The Maintenance of Conceptual Spaces Through Social Interactions

Max Peeperkorn

Graduation Thesis
August 2020

Supervisors: Rob Saunders and Oliver Bown

Media Technology MSc Programme
Leiden Institute for Advanced Computer Science
Leiden University, Netherlands

max.peeperkorn@protonmail.com

Abstract
Creativity is a social phenomenon. As individuals share their perspectives, ideas emerge that none could have had on their own. These perspectives are embedded in an individual's conceptual space, which plays a central role in the search for novel ideas and artefacts. In the context of creativity, conceptual spaces are an abstract structure for understanding explorative and transformative creativity, while in a broader cognitive view, they are viewed as geometric structures for organising thought. Building upon those two views, a computational social creativity model is developed, demonstrating the use of Variational Autoencoders for the mechanisation of conceptual spaces. The simulation explores how they are maintained through social interactions and influenced by the individual's preference for novelty and different ideologies of a creative society. Finally, several challenges, ideas and directions for future work are discussed.

Keywords: Computational Social Creativity · Variational Autoencoder · Conceptual Spaces · Systems View of Creativity · Artificial Intelligence · Agent-based Social Simulation

1. Introduction

Creativity is a fundamental aspect of human intelligence, which has frequently been attributed to the individual. The romantic notion of creativity views the individual as the sole instigator of novel ideas and artefacts, and in everyday experiences, it often appears this way. However, we draw inspiration from others, and our imagination is sparked by existing ideas and artefacts (Vygotsky, 2004). When engaging in social activities, such as brainstorm sessions, ideas emerge that none of the individuals could have had on their own. Creativity is not just an individual but a social and cultural phenomenon.

A widely accepted view of creativity in a socio-cultural context is the Systems View of Creativity (Csikszentmihalyi, 1988). The Systems View proposes that creativity can be observed in the interaction between its three forces: the *domain*, the *individual, and* the *field* (Fig.1). The domain is an abstract cultural repository where the existing knowledge is held. The individual produces some variation for the domain based on the existing knowledge, and the field is the social space for the individual and acts as a gatekeeper that selects which variations are worthy of circulation in society and preservation in the domain. In a social context, many individuals share their perspectives on a domain, and while the perspectives can have similarities, it is the differences that are interesting for generating new ideas and artefacts. These perspectives, or styles of thought, are embedded in an individual's conceptual space and play a central role in our search for new ideas (Boden, 1998).

For this research, a model of Computational Social Creativity (Saunders & Bown, 2015) was developed, which uses Variational Autoencoders to model conceptual spaces and investigate how they are maintained through social interactions. In parallel, we investigate the influence on this maintenance for different ideologies of the field and the novelty preference of the individuals.

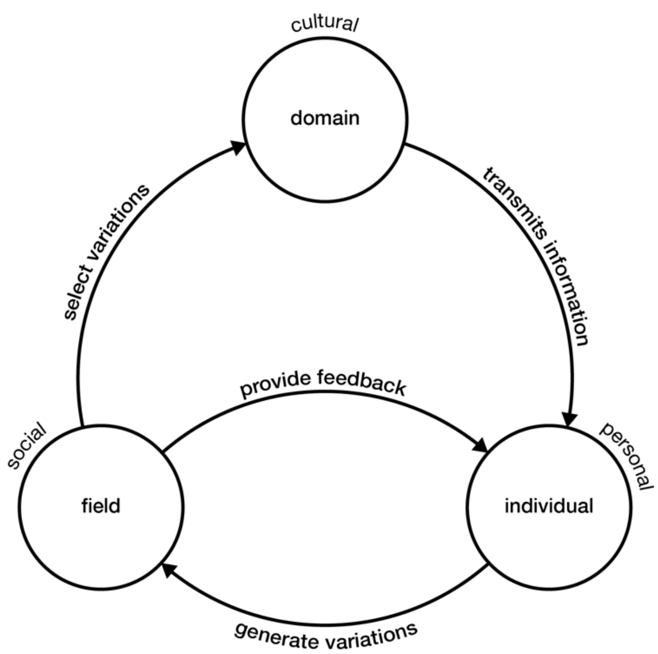

Figure 1 – Csikszentmihalyi's Systems View of Creativity. Creativity is observed in the interactions between the three forces: the domain, the individual, and the field. Illustration adapted from original as seen in Saunders (2012).

2. Conceptual Spaces

In a general cognitive view, conceptual spaces are mental structures that organise thought. Gärdenfors (2000) proposed conceptual spaces as a geometric structure and intended to bridge the symbolic and the sub-symbolic. What conceptual spaces allow us to do is to find similarities between symbols that cannot be derived from the symbolic level. For example, two binary strings that differ only one bit are symbolically similar but can have a very different meaning. Conceptual spaces describe properties and concepts that exist within the domain by assigning different qualities to its dimensions. The geometric structure allows for the organisation of objects in space along several dimensions, identifying similar objects based on the properties and concepts associated within their region in the conceptual space. An example given by Gärdenfors (2000) is the conceptual space of the domain of colour, consisting of three (physiological) dimension: hue, saturation, and brightness (Fig. 2).

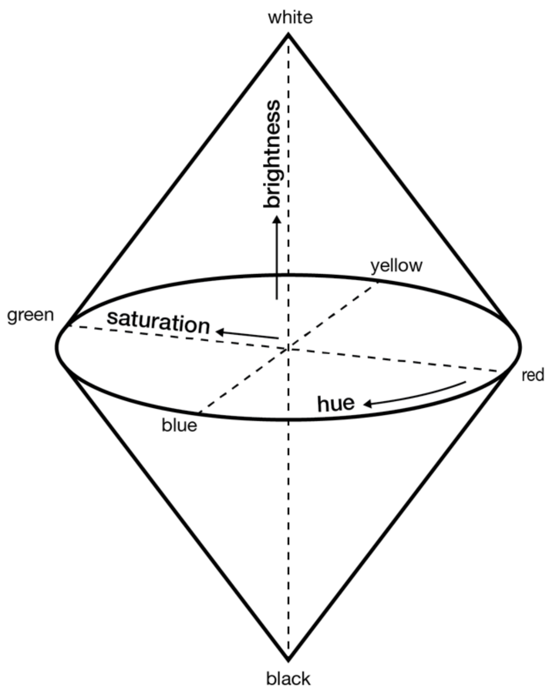

Figure 2 – An example of a geometric conceptual space given by Gärdenfors is the colour spindle for the domain of colour. Illustration adapted from the original as seen in Gärdenfors (2000).

In the context of creativity, Boden's view of conceptual space is well known and is a central aspect of her framework for creativity (2004). However, her exact definition of conceptual spaces remains elusive. Boden argues that conceptual spaces are simply a set of ideas and artefacts that follow the set of rules of a domain. Subsequently, it is mainly used as a metaphor, "map for the mind", to support the latter two of the three modes of creativity: exploration and transformation—the other being combination, which is viewed as a different creative strategy. Explorative creativity is a guided search (i.e. using a map) for novel ideas and artefacts within the known boundaries of the conceptual space. However, sometimes we encounter ideas that are beyond those boundaries. Coming up with these ideas is considered to be transformative, as it requires an update of the set of rules to fit the novel idea.

