Cognitive Map Dimensions of the Human Value System Extracted from Natural Language

Alexei V. SAMSONOVICH[a] and Giorgio A. ASCOLI[b]
[a]*Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA 22030.*
[b]*Krasnow Institute for Advanced Study and Psychology Department, George Mason University, Fairfax, VA 22030-4444*

Abstract. The notion of a human value system can be quantified as a cognitive map, the dimensions of which capture the semantics of concepts and the associated values. This can be done, if one knows (i) how to define the dimensions of the map, and (ii) how to allocate concepts in those dimensions. Regarding the first question, experimental studies with linguistic material using psychometrics have revealed that valence, arousal and dominance are primary dimensions characterizing human values. The same or similar dimensions are used in popular models of emotions and affects. In these studies, the choice of principal dimensions, as well as scoring concepts, was based on subjective reports or psycho-physiological measurements. Can a cognitive map of human values be constructed without testing human subjects? Here we show that the answer is positive, using generally available dictionaries of synonyms and antonyms. By applying a simple statistical-mechanic model to English and French dictionaries, we constructed multidimensional cognitive maps that capture the semantics of words. We calculated the principal dimensions of the resultant maps and found their semantics consistent across two languages as well as with previously known main cognitive dimensions. These results suggest that the linguistically derived cognitive map of the human value system is language-invariant and, being closely related to psychometrically derived maps, is likely to reflect fundamental aspects of the human mind.

Keywords. Cognitive architectures, affective dimensions, psychometric.

Introduction

Neuromorphic cognitive map is a functional unit that plays a central role in the Biologically Inspired Cognitive Architecture developed at George Mason University (BICA-GMU) by our research team [1-4]. The notion of a cognitive map, however, is not limited to the field of artificial general intelligence (AGI). The term "cognitive map" had been used in cognitive sciences for several decades with various meanings [5, 6]. The modern notion of a cognitive map was introduced by O'Keefe and Nadel [7] based on the hippocampal place cell phenomenon discovered at that time [8]. This notion was subsequently extended to include cognitive mapping of non-spatial features of contexts and paradigms, based on the spatial analogy [e.g., 9-12].

In the present work, a *cognitive map* is understood as a mapping from a set of cognitive representations (e.g. concepts, words) to an abstract continuous metric space, such that semantic relations among representations are reflected in geometric relations in the indexing space. This intuitive definition unfolds as follows. In a spatial cognitive map, the metrics are proportional to the perceived distances between associated landmarks in the physical world. In this case the cognitive map is essentially a model of perceived space [7]. Similarly, a temporal cognitive map, if it were found in the brain as a separate functional unit, would be a model of a perceived timeline. Another example of a cognitive map is a color space, in which locations correspond to perceived colors [13].

Speaking more generally, we distinguish various kinds of cognitive maps (Figure 1), based on the semantics they represent (logic, values, feelings, qualia) and on the representation systems they map (e.g., one may distinguish contextual and conceptual cognitive maps). While the idea of mapping representations of concepts onto an abstract space is not new [14], cognitive maps beyond spatial and temporal dimensions remain unexplored terrain in cognitive neurosciences. How to design a cognitive map: its topology, geometry, and the associated semantics? How to allocate representations on a map? Can this be done objectively, and/or from the first principles?

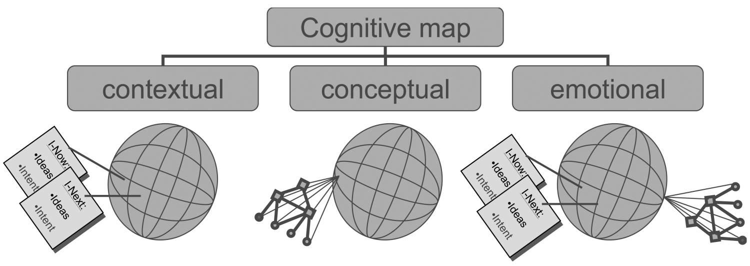

Figure 1. Cognitive maps are abstract metric spaces that reflect semantics of associated symbolic representations. Different kinds of cognitive maps may represent different aspects of semantics and/or map different kinds of representations.

Here we focus on a particular kind of cognitive maps: *conceptual value maps*. We believe that the notion of a human value system can be made more precise and more useful when it is represented with a cognitive map, the dimensions of which capture the values of concepts. This kind of a map can be constructed, if one knows how to define the dimensions of the map and how to allocate concepts in them.

The idea of the approach pursued in the present study is to use linguistic corpora as a source of data about the semantics of concepts (represented in this case by words). The hope is that self-organization may help us find the principal dimensions and to allocate concepts automatically. We expect that the problems of cognitive map creation can be solved in this case using available linguistic data. Therefore, we select a dictionary of words as our study material, keeping in mind that words represent concepts, and concepts are associated with values.

