2014 IEEE World Congress on Computational Intelligence, Beijing, China, 6-11.07.2014, IJCNN, IEEE Press, pp. xxx-yyy.

Memetics and Neural Models of Conspiracy Theories

Włodzisław Duch

Abstract—**Conspiracy theories, or in general seriously distorted beliefs, are widespread. How and why are they formed in the brain is still more a matter of speculation rather than science. In this paper one plausible mechanisms is investigated: rapid freezing of high neuroplasticity (RFHN). Emotional arousal increases neuroplasticity and leads to creation of new pathways spreading neural activation. Using the language of neurodynamics a meme is defined as quasi-stable associative memory attractor state. Depending on the temporal characteristics of the incoming information and the plasticity of the network, memory may self-organize creating memes with large attractor basins linking many unrelated input patterns. Memes with fake rich associations distort relations between memory states. Simulations of various neural network models trained with competitive Hebbian learning (CHL) on stationary and non-stationary data lead to the same conclusion: short learning with high plasticity followed by rapid decrease of plasticity leads to memes with large attraction basins distorting input pattern representations in associative memory. Such system-level models may be used to understand creation of distorted beliefs and formation of conspiracy memes, understood as strong attractor states of the neurodynamics.**

I. Introduction

BELIEFS in conspiracy theories are a part of much wider subject: formation of beliefs, distorted memories, twisted worldviews, and in general investigating ways in which learning fails to represent the data faithfully. Neural networks community tries to achieve perfect learning, but there is another, neglected side of learning and memory formation. When observations are not learned perfectly, what types of errors one may expect, and how it may influence beliefs and actions? Which observations will be neglected, and which will be remembered, transformed into memes, and are likely to be transmitted in a distorted form to other people?

Memetics, introduced in the 1976 book "The Selfish Gene" by Richard Dawkins [1], tried to explain cultural information transfer and persistence of certain ideas in societies. Memes may be understood as sequences or information structures that tend to replicate in a society. Despite great initial popularity of memetics, and the desperate need of mathematical theories to underpin social science, theories connecting neuroscience and memetics have not been developed. The Journal of Memetics was discontinued in 2005 after 8 years of electronic publishing. Memetic ideas were relegated into a set of fuzzy philosophical and psychological concepts of little interest to neuroscience. The lack of efforts to understand distortions of information transmission and memory storage in biological learning systems is certainly related to the lack of theoretical models, and to the experimental difficulties in searching for memes in brain activity.

Recently McNamara [2] has argued that neuroimaging technology may be used to trace memes in the brain, and to measure how they change over time. Following Heylighen and Chielens [2] *memotype* and *mediotype* distinction he proposes to distinguish *i-memes*, internal activation of the central nervous system, from the external transmission/storage of information structures, the *e-memes* existing in the world (for example, created by marketing and various media). We should clearly distinguish abstract information structure of memes, and their implementation in the brain or in some artificial cognitive system. Internal representation of i-memes is created by forming memory states that link neural responses resulting from e-meme perception to behavioral (motor) responses that are necessary for replication of memes, linking sensory, memory and motor subsystems. Sets of memes forming *memeplexes* determine world views, including culture, values and religions, predisposing people to accept and propagate selected memes.

In the fascinating book "Why people believe weird things" Michel Shermer writes about 25 fallacies that lead people to believe conspiracy theories and other weird things [4]. This is certainly a very complex topic: brains are predisposed to perceive various observed patterns as meaningful information, forming theories and searching for explanations, referring to the long-term episodic and semantic memory. The whole conceptual framework that is needed to interpret new observations includes memes, activated by the resulting associations. Once such theories are established it becomes hard to avoid the observer effect. Observations that agree with established beliefs will strongly activate brain networks, thanks to the mutual co-activations of memeplex patterns, creating additional memes that make the whole memeplex even stronger. Contradicting observations will arouse weak activations and be ignored. Science systematically tries to falsify hypothesis by performing experiments, but from the evolutionary perspective falsification is simply too dangerous. In slowly changing environment stability of beliefs is more important, even at the price of wide acceptance of meaningless taboos and superstitions. Even today educational systems in most countries do not encourage skeptical thinking. Religious leaders and conservative politicians are all against instating skepticism into the educational system, in fear of destabilization of established world views. There is little or no penalty for accepting false beliefs, distorted views of reality and conspiracy theories.