While conceptual spaces as maps are effective for making a case for exploration and transformation, there are two issues with the definition given by Boden. The first is that exploration and transformation are, to a large extent, interchangeable and can be seen as the same process (Ritchie, 2006). Even in Boden's own words, "artists map and remap their territory as they go" and "mental geography is changeable, unlike terrestrial geography." The difference is that some novel ideas and artefacts have the potential to require one to completely "redraw" the "map" and therefore cause a transformation of the conceptual space. If this is the case, then any artefact can potentially transform a conceptual space. The second problem with Boden's definition is that it is too abstract to be useful for computational purposes. Research has been done to mitigate this problem (Wiggins, 2006a; Wiggins, 2006b; Thornton, 2007) and Forth et al. (2010) propose a unification of Gärdenfors' and Boden's views for concept formation in the domain of music. However, the concern of this research is not with the details within the conceptual space but to take a stance which shares aspects with both views. Taking Boden's view for examining the explorative (and transformative) creative acts, but also use Gärdenfors' geometric structure to traverse the space. This research proposes to use the Variational Autoencoder (VAE) as a means to model conceptual spaces for simulation purposes.

3. Mechanising Conceptual Spaces

Due to their probabilistic nature and their compression and generative capabilities, Variational Autoencoders (Kingma & Welling, 2014) are conceptually a natural fit for mechanising conceptual spaces. A VAE is a self-supervised technique that learns fuzzy relations present in data. In contrast to regular autoencoders, the VAE is a variant that enables the mapping of data onto a smooth latent space–which is reminiscent of Gärdenfors' conceptual space. However, the concepts and properties embedded in VAE cannot be identified. Based on its characteristics, this research assumes that the VAE is an accurate abstraction for their formation.

3.1 Perception, Interpretation, and Production

The conceptual space, as a structure for thought, is involved in the individual's perception, interpretation and production. In people's minds, they describe the artefact according to its characteristics, which are properties and concepts like colour or tone. Then by definition, perception is the process of representing an object in a lower dimension. Analogous to this act of perception, the VAE encodes an artefact (one sample) in its latent spaces. Based on the existing information embedded in the latent space of that domain, it encodes the artefact to its latent representation (Fig. 3a). In line with Gärdenfors' view, two artefacts that are close together in the latent space have similar properties.

Perception and interpretation are closely related, but the difference is that the latter requires an extra effort to learn what the artefact is and what it conveys. In other words, the conceptual space adapts to the relevant information within the artefact. In terms of the VAE, this means that training the VAE on new samples can be considered interpretation, updating the weights to represent the artefacts appropriately in relation to existing the latent space appropriately (Fig. 3b).

Past experiences and existing knowledge have a large influence on the generation of new ideas and the production of artefacts. For this reason, the artefacts that are generated should be in line with the artefacts perceived and interpreted. By decoding a random latent sampling, new artefacts can be produced from the current latent space (Fig. 3c). The production can be influenced by changing the mean and standard deviation when sampling the latent distribution. In this research, the VAE continuously adapts, and therefore, the range of the possible generations should expand as its latent space is explored and shift as it is transformed.

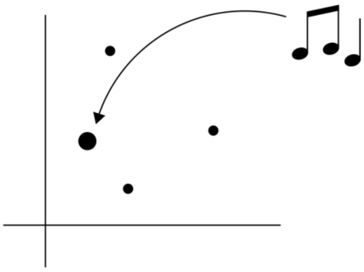

Figure 3a – Perception: Lookup an artefacts position in the latent space

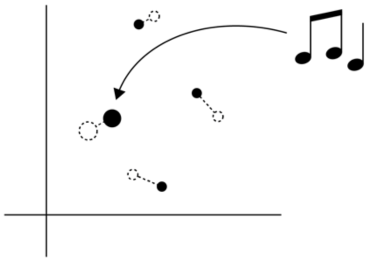

Figure 3b – Interpretation: Update the latent space with the new artefact.

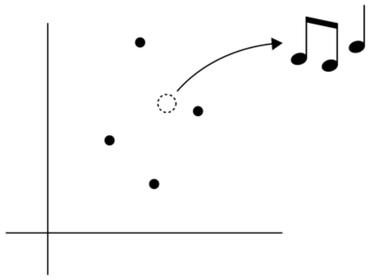

Figure 3c – Production: Generate a new artefact from the latent space.

3.2 Using Variational Autoencoders in Simulations

Modern machine learning algorithms are designed to process large amounts of data in batches. The VAE is no different, usually requiring extensive training to learn any features and to build a smooth and stable latent space. This implies that for the VAE to be useful for social simulations, they need to be correctly initialised and operated at the correct scale.

Agent-based models start from an initial state and observe what happens according to the implemented policies. A VAE cannot just start learning from a random initialisation one artefact at the time. It needs to find relevant features of the domain to produce sensible outcomes, and that requires data and training. This problem is easily overcome by pre-training the VAE's, which is analogous to each individual having a basic education of the domain. However, there is a delicate balance that needs the be considered. Pre-training too much could cause the model to ignore any modelling policies and just use the initially learned variation throughout the simulation. Conversely, training it too little might lead to a random walk, generating artefacts that do not make sense. Within the current research, a good heuristic for determining the balance has not been found. It is likely that the exact balance is dependent on the data and the domain used for the simulation.

In this social simulation, the VAE needs to adapt to new experiences using additional training on a dynamic dataset selected by the implemented policies. As its learning, it is crucial to make sure that the VAE does not forget past experiences or to collapse on a small subset of the domain.

3.3 Data Representation and VAE architecture

The domain used for the simulation defines the required network architecture and type of VAE. This research developed a simplified musical representation of short synthetic sequences of pitches. These short melodies consist of 16 timesteps (1 bar) of 12 possible pitches (chromatic scale). This domain was chosen because the categorical data allows for exact matching at each timestep of the reconstruction.

A simple Variational Recurrent Autoencoder (VRAE) architecture was designed based on the work by Fabius & van Amersfoort (2014). Instead of simple recurrent layers, Long Short-Term Memory (LSTM) layers were used. LSTM networks are a type recurrent neural network, where information in the data is better preserved over time, by explicitly passing the current cell state to the next timestep (Hochreiter & Schmidhuber, 1997). As a result, LSTM's outperform simple recurrent layers and converge faster. The VRAE network has a 32-dimensional latent space and two hidden LSTM layers with 128 nodes each (Fig. 4). Although LSTM's are designed to work with longer sequences, research indicated they their performance for VRAE's is limited, and a method for improving the performance of the VRAE is to anneal the Kullback-Leibler (KL) term, which penalises the VRAE so that it maintains a smooth latent space, in the earlier stages of training (Bowman et al., 2016). This technique allows the VRAE to extract more informative features early on before it is penalised by the full KL divergence term to smooth the latent encodings. Although the sequences used in the current domain are still considered short, an improvement in the latent distributions was observed. The KL divergence term is annealed using a logistic function over the first 25 epochs. A lower dimensional latent space would have been preferable (two dimensions would be ideal for visualisations), but it proved hard to train, even with the addition of KL annealing.