Simple as it is, the idea of applying the notion of a cognitive map to linguistic corpora appears to be unexplored, while multidimensional metrics were used to characterize semantics of words, concepts and feelings in many cognitive studies. Examples include theoretical models of emotions [e.g., 15, 16] that go along with experimentally derived psychometric dimensions of words [17, 18] and partially overlap with abstract studies of "quality dimensions" [14]. All of the above influenced modern cognitive architecture designs that make use of principal cognitive dimensions (see, e.g., Adams, this volume). In the aforementioned experimental studies, the choice of principal dimensions, as well as related scoring of concepts, was based on subjective reports or psycho-physiological measurements. Can a cognitive map of human values be constructed without testing human subjects? Here we show that the answer is positive, using generally available dictionaries of synonyms and antonyms.

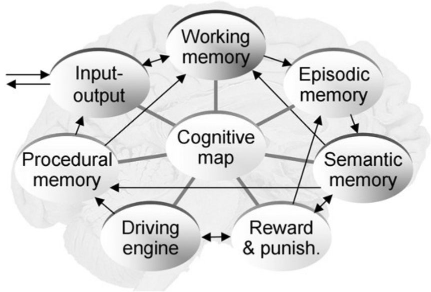

Figure 2. A bird view of BICA-GMU, as described in [2].

The present study of cognitive maps is best framed in the context of our cognitive architecture design: BICA-GMU (Figure 2). Information processing in BICA-GMU occurs at a higher symbolic level, based on new building blocks called “a schema” and “a mental state” [2]. Intuitively, the notion of a schema can be associated with that of a concept, while the notion of a mental state can be associated with that of a context. Multiple instances of schemas and mental states fill in the three main memory systems in BICA-GMU: working, semantic and episodic memory. When new information comes to the system through the input-output buffer, it gets represented by instances of schemas. This is done with the help of procedural memory that "knows" what schemas should be used for each given kind of input. The rest of information processing does not have a pre-defined solution and may involve search of the entire semantic and/or episodic memories at each step. Thus, filtering of the exploding tree of possibilities becomes vital for successful operation of BICA-GMU in practical scenarios. This filtering is one of the main functions of neuromorphic cognitive maps in BICA-GMU. In general, it can be understood as a task to suggest a preferred choice of a schema that will be used by the architecture at the next step. Filtering by a cognitive map also constrains the semantics of admissible schemas to a narrow domain in the cognitive space. This mechanism could be used, e.g., in analogy search or in classification of memories. Another aspect of the same cognitive map function is evaluation (how good, how plausible, how exciting, etc.) of a given concept. In this sense, cognitive maps provide an otherwise undefined “metric system” in the field. While the cognitive map is expected to develop its metrics through self-organization, the primary values for a limited number of selected concepts need to be provided by an external source. In the case of BICA-GMU they are provided by the reward and punishment system (Figure 2).

1. Materials and Methods

1.1.Linguistic Corpora

The study presented here was conducted using two linguistic corpora: dictionaries of synonyms and antonyms available as parts of the thesaurus in Microsoft Word 2003 (MS Word). The two corpora apparently have independent origin[1] and different characteristics. The total size of each of them is above 200,000 entries. The core dictionaries used in this study and the corresponding matrices *W* of synonym-antonym relations were extracted automatically following the algorithm described below.

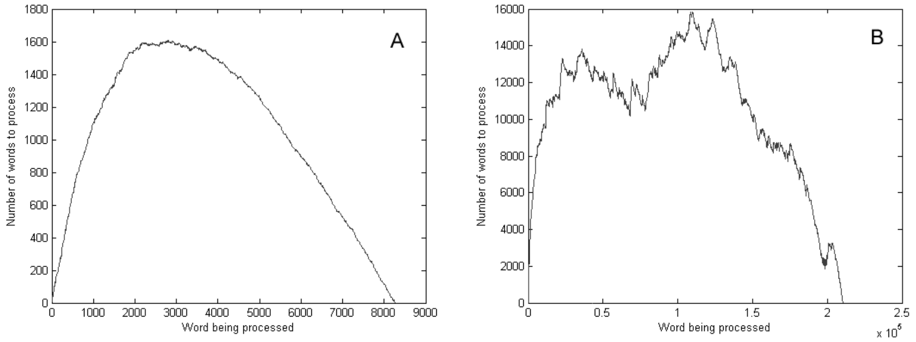

Figure 3. Extraction of the dictionary from MS Word. **A:** English, **B:** French. Abscissa: the word that is being processed. Ordinate: the number of unprocessed words in the retrieved list at the current step.

1.1.1.Extraction of the Core Dictionary

The following algorithm was used to extract a core dictionary from MS Word.

