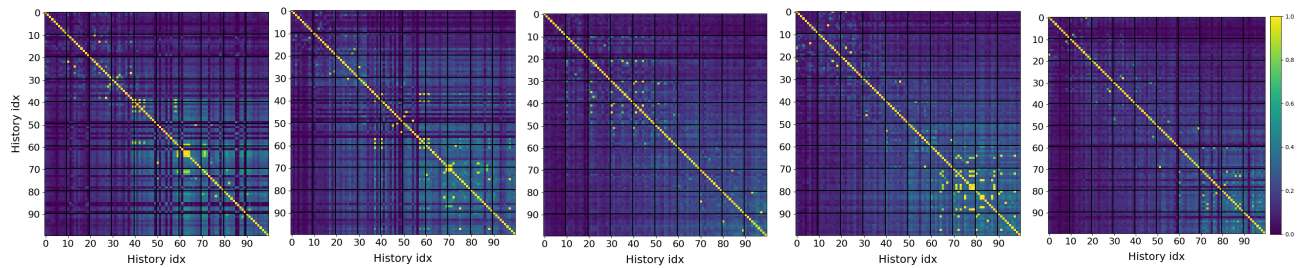




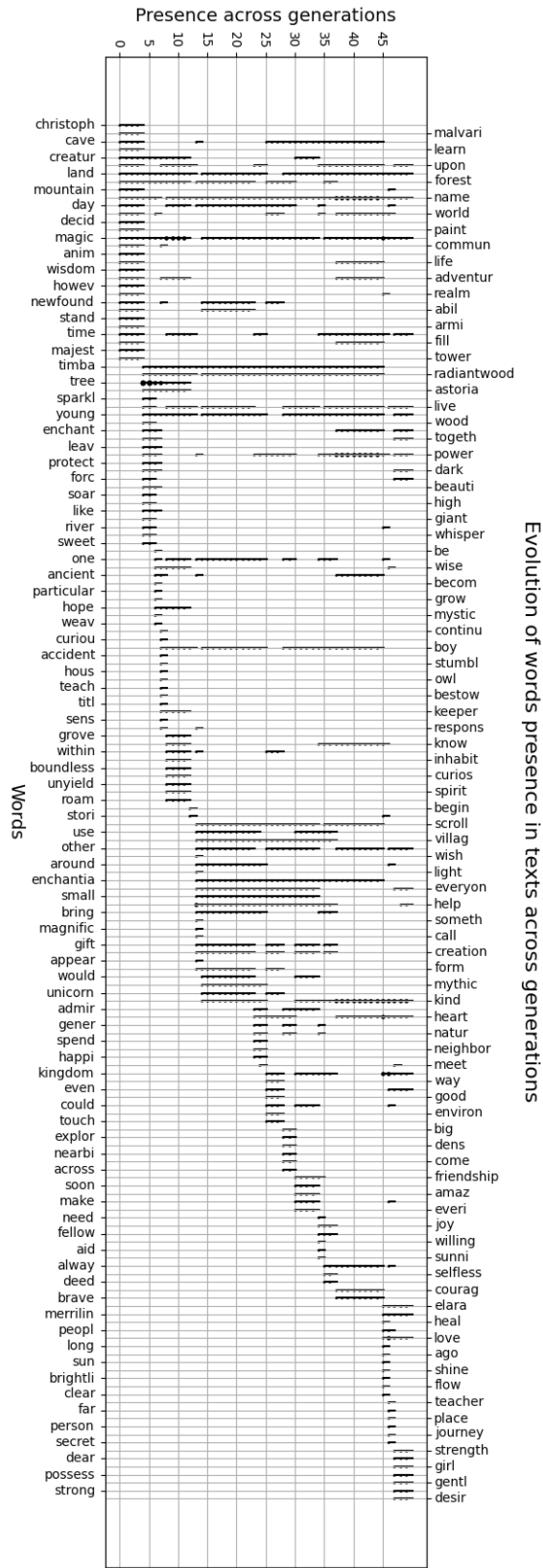
Supplementary Figure 12: All similarity Matrices for the Creative personality



Supplementary Figure 13: All similarity Matrices for the NotCreative personality



Supplementary Figure 14: All similarity Matrices for the Mixed Population



Supplementary Figure 15: Complete rotated word chains plot of fig .2