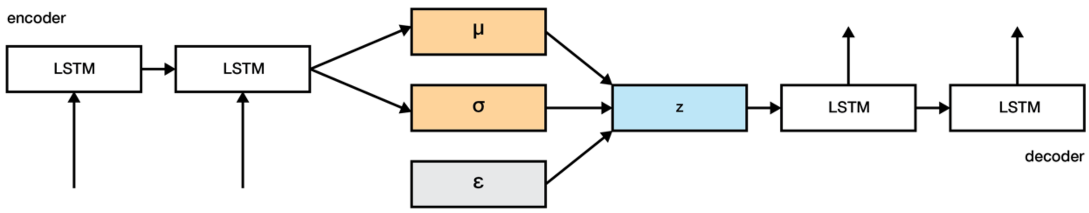

Figure 4 — The Variational Recurrent Autoencoder architecture. The LSTM layers consist of 128 nodes each, with two parallel fully-connected linear layers (the orange blocks) consisting of 32 nodes each, providing the means and standard deviations to sample the latent encoding (the blue block) for each artefact.

4. The Simulation

The simulation is structured according to Csikszentmihalyi's Systems View of Creativity (1988), which lends itself well as a basis for modelling societies of creative agents and has been proposed (Liu, 2000; Saunders, 2012) and used for such modelling experiments in the past (Hanna, 2005; Saunders, 2011). In previous models, the domain has been modelled as a repository holding the artefacts selected by the field. Although the agents keep track of what artefacts they have created in the current simulation, those repositories are not where the knowledge of the domain is held. The variations are produced by sampling the VRAE's, implying that the knowledge of the domain is actually held within the conceptual spaces of the agents. The domain is therefore distributed across the agent's conceptual spaces and updated through additional training.

Every agent has its own unique conceptual space, and as a result, the produced variations cannot be compared on the individual level. A 'global' VRAE is introduced to provide a latent space in which the artefacts can be compared. This way, it serves as an Archimedean point that enables the analysis of the distributed domain. The global VRAE operates outside the scope of the simulation and is not trained during simulation, establishing a stable overview of the distributed domain as a whole upfront.

Similarly, a common ground is required during the simulation, it needs to be determined which agents are allowed to share their artefacts, and the field needs information of the domain to enforce its ideology. For this purpose, a recommender system is introduced within the scope of the simulation. Recommender systems, which are prevalent in many online consumer services and platforms, provide suggestions, and other objects of interest, based on the users' preferences. Due to this characteristic, the recommender system serves as a matchmaker between the agents and proxy of the socio-cultural gatekeepers (e.g. galleries and professional critics) of the field. By using a shortcut to the global VRAE, it can determine every agent's current location based on what they produced and if required, retrieve the relevant information required to apply the fields' ideology. The agents do not have direct access to the global VRAE.

4.1 Establishing the domain: Initialising the Variational Autoencoders

While the agents' VRAE's are kept simple, their initial training requires a specific set-up. A possible strategy might be to generate the dataset and uniformly divide into a subset for each agent. The global VRAE is then trained on the whole dataset. Uniform sampling of the domain, however, would result in each agent having a similar view of the domain. While this might be a valid choice for some investigations, this research investigates how the latent spaces are maintained over time, to allow for this observation, each agent in the simulation requires a unique starting 'locations' in the domain. Instead of subdividing the dataset randomly, each sample in the dataset is assigned to a slice for every individual according to its latent encoding in the global VRAE (Fig. 5a). While the slices appear to be hard cuts, the data itself is much blurrier. The individuals are trained on their respective slice. After a sampling (with a small standard deviation of 0.25) of their latent space reveals and verifies their unique starting locations (Fig. 5b).

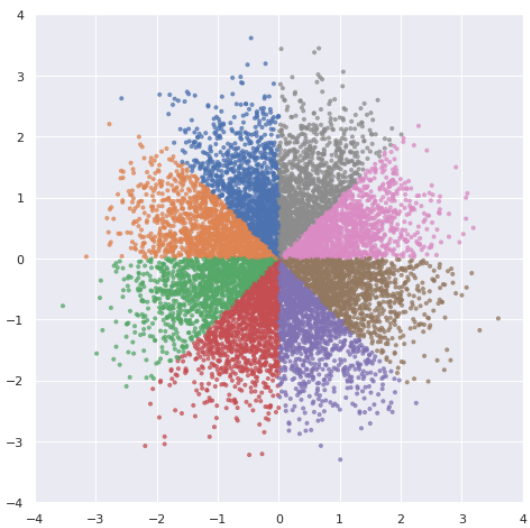

Figure 5a – After constructing the Global VRAE's latent space, all artefacts are encoded and project on two-dimensions using PCA. This projection is then used divided the training data amongst the (eight) agents.

Figure 5b – After pretraining the agents, a sampling of their latent spaces and projected on two dimensions using t-SNE (van der Maaten & Hinton, 2008) in the global VRAE's latent space, clearly reveal clusters–that indicate the unique starting 'locations' of the agents.

4.2 Ideologies of the Field

The field acts as a gatekeeper for what artworks are selected for circulation, according to the ideology of the society (Csikszentmihalyi, 1988). Different ideologies use different selection criteria and subsequently influences the social interactions taking place in the domain. For example, a society might prefer artefacts that are similar to what is currently in circulation or perhaps it only wants to display artefacts of certain style. In the simulation, ideologies are modelled by determining the selection probability for each artefact available for circulation. This research proposes three different methods for calculating these probabilities each resembling a different ideology.

The first is a field without ideology, in which every artefact has an equal chance of being selected for circulation. This uniform approach is a neutral ideology, which also makes it suitable as a baseline for evaluation. The second is a conservative ideology that favours recently selected artefacts. A well-known heuristic that models this behaviour is frecency, a portmanteau of frequency and recency. With every round, the frecency value is decreased according to a simple exponential decay function (Eq. 1). Conversely, each time an artefact is selected by the field, its frecency value is increased by 1. The frecency value is clamped to a minimum of 0.01, to prevent any artefacts from being discarded entirely. New artefacts are initialised with a frecency value of 10.

$$N(t) = \mathrm{N} \cdot e^{-\lambda t}, \text{ where } \lambda = 0.1 \quad (1)$$

Finally, the third ideology is based on the density of a specific area in the global VAE's latent space. This ideology favours artefacts that exist in less explored areas of the domain. This progressive ideology pushes for novelty on the field level. Compared to the first two ideologies, this ideology utilises the current state of the domain, via the recommender systems which retrieve the relevant information from the global VRAE. Artefacts that exist in a low-density area (with few artefacts in the neighbourhood) have a higher chance of being selected than artefacts in a high-density area. The density is calculated using Kernel Density Estimation (KDE), which estimates the probability density distribution of the latent space, based on the artefacts produced so far. Subsequently, the density of the area surrounding each artefact is approximated. The probability density distribution is recalculated every round as new artefacts are introduced. For simplicity, the bandwidth for KDE is determined using Gaussian approximation (Silverman, 1986)

4.3 An Individual's Preference for Novelty

Due to the way the domain and each agent were established starting out at different locations, a natural diversity emerges that in itself could lead to creative behaviour. Hanna (2005) developed a model for creativity, demonstrating that novelty is not a requirement for the emergence of creative behaviours and that the diversity of the individual is enough. However, for the purposes of this research, it is still relevant to investigate the effect of a preference for novelty. If we consider Boden's map of the mind metaphor, then Martindale's (1990) alternative definition of novelty as the disruption of expectedness, is a good basis for modelling novelty preferences of individuals. VRAE's build normally distributed latent spaces as enforced by the KL term. As a result, most information is kept near the origin. By moving away from the origin, it can be expected that more *unexpected* artefacts are generated, because there is less information in those regions. It is proposed that by setting different standard deviations, we can model the individual's preference for novelty, a lower standard deviation makes it more likely the agents will have a conservative approach and work with what it knows best. The VRAE imposes a natural limit of one standard deviation, and for the experiment, three values (0.25, 0.5, and 0.75) were selected to explore the effect of the novelty preference.