Step 1: Start with one word in the dictionary. Take the next word from the dictionary. Retrieve its synonyms and antonyms from MS Word. Merge them into the dictionary (avoid repetitions). Repeat until the retrieved dictionary is processed.

Step 2: Eliminate duplicates that were not detected during Step 1. Recursively remove all words with less than two connections (see Figure 4 A). The remainder by definition constitutes the "core" dictionary. Symmetrize the relation matrix *W* by making all synonym and antonym links bi-directional[2].

The starting word for English was "first", for French "premier". The resultant sets of words never changed by more than a few words when we tried different starting words.

1.1.2.Characteristics of the Core Dictionaries

The extracted English core has 8,236 words. An average word in it has 3.0 synonyms (1.8 before symmetrization) and 1.4 antonyms (0.8 before symmetrization). The

[1] English thesaurus was developed for Microsoft by Bloomsbury Publishing, Plc. French thesaurus is copyrighted by SYNAPSE Development, Toulouse, France.

[2] Symmetrization is necessary for the energy function (*) to be Hermitean, in which case the relaxation process (**) converges to a fixed point rather than a cycle.

extracted French core has 87,811 words. An average word in it has 6.4 synonyms (3.9 before symmetrization) and 7.5 antonyms (3.9 before symmetrization). The total average number of connections (degree) per word is 4.3 for English core and 14.0 for French core. In each case, the extracted core is a small part of the entire thesaurus.

The graph of synonym-antonym links for the English core is nearly scale-free [19] (Figure 4, cf. [20]), but cannot be used as a cognitive map. For example, considering the graph of English synonyms only, one can see that distances on the graph (measured in the number of links of the shortest path) are semantically misleading:

- Average distance between words = 6.7
- Distance (true, false) = 8
- Distance (big, small) = 5
- Distance (happy, satisfaction) = 7
- Distance (cloth, dress) = 5

This happens because only very few of all synonyms and antonyms of a given word are actually listed in the corpus (the graph is very sparse).

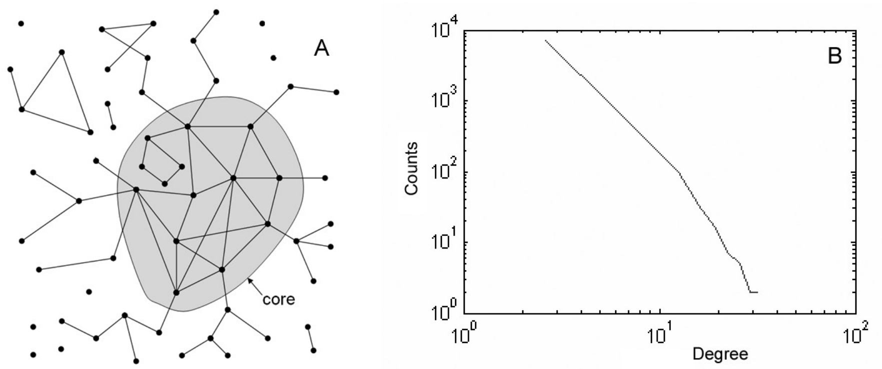

Figure 4. Post-processing of the extracted dictionary. **A:** The notion of a core dictionary. Only nodes that have at least two connections to other core nodes are left in the core. Links are bi-directional and could be synonym-synonym or antonym-antonym connections. **B:** Scaling law analysis [19] shows that the extracted English core dictionary forms a nearly scale-free graph, which is typical for word-association networks [20].

1.2.Statistical-Physics-Inspired Approach to Self-Organization of a Cognitive Map

The idea of our approach to constructing a cognitive map by self-organization is to represent the evolving cognitive map by a statistical mechanical model and to find its ground state that is expected to capture the semantics. Therefore, the heart of our method is the following algorithm, which was designed through trial and error[3].

1. Randomly allocate N words as "particles" (vectors) in R100 in a unit ball.

2. Arrange for attraction between synonyms and repulsion among antonyms by defining an energy function of the system based on the relation matrix W:

$$H(x) = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij}\, x_i \cdot x_j + \frac{1}{4}\sum_{i=1}^{N} |x_i|^4 , \quad x \in R^N \oplus R^{100} .$$

[3] We tried variou ... th the dimensionality ranging from 1 to 100, and selected an optimum.

(1)

3. Simulate thermodynamical relaxation of the system to its ground state (106 iterations) based on the following stochastic equation (is a Gaussian noise):

$$\boxed{\dot{x}_i = -\frac{\partial H}{\partial x_i} + \eta_i(t)\ ,\ \ \langle \eta(t)^2 \rangle \to 0\,.} \qquad (2)$$

4. Rotate the resultant distribution in R^{100} to its principal components (PCs).

5. Identify semantics of the coordinates (PCs) by sorting words along them.

The symmetric relation matrix W in (1) has "+1" entries for pairs of synonyms and "–1" entries for pairs of antonyms, all other matrix elements are equal to zero. During the construction of W, different forms of the same word were treated as synonyms. Convergence of (2) to a ground state was assessed by measuring the maximal displacement of any "particle" in one step of simulated dynamics.