W. Duch is with the Department of Informatics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Toruń, Poland (e-mail: wduch@is.umk.pl). This work was started when he worked as Nanyang Visiting Professor in the School of Computer Engineering, Nanyang Technological University, Singapore.

The complexity of the belief formation processes has discouraged scientists from approaching this important problem. Obviously no simple computational model is going to explain all facts related to formation and preservation of beliefs, and in particular conspiracy theories. This should not discourage us from forming testable hypothesis based on neurodynamics. After all simple neural network models introduced by Hopfield and Kohonen, despite being only loosely inspired by neurobiology, have found a number of applications in computational psychology and psychiatry.

The next section is devoted to memetics and representation of information in the brain. It is followed by a section on competitive learning models of memory formation. These models are used to illustrate some mechanisms of memory distortions. Remarks about implications of network simulations for the theory of memetics are presented in section four, and the final discussion in section five.

II. MEMETICS AND INFORMATION IN THE BRAIN

A. Subjective information

Ultimately all thoughts and beliefs result from neurodynamics. The flow of neural activation through neural systems is determined by many biological factors, including brain connectivity, concentration of neurotransmitters, emotional arousal, priming effects, brain stem activity. Information is acquired and internalized in the brain through direct observation of patterns in the world, direct communication with people and animals, and indirectly through various media, texts and physical symbols of all sorts. Brains provide material support for mental processes, understanding and remembering symbols, ideas, stories. Memes are units of information that spread in cultural environment, information granules that prompt activation of patterns in brains groomed in particular subculture. Therefore the same information may become a meme for some brains, and may be ignored by others. Understanding is a process that requires association of new information with what has already been learned.

New things are learned on the basis of what is already known to the system. This is a general principle behind brain activity [5]. Patterns are encoded in memory depending on the context, sequence of events, association with known facts, properties of already encoded experiences, attention and general mental state. Definition of Shannon information as entropy does not capture the intuitive meaning of the value of information for the cognitive system. The amount of optimal (in the minimum length description sense [6]) restructuring of the internal model of the environment resulting from new observation (new meme added to the memeplex) is a good subjective measure of the quantity of meaningful information carried out by this observation. **Pragmatic information** that captures the subjective meaning of information is based on the difference between algorithmic information before and after observation is made [5]. Itti and Baldi used similar idea to define the amount of surprise as the relative entropy or Kullback-Leibler (KL) divergence between the posterior and prior distributions of beliefs in Bayesian models [7].

B. Memes in brains

In memetics information structures that reflect part of mental content based on a network of memes are called memeplexes. They evolve in response to enculturation and exposure to observed patterns. Specific cultural behaviors, learned concepts, word meanings, collocations or phrases describing ideas, may be treated as memes, but some are very rare and difficult to acquire, while others spread quickly with ease. Mental content can be much wider than just the network of memes. Memetics should position itself in respect to the theory of communication, language acquisition and learning.

Using the language of neurodynamics **a meme is** defined as **quasi-stable associative memory attractor state, with robust attractor basin**. Brain activation $A(w)$ after hearing a word w (a set of words, experiencing a cue) may rapidly evoke activation corresponding to meme $A(M(w))$. Such state may be activated by many associations. For simple visual percepts, such as shapes of objects, similarity between brain activations $A(M)$ in the inferotemporal cortical area have been directly compared, using fMRI neuroimaging, to the similarity of the shapes of these objects [8]. Significant similarity has been also found in the whole brain activity when people think about various objects [9], showing how meanings are encoded in distributed activity of the brain. Similarity between memes corresponding to perceived objects $M_i \Leftrightarrow O_i$, may be roughly compared to some measures of similarity between object properties. Therefore similarity between brain activities $A(M_1)$ and $A(M_2)$ that represent two memes M_1 and M_2 evoked by objects O_1,O_2 (percepts, cues, words) should be comparable to some measures of object similarity:

$$S_a(A(M_1),A(M_2)) \sim S_o(O_1,O_2). \quad (1)$$

McNamara [2] hopes to detect the signature patterns of new memes by analyzing the neurodynamics of learning novel name–action associations for abstract category names, looking at the changes of the brain connectivity profiles. This may be a useful strategy for abstract categories, or for simple percepts, but general search for signatures of memes using neuroimaging techniques will be very difficult. Activation patterns may significantly differ for individual people, depending on their memeplexes. For the same person distribution of fMRI activations may change at different times of the day. Transcranial magnetic stimulation (TMS) disrupting the function of the left inferior frontal gyrus has already been used to alter belief formation in favor of remembering more bad news [10]. Such brain stimulation may be used to change acceptance of memes that would normally be ignored.

Many concepts gradually change their meaning over time. Genes, once defined as sequences of DNA base pairs, are now acknowledged as distributed DNA and RNA templates, with exons on different chromosomes, “encoding a coherent set of potentially overlapping functional products” [11]. Precise definition of a gene is difficult because they are structures of partially mutable highly organized molecular matter living in specific network of complex processes. They exist because specialized environment facilitates their replication. Strong coupling with the environment makes the concept fuzzy: it is not just DNA sequences, but complex pattern in the whole network. The whole system is responsible for replication of information. Memes are even more difficult to extract from the whole network of brain activities.

Understanding how brain connectivity and other factors determine neurodynamics, encode beliefs, filter incoming information, distort it and transmits it further, is certainly a grand challenge. Complex information processing in the brain has not yet been understood in sufficient details to allow for development of comprehensive theories of such processes, but some insights based on simple memory models may be gained. New information added to the memeplex (existing pool of interacting memes, or attractor states) becomes distorted, changes the memeplex, and is replicated further. Once a set of distorted memory states is entrenched it becomes a powerful force, attracting and distorting new thoughts associated with them. Deep encoding of information that enhances the memeplex is one of the reasons why conspiracy theories are so persistent.

C. *Concepts in brains and computers*

The vector model, popular in the Natural Language Processing (NLP), represents word meaning using correlations between co-occurrence with other words in some window covering the text around a given word. Vectors $\mathrm{C}(w)$ represent words w by averaging over many contexts restricted to a specific meaning of a given word. Faithful representation of word meaning demands similar ordering of distance relations $\mathrm{D}(\mathrm{C}(w_1),\mathrm{C}(w_2))$ between vectors $\mathrm{C}(w_1),\mathrm{C}(w_2)$ representing words w_1, w_2, as shown by similarities between brain activations when concepts associated with these words are contemplated:

$$\mathrm{S}_a(A(w_1),A(w_2)) \sim \mathrm{D}(\mathrm{C}(w_1),\mathrm{C}(w_2)) \qquad (2)$$

Each vector $\mathrm{C}(w)$ attempts to approximate meaning of the word that is encoded in the distribution of brain activity [9]. Without priming effects [12] and association of words with memes only very coarse representation is possible. Brain activations strongly depend on context, and thus also distance functions $\mathrm{D}(\mathrm{C}(w_1),\mathrm{C}(w_2);\mathrm{cont})$ should be context dependent. The whole process is dynamic, with spreading of neural activations responsible for priming related concepts. Meaning is thus connected to the activation of many subnetworks in the brain, memory of sensory qualities and motor affordances. Dynamical approach to the NLP vector model has not yet been fully developed although some steps in this direction have been made [13][14]. Despite our efforts (Duch, unpublished) to describe dog breeds in terms of skin, head and body features derived from databases and semi-structured texts describing dogs, it was not possible to identify accurately dog breeds by asking questions about their features. Using images (or just silhouettes) of dogs leads to more accurate and faster identification of dog breeds. Brain activity evoked by hearing or reading words is connected to internal imagery at relatively high level of invariant, multimodal object recognition. Similarity functions between objects $\mathrm{S}_o(O_1,O_2)$ based only on correlations between verbal descriptors, cannot do justice to estimations of similarity of brain activations. Finer discrimination may require recall of lower-level sensory qualities, referring to particular shapes, colors, movements, voice timbre or tastes. This shows the need for representation of sensory imagery in NLP systems. Vector representation based on word correlations does not reflect essential properties of the perception-action-naming activity of the brain [15], it does not even contain structural description in terms of object features or phonology. More details on word representation in the brain and its relation to the vector model may be found in [13][14].