4.5 Completing the Model

After initialising the global VRAE and providing the agents with their basic education and starting position, the simulation starts a loop (Fig. 6) iterating over three stages for a given number of rounds. Each stage resembles a force in the systems view of creativity, the domain, the individual and the field, respectively. At the start of the simulation from each agent's VRAE, a number of artefacts are sampled from their respective latent spaces, which serve as the initial selection of the field. Each round iterates over the following steps:

1. The agent receives *m* selected artefacts from the field. Each agent then trains their conceptual space for the given number of epochs (the budget) to extract the new artefacts' features. After this, *n* artefacts are constructed from their updated latent space using a given novelty preference as the standard deviation for sampling.
2. For each agent, the recommender system retrieves the latent encodings for each agents' *n* artefacts and uses the mean of those encodings for the current position.
3. The *k* nearest neighbours are selected to share their artefacts with the current agent to form a pool of artefacts. Subsequently, for each agent, the field selects *m* artefacts from their pool for use at step 1 in the next round. The artefacts are selected using the probabilities depending on the field's ideology.

The above steps tie together the mechanisms of the domain, the individuals and the field. With each iteration, the agents adapt their conceptual spaces and produce new variations that eventually spread through the small creative society according to the fields' ideology, expanding and updating the domain.

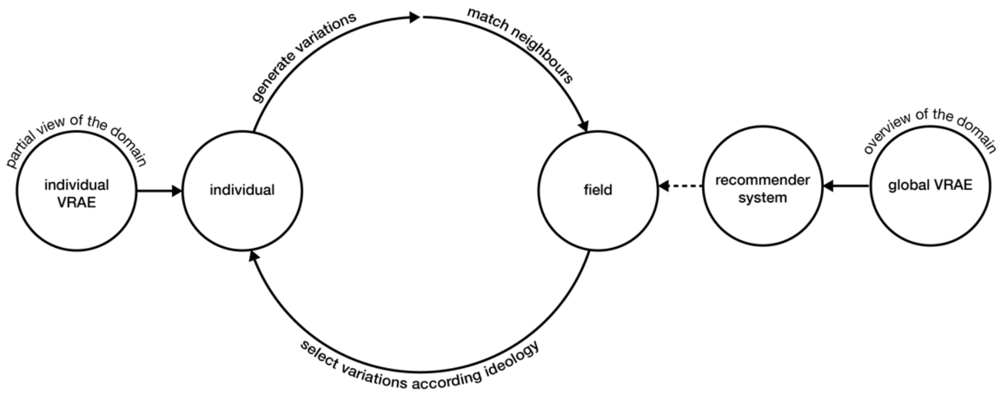

Figure 6 — Illustration of the simulation loop. Each individual agent only has access to its own VRAE, a partial view of the domain. The field uses the recommender system as a shortcut to the global VRAE to match the neighbours and select variation according to its ideology.

5. Results

Besides the evaluation of the VRAE's, this research has set up two experiments, consisting of several runs with the default settings (Table 1) investigating the ideologies of the field and the novelty preference of the individuals. All simulations are run for 250 rounds and with eight agents, selecting one neighbour each round. It is important to keep the number of neighbours in proportion to the number of agents relatively low. Otherwise, too much information is shared amongst all agents, preventing the emergence of interesting behaviours. The agents were pre-trained on a dataset of roughly 1250 samples (as a result of the dataset subdivision described in section 4.2) for 250 epochs, providing a basis for the agents and resulting in average reconstruction error of 0.79 and average accuracy of 70% on reconstructing the training sequences. Global VRAE is trained on a synthetic dataset of 10000 samples for 1000 epochs because its goal is to establish an overview of the domain. It achieved close to 87% accuracy and a reconstruction error 0.38, which appears to be the upper limit for the current VRAE architecture and synthetic data. Training beyond this point did not provide any increase in performance.

Parameters	*Options*
Rounds	250
Number of Agents	8
Number of Artefacts Selected	64
Number of Artefacts Produced	4
Number of Neighbours	1
Pre-training Epochs Individuals	250
KL Annealing Epochs	25
Budget	5
Field Ideology	**Uniform** | Density | Frecency
Novelty Preference	**0.25** | 0.5 | 0.75

Table 1 — The model's parameters and options used in the experiments. Where applicable, default values are in bold.

5.1 Evaluation

In the initial set-up of the simulation, the intent was to adapt the conceptual space with ten artefacts each round. As mentioned earlier, the choice of categorical data was initially chosen because of possibility to check for exact matches with the original and the reconstruction, and therefore, can be used as a training target for the VRAE. With this mechanism, it would train until all artefacts were perfectly reconstructed, or the training budget of 100 epochs was consumed.