In the next section competitive learning models are introduced, and then used to illustrate the process of learning that leads to memes based on distorted relations.

III. Competitive Learning and Weird Beliefs

Conspiracy theories have serious consequences for politics, especially environmental policies and health, they facilitate growth of political extremists and dangerous religious sects. Conspiracy theories are investigated mainly by sociologists, focusing on hidden networks controlling political and economic relations. They may be viewed as a particular form of weird beliefs that cover all kinds of distorted views. In the past secret societies were rather rare, but now media try to stir controversy discussing GMO, vaccines, AIDS, miracle cures, UFOs, prophecies, assassinations, airplane crushes and other issues, although there is scientific or common sense consensus. While there are many psychological reasons for formation of such beliefs so far there have been no attempts to create a cognitive theory supported by computational models, generating hypothesis for testing.

The language of memetics has not helped to explain deeper reasons for such beliefs [1],[16]. Conspiracy theory may be treated as a memeplex that is easily activated by various pieces of information, giving it meaning consistent with the memeplex responses. From neurobiological perspective learning requires adaptation, changing connectome, physical structure of the brain. Learning is thus difficult and energy-consuming. Simple explanation of complex phenomena have therefore great advantage, as long as they do not lead to behaviors that significantly decrease chances for reproduction. Evolutionary Darwinian adaptations require many generations and to have noticeable influence on human beliefs they should affect large subpopulations. Evolution may explain changes in understanding such concepts as human freedom, abandonment of slavery, caste and racial divisions, attitudes towards children (selling children into slavery continued to 19th century). Distortions of the learning process provides more plausible explanation for rapid learning leading to weird beliefs. However, the field of neural networks aiming at achieving perfection in learning paid little attention to understand distortions of learning.

There are many scenarios leading to formation of conspiracy theories. A rather common situation is due to the rapid freezing of high neuroplasticity. Initial uncertainty of important information (there are rumors that something strange or dangerous has happened) leads to confusion and strong anxiety (perhaps the news are not true, who knows what has really happened). High emotions and stress leads to release of neurotransmitters and neuromodulators from the brain stem nuclei, through the ascending pathways, activating serotonin, norepinephrine, and dopamine systems. In this period of strong arousal increased brain plasticity allows for rapid learning. It is not yet clear what information is worth encoding, so all facts and gossips are memorized. Once the uncertain situation is resolved, either in positive or in negative way, there is no need for further learning. Strong emotions have depleted neurotransmitters, therefore neuroplasticity is rapidly decreased.

This scenario may be reproduced in many unsupervised competitive learning models, including ART model that has vigilance parameter [18] to regulate neuroplasticity. Many other competitive learning models based on Hebbian learning have been presented in [19]. DemoGNG 2.2 Java package, written by Bernd Fritzke and Hartmut S. Loos [20], implements winner-take-all learning in Self-Organizing Map (SOM), Competitive Hebbian and Hard Competitive Learning, Neural Gas, Growing Neural Gas, Growing Grid, and other algorithms. In all these algorithms activity of units representing neurons is compared with the input, and those units with the best match adapt their parameters increasing their activation. Neurons in the neighborhood of a winner are also allowed to adapt, depending on their distance from the winner. If there is no clear match constructive algorithms add new neurons allowing the network to grow.

The rapid freezing of high neuroplasticity (RFHN) model described here is based on the following assumptions:

- Emotions and uncertain stressful situations at the beginning of learning lead to high neuroplasticity.
- High neuroplasticity is imitated in the model by large learning rates (due to the primary neurotransmitters), and by a broad neighborhood of the winner neuron (due to the diffuse neuromodulation and volume learning).
- The network tries to reflect associations between input vectors, adapting neuron parameters (usually codebook vectors) to approximate distribution of information contained in the presented input vectors.
- Sudden decrease of the uncertainty and emotional arousal is mirrored by the decrease of learning rates and neighborhood sizes, leading to distortions of complex relations between input items.
- Slow forgetting that follows rapid freezing is based on memory reactivations, and contributes to the retention of memory states represented by the highest number of neurons only, forming clusters of nodes with large and strong basins of attraction that link many states.
- Clusters of neurons that are frequently activated and thus easily replicated represent memes.
- Conspiracy theories are characterized by a number of strong memes, with many neurons encoding information that has never been presented, forming distorted associations between facts.