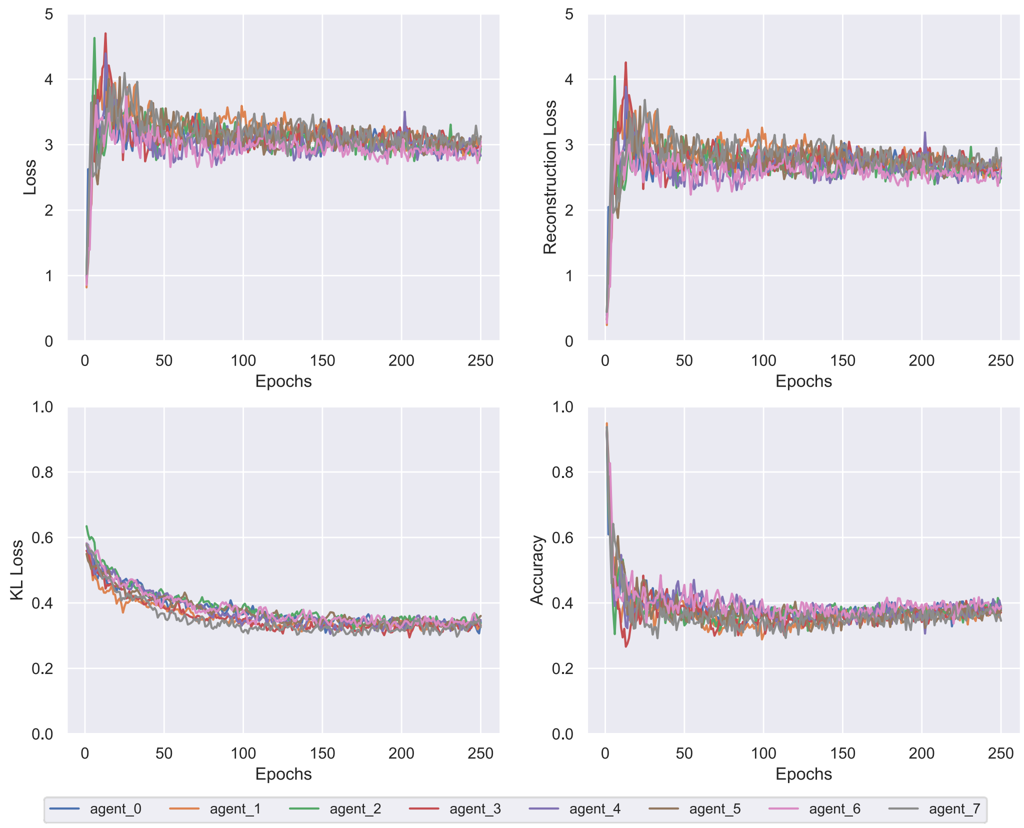

Figure 7 – The initial set-up of the simulation selected only ten artefacts each round. Looking at the immediate rise of the reconstruction error and the low accuracy, the VAE quickly destabilises. As a result, the agents do not learn to incorporate the information provided through social interaction with this configuration.

This set-up is conceptually interesting but resulted in unstable VRAE's (Fig. 7). Unsurprisingly, it overfitted on those few samples every round, causing the VRAE to lose most of its initially embedded information. In hindsight, the budget of 100 epochs is obviously too high and the main cause of deterioration due to heavy overfitting. Unless the presented artefacts were already very close to the current latent space of VRAE, it would repeatedly shift the weights too much towards the new artefacts, losing any general knowledge.

In order to resolve this problem, the decision was made to operate the simulation on a larger scale, and training target mechanism was removed. The number of selected artefacts was increased to 64, and the budget was lowered to 5. To allow the model to adapt and to maintain a stable latent space, this also meant that a larger number of new samples needed to be generated each round to keep a balance between new and old information, for this reason, the number of artefacts produced was increased to four. These numbers are based on observations during development and several test runs. From these tests, it was also observed that higher values appear to work equally well, but it requires exponential greater computational resources. Due to the number of simulations to be run, the current settings were chosen. The evaluation of the VRAE's is done on all artefacts stored in each agent's repository. If the agent adapts well to the selected artefacts, a rise in accuracy and drop in the reconstruction error should be observed. If the outcome moves in the opposite direction, the past experiences are not preserved, and the VRAE becomes unstable.

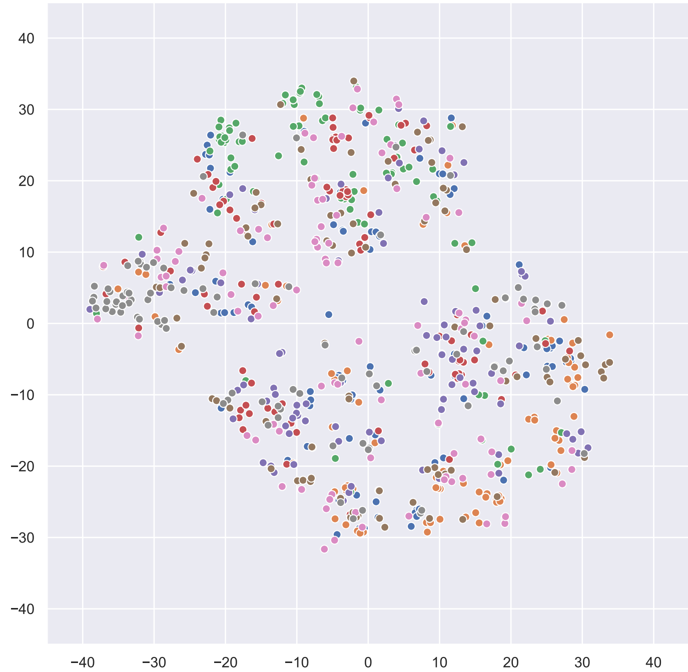

Figure 8 – After the simulation, a sampling of each agents' latent spaces projected on two-dimensions using t-SNE (van der Maaten & Hinton, 2008), suggests the conceptual space have adapted according to the social interactions.

The evaluation shows, after an initial rise in the loss, a stable performance. The results show that VRAE's have consistent behaviour for each agent, and as expected, with small differences as a result of the fields' policies (Fig. 8). Given the agents starting positions (Fig. 5b), they now cover a wider area of the domain. Using t-SNE (van der Maaten & Hinton, 2008), figure 7 shows a two-dimensional projection of a random sampling of the agent's latent spaces after the simulation. Suggesting that the agents adapt their conceptual spaces as a result of the social interactions (Fig. 8). A minor detail in the evaluation is the KL loss, which is going down–which this is not necessarily a problem. The KL loss only drops a slow rate, and the other metrics behave as expected. It should be pointed out that if this were to drop to 0, the VRAE's suffer from posterior collapse–which results in the loss of information in the latent space, and subsequently the VRAE starts to produce random artefacts.

5.2 Experiment 1: Ideologies

The first experiment is aimed at investigating how different ideologies for selecting artefacts affect how the agents behave and how their conceptual spaces are maintained. The experiment is run using the default settings (Table 1) and a novelty preference for the individuals set at 0.25. Figure 9 shows the influence of the different ideologies on social interactions. The ideology using the density policy displays the most distributed communication, while the uniform ideology appears to favour existing relations even though there is no clear clique formation. Interestingly, the conservative ideology using the frecency heuristic has collapsed onto a single agent. This agent (zero) appears to become very popular with its peers towards the end of the simulation.

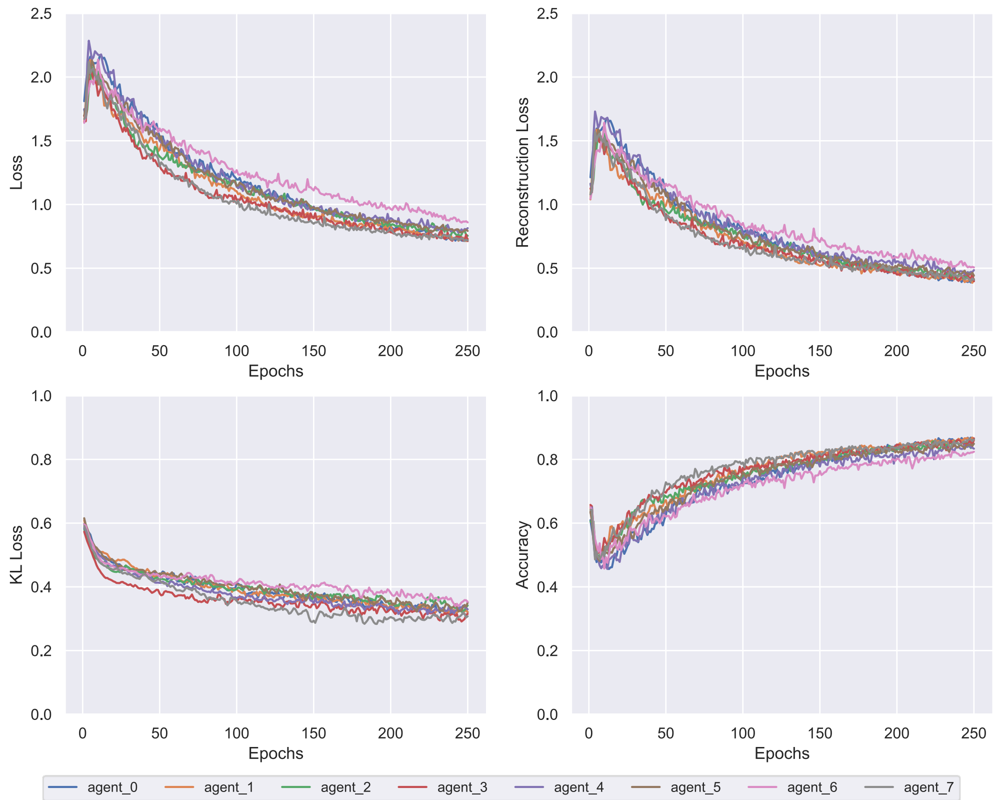

Figure 9 —The performance of the VRAE's during the simulation is shown. The model behaves as expected and similar to normal training. Every agent adapts well to artefacts selected by the fields' policies.

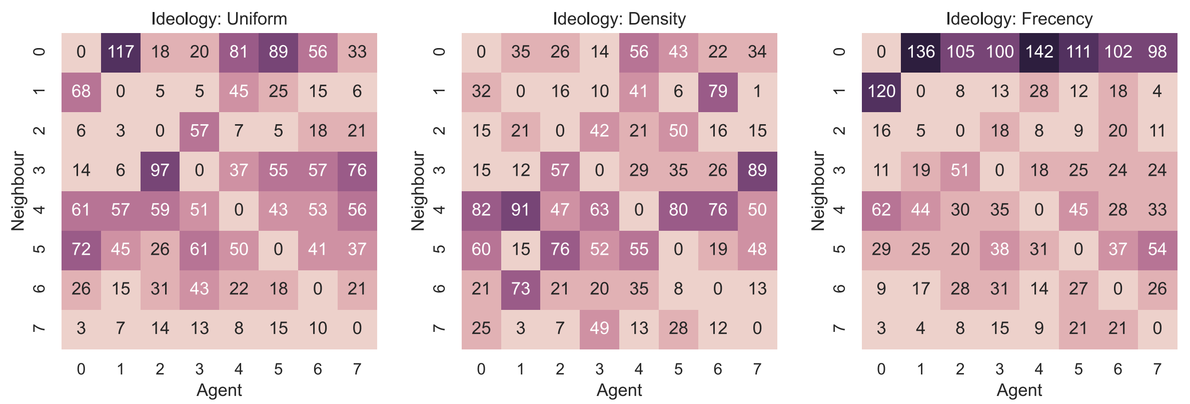

Figure 10 – The communication matrices indicate the number of interactions at the end of the simulation for different ideology settings. The ideology using density shows the most distributed interaction, likely because it favours artefacts in less explored areas. The frecency ideology collapses and favours one agent for their interactions.

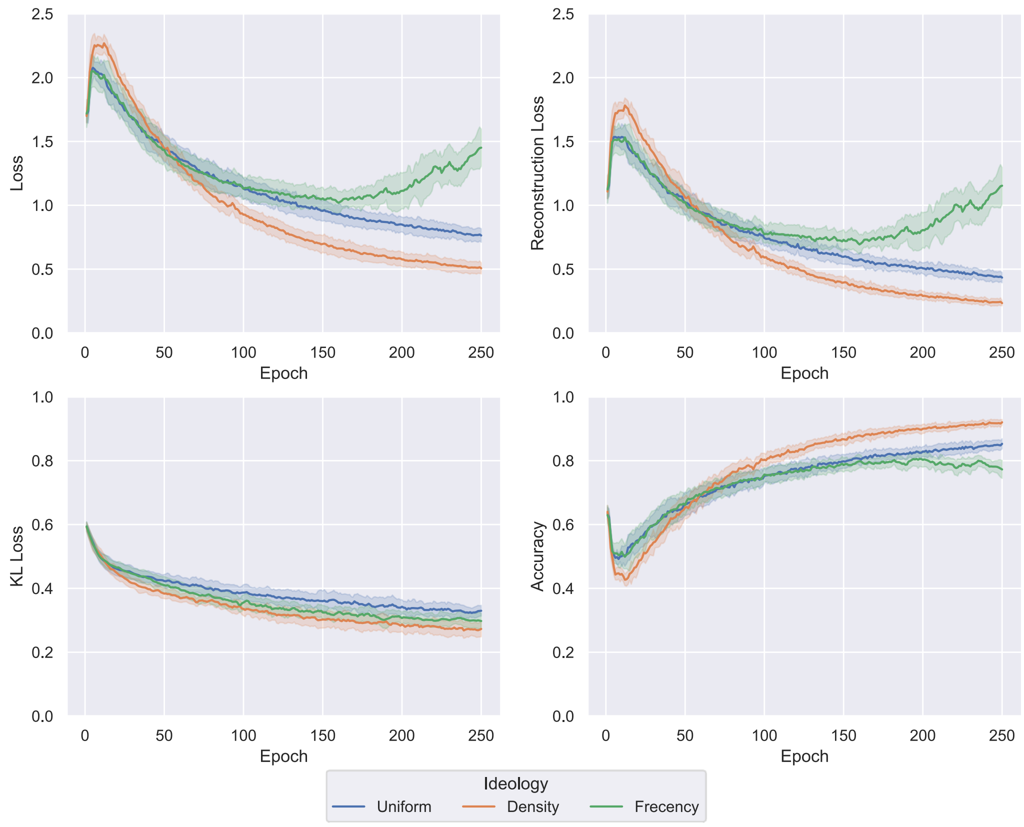

Figure 11 – The ideology influences how the VRAE's are performing. In comparison to the neutral uniform ideology, we see that the density ideology yields the best performing conceptual space, while the frecency ideology causes a collapse.

Looking at the performance of the VRAE's (Fig. 11), the results align with outcomes of the communication matrices. The ideology using the density measure achieves the best results because it favours artefacts which are considered novel by the field. As a result, it introduced more variation and diversity in the selection, leading to better maintenance overall of the conceptual space compared to the other ideologies. The field with the uniform ideology, is in the middle as expected, fulfilling its neutral position. With frecency ideology, it clearly starts to deteriorate around epoch 175, and the agents are unable to maintain useful conceptual spaces causing the collapse observed in the respective communication matrix. While it is likely that this ideology would move towards a smaller subset of the domain, such a strong effect was not expected. It is possible that the exponential decay function is too weak. Another reason could be the frecency value assigned to new artefacts is too low, and therefore, artefacts might not be given a chance to be selected for circulation.