As a result these networks do not reflect real observations. The RFHN model may be simulated using several competitive learning models. In fact all such models show similar behavior, therefore only the Self-Organized Maps [17], and Neural Gas model with Competitive Hebbian Learning (NG-CHL) [21] are shown here for illustration.

The basic idea of competitive learning is to approximate the activity of neural cell assemblies by neurons (units) that serve as codebook vectors $\mathbf{W}(t)$. They represent receptive fields, adapting to the probability density of the incoming signals. Each neuron receives input signals and competes with other neurons using the winner-takes-most (or takes all) principle, leaving only a small subset of active units that are updated. The winning neural assembly is represented by a vector $\mathbf{W}^{(c)}(t)$ and a small group of vectors in its direct neighborhood $O(c)$. SOM starts with a fixed two-dimensional grid of neurons. Learning proceeds by identifying most similar codebook vector to the current observation $\mathbf{X}(t)$, and updating the codebook vector and vectors in its immediate physical neighborhood according to the formula:

$$\text{For } \forall i \in O(c)$$

$$\mathbf{W}^{(i)}(t+1) = \mathbf{W}^{(i)}(t) + h(r_i, r_c, t)\left[\mathbf{X}(t) - \mathbf{W}^{(i)}(t)\right] \quad (3)$$

where the neighborhood is assumed to be Gaussian:

$$h(r, r_c, t, \varepsilon, \sigma) = \varepsilon(t)\exp\left(-\left\|r - r_c\right\|^2 / \sigma^2(t)\right) \quad (4)$$

The size of this neighborhood is decreased from the initial value of dispersion σ_i to the final value σ_f according to the formula:

$$\sigma(t) = \sigma_i \left(\frac{\sigma_f}{\sigma_i}\right)^{t/t_{\max}} \quad (5)$$

The maximal age t_{max} determines the annealing schedule. The learning rate is similarly decreased by:

$$\varepsilon(t) = \varepsilon_i \left(\frac{\varepsilon_f}{\varepsilon_i}\right)^{t/t_{\max}} \quad (6)$$

SOM model has been used with success in comparison to other models [22] to explain orientation and ocular dominance columns in the visual cortex.

The NG-CHL algorithm does not have a fixed initial grid topology as SOM, neurons float like gas particles. At each adaptation step a connection between the winner and the second-nearest unit is created, if it does not already exist. The newly created or existing selected edges are refreshed receiving age=0, while the age of other edges emanating from the winner neurons are increased by 1. The reference age is gradually changed from T_i to T_f according to:

$$T(t) = T_i \left(\frac{T_f}{T_i}\right)^{t/t_{\max}} \quad (7)$$

Edges that are not refreshed for more than $T(t)$ steps are removed. This simulates forgetting mechanism.

Following computational experiments have been done:

- Training SOM and NG-CHL on stationary data concentrated in two distinct areas, with initial high plasticity and rapidly decreasing learning rates.
- Training SOM and NG-CHL on non-stationary data from observations that move and suddenly change, with initial high plasticity and rapidly decreasing learning rates.

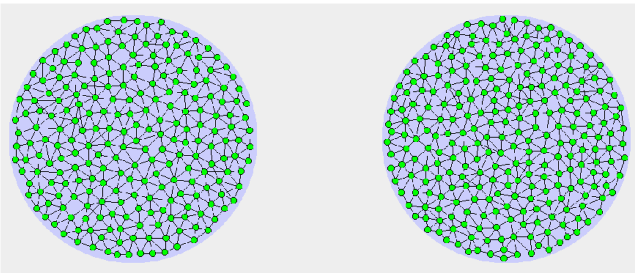
- Retraining the model after malformation of relations has already occurred, using temporally increased plasticity.