5.3 Experiment 2: Novelty Preferences

The second experiment focussed on the influence of different individual's novelty preference on the latent spaces. By changing their ability to producing artefacts at the edge of their current knowledge, the aim is to see if a higher novelty preference leads to further exploration. For these experiments, the uniform ideology is used because of its neutral characteristics. Figure 12 shows that a higher novelty preference leads to a VRAE with a lower performance. The result is not unexpected, because using wider sampling yields more variation and provides the VRAE's with more variation, ultimately exploring more of the domain as a whole. What is encouraging to see is that the metrics stabilise and do not deteriorate, albeit with different performances. The communication matrices (Fig. 13) draws a similar picture, a higher novelty preference leads to more distributed interactions between the agents, while at lower novelty preference, the evidence shows that some agents favour one or two agents. Likely because the agents push less towards the edges of their latent spaces, and therefore explore less of the entire domain.

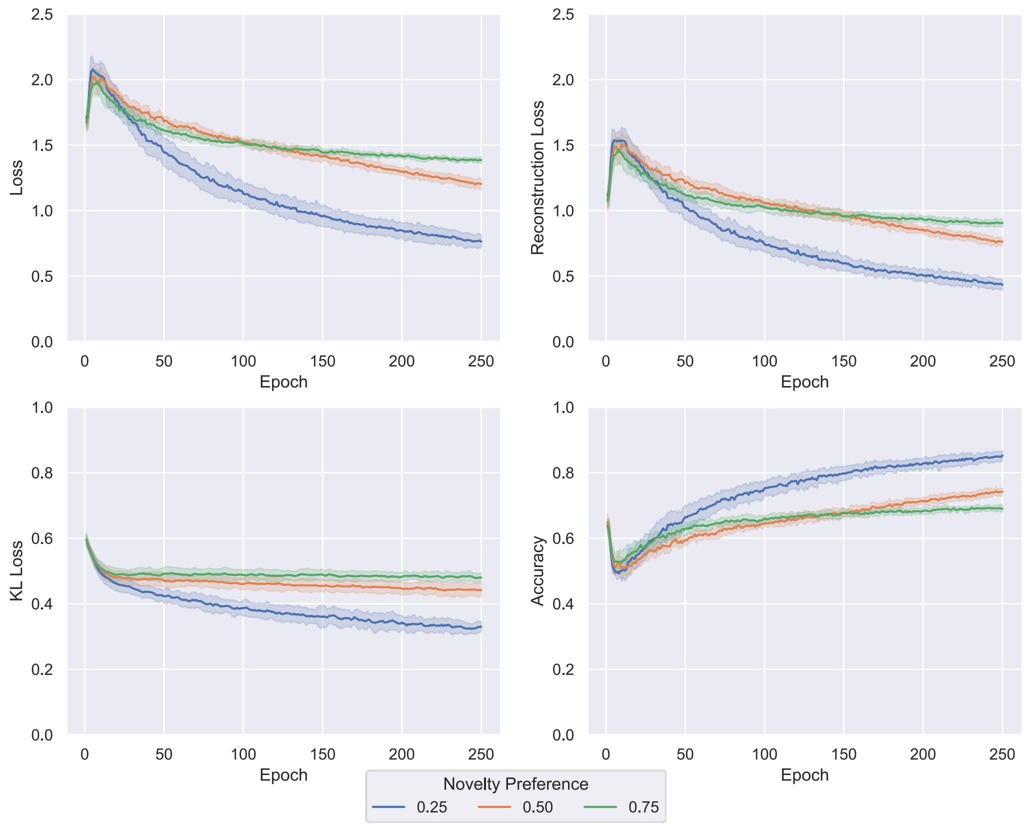

Figure 12 – Different novelty preferences lead to different maintenance stabilities of the VAE's. A higher preference for novelty leads to a wider sampling of the agent's latent space. As a result, the agents are exposed to a wider range of artefacts.

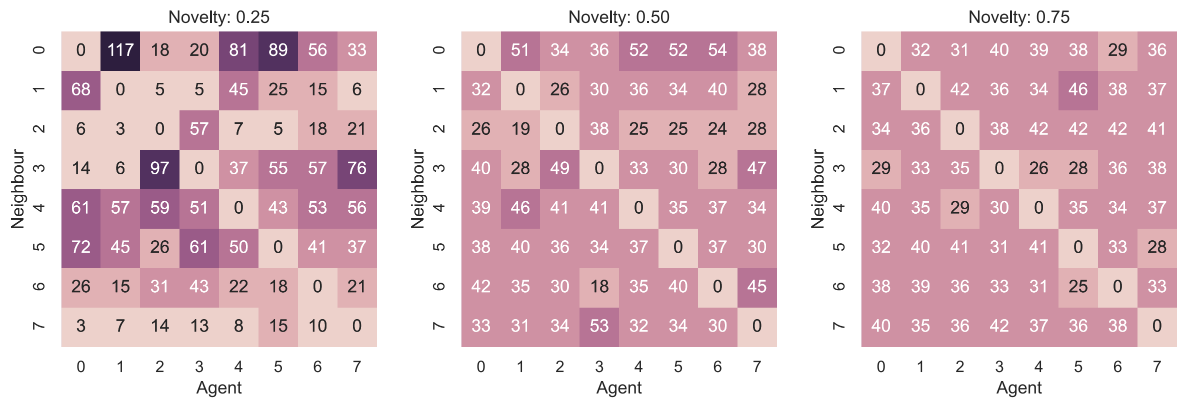

Figure 13 – The communication matrices indicate the number of interactions at the end of the simulation for the different novelty preferences. With a higher novelty preference, the agents have more distributed interactions, as they are able to explore the edges of their latent spaces.

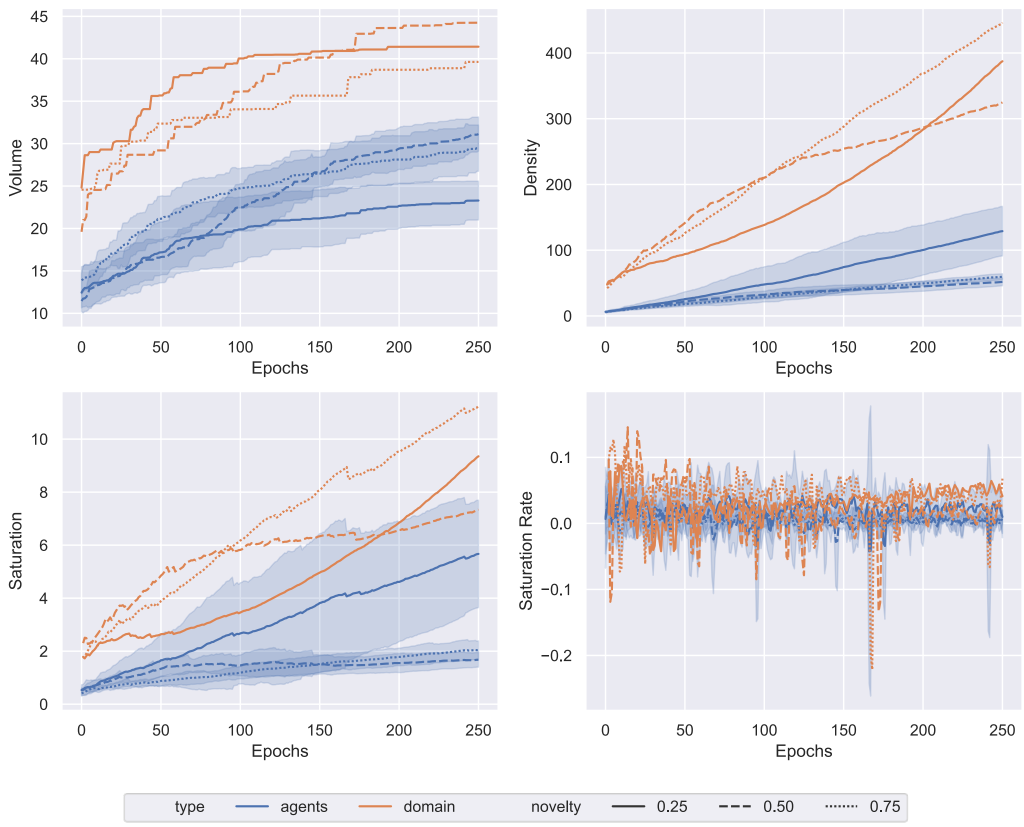

Figure 14 – The results of the saturation measure. Originally, this measure was intended to be included in the evaluation but proved hard to draw conclusions from in its current state. The saturation is the density in proportion to the volume of the conceptual space. The saturation rate is the gradient of the saturation.

6. Conclusion

This research demonstrates the utility of Variational Autoencoders as a computational model for conceptual spaces, both for use in simulations and as a tool for analysis. In the process, the limitations and solutions have been identified. Additionally, three different ideologies and individual's preference and its effects on the maintenance of conceptual spaces have been explored, and the results display the alignment of the VAE evaluations and corresponding social interactions. Concluding that the use of VAE's as conceptual spaces is a viable technique for social simulations and provides new opportunities to investigate creative behaviours and social interactions. This contribution suggests the possibility of evaluating creative output on learned relations and without the use of predetermined rules. It could provide opportunities to work with ill-defined domains where the rules cannot be easily described.

7. Discussion

As with any modelling experiment, many hyperparameters were introduced that can be tweaked and tuned to investigate how the model works, and there are numerous experiments that can be run, and many aspects left to explore and verify.

The research suggests that VAE's adapt as a result of the social interactions between the individuals but has not included a statistical approach to verify the correlation and does not explore to what degree maintenance of the conceptual space is stable. More importantly, how social interactions influence the outcome is not investigated, and this is identified as an essential candidate for future research. Likewise, the degree of influence of the field's ideology and the individual's novelty preference can be further explored to gain insights into the exact causes of the current outcomes. Other elements of the current model that are interesting to explore are the effects of sharing with multiple neighbours, using a larger population or investigate what happens over longer simulation runs. With those possibilities, it cannot be stressed enough that finding the correct balance of VAE architecture, training, selection, and production, before and during the simulation, is essential for success.

By using VAE's as a conceptual space, the geometric space provides opportunities for additional measures to be explored. One of those other measures could be saturation. During development, this research planned to use this measure, which is calculated from the volume of a convex hull in proportion to the density, but it was found that we could not conclude anything from these outcomes yet (Fig. 14). At this moment, individuals only add more artefacts, and no information is discarded. As a result, the saturation would always rise. A way to resolve this is to implement an evolutionary approach, and it might lead to additional interesting research opportunities. The cultural evolution perspective is relevant to explore and could build upon the work of Boyd & Richerson (1985). It could be interesting to investigate the emergence of the ratcheting effect (Tomasello, 1999), or paradigm shifts (Kuhn, 1970) as the conceptual space of the individual adapts to new situations over time. Additionally, the introduction of an evolutionary mechanism which would allow for the dynamic of "young geniuses and old masters", a theory proposed by Galenson (2009). This could be interesting in light of clique formation, which can be likened to art movements or scientific revolutions. It is expected that this would introduce new perspectives every generation less biased by previously generated artefacts and push the conceptual spaces in different directions.

Finally, the current simulation is also a demonstration of training several generative networks together using a dynamic dataset based on social policies. While the network set-up was very simple, training networks using social interactions could be an interesting research goal. Humans get their best ideas by sharing with and learning from others. Perhaps creativity in machines can also benefit from such interactions.

References

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial intelligence*, *103*(1-2), 347-356.

Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*. Routledge.

Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A. M., Jozefowicz, R., & Bengio, S. (2016). Generating sentences from a continuous space. *arXiv preprint arXiv:1511.06349*.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process.* University of Chicago Press.

Csikszentmihalyi, M. (1988). Society, culture, and person: a systems view of creativity, in R. J. Sternberg (ed.), *The Nature of Creativity*, Cambridge University Press, Cambridge, 325–339.

Forth, J., Wiggins, G. A., & McLean, A. (2010). Unifying conceptual spaces: Concept formation in musical creative systems. *Minds and Machines*, *20*(4), 503-532.

Hanna, S. (2005). Where creativity comes from. *Proceedings of Human Interaction 2005*, 45-70.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Fabius, O., & van Amersfoort, J. R. (2015*). Variational recurrent auto-encoders*. arXiv preprint arXiv:1412.6581.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press.

Galenson, D. W. (2009). Old masters and young geniuses: The two life cycles of human creativity. *Journal of Applied Economics*, *12*(1), 1-9.

Kingma, D.P., & Welling, M. (2014). Auto-encoding variational Bayes. *International Conference on Learning Representations 2014.*

Kuhn, T. S. (2012). *The structure of scientific revolutions.* University of Chicago Press.

Liu, Y. T. (2000). *Creativity or novelty?: Cognitive-computational versus social-cultural*. Design Studies, 21(3), 261-276.

Martindale, C. (1990). *The clockwork muse: The predictability of artistic change*. Basic Books.

Ritchie, G. (2006). The transformational creativity hypothesis. *New Gener Comput* **24,** 241–266

Saunders, R. (2011). Artificial Creative Systems and the Evolution of Language. In *ICCC* (pp. 36-41).

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

Saunders, R., & Bown, O. (2015). Computational social creativity. *Artificial life*, *21*(3), 366-378.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press.

Thornton, C. (2007). A quantitative reconstruction of Boden's creativity theory. *Artificial Intelligence*.

Tomasello, M. (1999). The human adaptation for culture. *Annual review of anthropology*, *28*(1), 509-529.

Van der Maaten, L. & Hinton, G. (2008). Visualising data using t-SNE. *Journal of Machine Learning Research*, *9*(11), 2579-2605.

Vygotsky, L. S. (2004). Imagination and creativity in childhood. *Journal of Russian & East European Psychology*, *42*(1), 7-97.

Wiggins, G. A. (2006a). A preliminary framework for description, analysis and comparison of creative systems. Journal of Knowledge-Based Systems, 19(7), 449–458.

Wiggins, G. A. (2006b). Searching for computational creativity. New Generation Computing, 24(3), 209–222.