The number of neurons in the brain is extremely large, therefore it is instructive to check how the number of network nodes in simulations will affect distributions. For the stationary experiments 10.000 nodes have been used, with initial parameters randomly distributed, and signals coming from two separated circular areas. This should represent two alternative situations that are monitored. For the non-stationary situation all parameters were initially concentrated in the rectangular patch, simulating situations in which restricted domain has already been learned and stable. Then the patch moves across the whole domain providing new input patterns (observations) from the areas it covers. When the edge of the domain is reached the patch jumps to the other side.

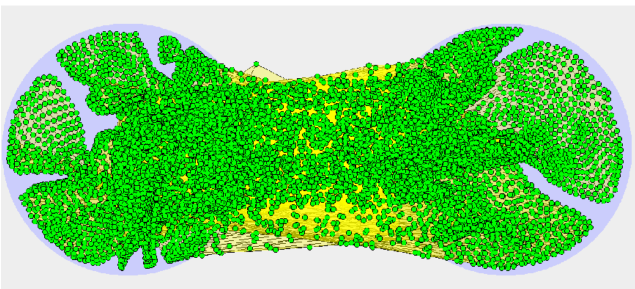
IV. CONSPIRACIES AND MEMORY DISTORTIONS

A. Stationary situation

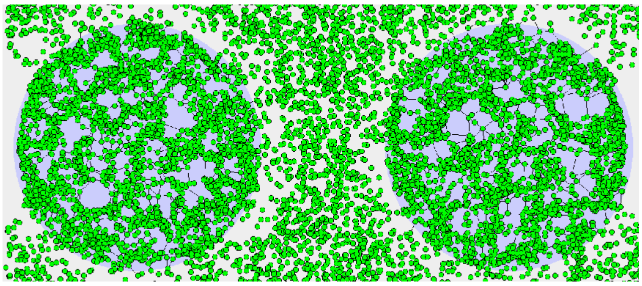
Perfect representation of all signals should cover two distinct circular areas. A good solution that requires slow learning with 500.000 steps is shown below.



Training 100x100 SOM network, with initial σ_i=5, σ_f =0.01, ε_i=1, ε_f =0.001, for 10.000 steps, did not pull all parameters of neurons towards data area. Despite high density of neurons some gaps have been left and were not removed by further retraining. This effect comes from the dynamics of learning with shrinking neighborhoods. There is a greater chance for neurons near the age to be pulled towards high density areas by many neurons that are selected as winners than to be pulled towards the data in the gap area.

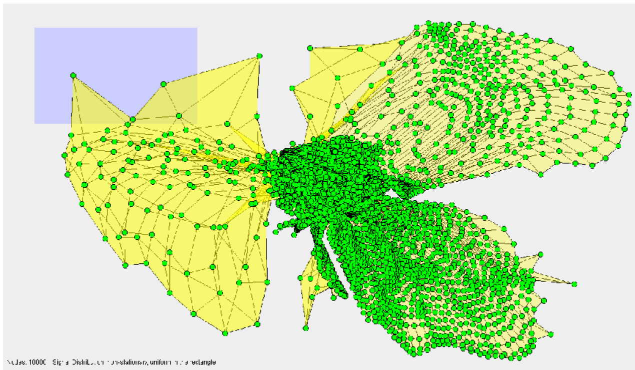


The NG-CHL model with initial high plasticity and rapidly decreasing learning rates has also produced big gaps and high density areas. Forgetting parameters have been set to $edge_i$=20 and $edge_f$ =200. Further retraining with fast forgetting creates even bigger gaps. Many input patterns are therefore associated with high density clusters acting as memes. Associations with other input patterns are based more on stereotypes (clusters) rather than faithful observations.

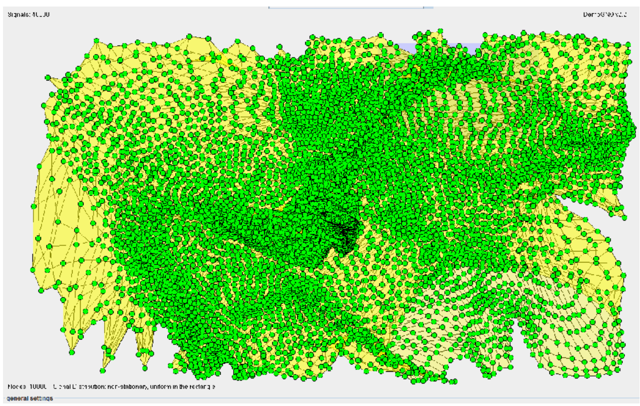


B. Non-stationary situations

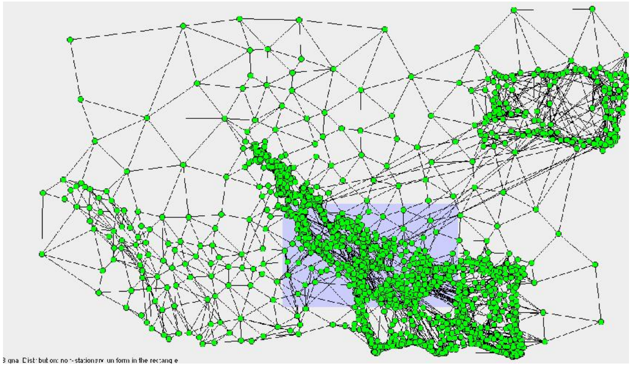
The nonstationary situation shows much stronger distortions. SOM starting from high plasticity (same parameters as for the stationary case) after fast decrease (10.000 steps) shows very strong concentration of neurons that point to the initial patterns, and did not learn much later.



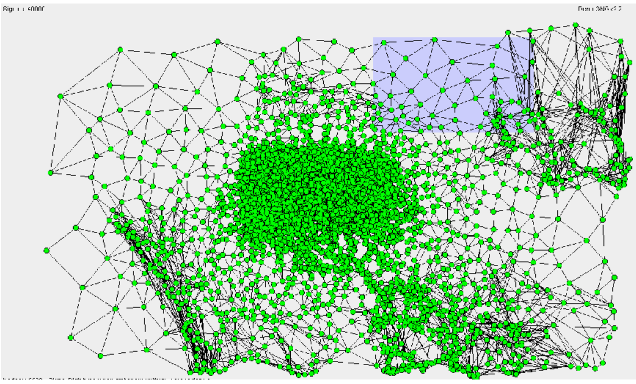
Further training with increased plasticity may somehow repair the distorted view, although it is very hard to remove strong meme that has been formed in the center. Large basin of attraction for this meme will lead to its activation frequent activation even by irrelevant input patterns.



The NG-CHL algorithm also creates completely distorted representation. After 40.000 steps it has created two separate memplexes, each with several strong memes that are used to interpret all incoming patterns.



Training with faster forgetting creates 4 larger memeplexes, but still has completely distorted view of the input patterns. It is quite difficult to create faithful representation of input patterns for non-stationary signals. Very long training times with several hundred thousand iterations are needed to achieve it. In case of rapidly changing situations it is much more likely that a distorted view will be learned rather than faithful representation of reality.



V. Conclusion

Biological and psychological belief forming mechanisms are very complicated. Predispositions for accepting distorted views of reality may come as a side effect of education and life experiences and therefore are rather hard to investigate. Accepting simple explanations is rewarding, creates pleasant feelings of understanding. Complex explanations requires a lot of effort to understand and it takes time. It feels better to have a simple (although inadequate) explanation than to have no explanation at all.

Why do people believe in conspiracy theories? Because this is how the brain works. Neurodynamics helps to understand the conditions under which large basins of attractions, called memes, are created in memory networks. how and why they form memplexes that lead to the distorted associations. This is an important step towards linking memetics with theoretical and experimental brain science. Perhaps memes can be measured [2], and computer simulations should help to define most suitable experimental conditions.

What lessons can we draw from computational experiments with competitive learning? The rapid freezing of high neuroplasticity (RFHN) model presented here is very simple but it seems that all types of competitive learning models show similar behavior. More complex models with high-dimensional input patterns almost certainly will have even bigger problems with faithful representation of input patterns and will lead to large attractor basins that can be interpreted as memes. Only slow learning guarantees faithful representation. Analysis of formation of weird beliefs is very important, but so far there have been no attempts to link it to brain processes. Simulations presented here should draw attention to the need of analysis of the type of distortions that are common in neural networks.

Acknowledgment

This research was supported by the Polish National Science Foundation grant DEC-2013/08/W/HS6/00333.

References

[1] R. Dawkins, *The Selfish Gene*. New York: Oxford Uni. Press, 1976.

[2] A. McNamara, "Can we measure memes?", *Frontiers in Evolutionary Neuroscience* 3,doi:10.3389/fnevo.2011.00001

[3] F. Heylighen, F., and K. Chielens, "Evolution of Culture, Memetics," in Encyclopedia of Complexity and Systems Science, B. Meyer, Ed. Springer, 2009, pp. 3205-3220.

[4] M. Shermer, "Why People Believe Weird Things: Pseudoscience, Superstition, and Other Confusions of Our Time." Souvenir Press 2007

[5] W. Duch, "Towards comprehensive foundations of computational intelligence." In: W. Duch and J. Mandziuk, Challenges for Computational Intelligence. Springer Studies in Computational Intelligence, Vol. 63, pp. 261-316, 2007.

[6] J. Rissanen, "Modeling by shortest data description". *Automatica* vol. 14 (5), pp. 465–658, 1978.

[7] L. Itti, P. F. Baldi, "Bayesian Surprise Attracts Human Attention", In: *Advances in Neural Information Processing Systems*, vol. 19 (NIPS*2005), pp. 547-554, Cambridge, MA:MIT Press, 2006.

[8] H.P. Op de Beeck & C.I. Baker, "The neural basis of visual object learning." Trends in Cognitive Sciences, vol. 14, pp. 22-30, 2010.

[9] T.M. Mitchell, S.V. Shinkareva, A. Carlson, K.M. Chang, V.L. Malave, R.A. Mason, M.A. Just, "Predicting human brain activity associated with the meanings of nouns." Science. Vol 30;320(5880), pp. 1191-95, 2008.

[10] T. Sharot, R. Kanai, D. Marston, C. Korn, G. Rees, G. & R.J. Dolan, "Selectively Altering Belief Formation in the Human Brain." PNAS vol. 109(42), pp. 17058-17062, 2012.

[11] M.B. Gerstein, et al. "What is a gene, post-ENCODE? History and updated definition," *Genome Research* vol. 17, np. 6, pp. 669–681, 2007.

[12] T.P. McNamara, "Semantic Priming. Perspectives from Memory and Word Recognition", New York: Psychology Press, 2005

[13] S. Lamb, *Pathways of the Brain: The Neurocognitive Basis of Language*. Amsterdam & Philadelphia: J. Benjamins Pub. Co. 1999.

[14] W. Duch, P. Matykiewicz, J. Pestian "Neurolinguistic Approach to Natural Language Processing with Applications to Medical Text Analysis." Neural Networks vol. 21(10), pp. 1500-1510, 2008.

[15] F. Pulvermuller, *The Neuroscience of Language. On Brain Circuits of Words and Serial Order*. Cambridge, UK: Cambridge University Press, 2003.

[16] K. Distin, The Selfish Meme. Cambridge, UK: Cambridge University Press, 2005.

[17] T. Kohonen, *Self-Organizing Maps*. 3rd, ext. ed. Springer, T. 2001

[18] S. Grossberg, "Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world." *Neural Networks* vol. 37, 1-47, 2012.

[19] R. Xu, D. Wunsch II, "Survey of Clustering Algorithms", IEEE Transactions on Neural Networks, vol. 16(3), pp. 645-678, 2005

[20] B. Fritzke, H.S. Loos, DemoGNG 2.2, http://www.demogng.de/

[21] T. M. Martinetz, K.J. Schulten, "Topology representing networks." *Neural Networks* vol. 7(3), pp. 507-522, 1994.

[22] E. Erwin, K. Obermayer, K. Schulten, "Models of orientation and ocular dominance columns in the visual cortex: a critical comparison." *Neural Computations* vol. 7, pp. 425-468, 1995