1.3. Psychometric Data Used in This Study

In the analysis of our results we used the Affective Norms for English Words (ANEW) database [21] developed by the Center for the Study of Emotion and Attention (CSEA) at the University of Florida. This database contains 1,034 affective English words. The ANEW database was created using the Self-Assessment Manikin to acquire ratings of *pleasure*, *arousal*, and *dominance*. Each rating scale in ANEW runs from 1 to 9, with a rating of 1 indicating a low value (low pleasure, low arousal, low dominance) and 9 indicating a high value on each dimension. The ANEW database was kindly provided by Dr. Margaret M. Bradley (University of Florida, CSEA).

1.4. Software and Hardware

Algorithms were implemented using XEmacs and GNU C on Dell Optiplex GX620 running Fedora Core 5 Linux. Data preparation and analysis were performed using Microsoft Office 2003 Professional Enterprise Edition on Dell Optiplex GX620 running Windows XP Pro, also using Octave and Matlab 7.0.

2. Results

In all our numerical experiments, 10^6 iterations (and 10^5 iterations in most cases) were sufficient for convergence of the system to its ground state (assessed as described above). In the numerical implementation of (2), the time step was $\Delta t = 0.001$, the initial noise standard deviation $<\eta^2>^{1/2}$ was 0.5, and its value decreased inversely proportionally with time. Our key findings are presented in Figures 5-7 and Tables 1-2.

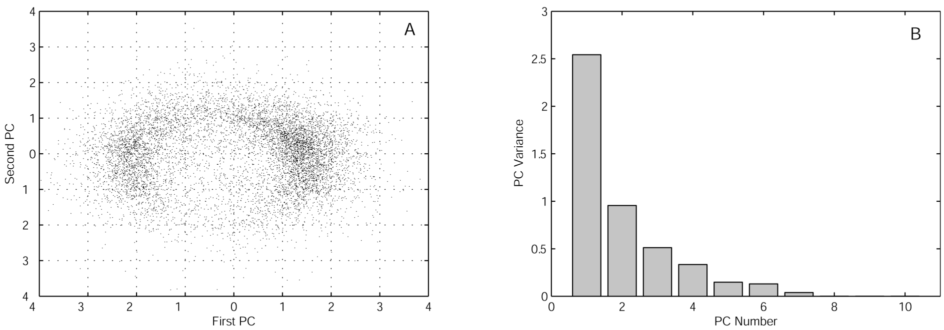

Figure 5. The English core in its "ground state". **A:** The "banana shape" of a final distribution of English words (visible after rotation to the main PC coordinates). The two fuzzy clusters and the corresponding horizontal dimension separate positive and negative values (*valence*). The vertical axis corresponds to another cognitive dimension: "calming vs. exciting" (*arousal*). D=100, H(x) is given by (*) in Section 1.2. **B:** Variances of the first ten PCs (Table 1). PCs are sorted by their variance.

Table 1. Sorted lists for the English core in a ground state. 10 PCs, 5+5 elements of each list. The order number and the variance of each PC are given in the left column.

PC # & Variance	Starting from the one end of the list:	Starting from the other end of the list:
1: 2.54	increase, well, rise, support, accept…	drop, lose, dull, break, poor…
2: 0.95	calm, easy, soft, gentle, relaxed…	difficult, harsh, hard, trouble, twist…
3: 0.51	start, open, fresh, begin, release…	close, delay, end, finish, halt…
4: 0.33	thin, edge, use, length, wet…	center, save, deep, dry, middle…
5: 0.15	essential, need, poverty, basic, necessary…	back, surplus, later, wealth, unnecessary…
6: 0.13	pull, private, receive, owe, keep…	dig, push, channel, ditch, national…
7: 0.04	over, top, on top of, above, impose…	base, under, below, underneath, beneath…
8: 5.6e-8	old, mature, adult, aged, previous…	young, child, dig, immature, new…
9: 1.7e-9	normally, carefully, modestly, frequently, often…	unusually, extremely, rarely, strangely, carelessly…
10: 6e-10	personally, carefully, for myself, amazingly, thoughtfully…	universally, carelessly, normally, generally, usually…

The shape of resultant distributions (Figures 5, 6) was found independent of the initial conditions and the realization of the stochastic noise η, which was sampled from a normal distribution. Interestingly, the geometric properties of shapes of the final distributions for the two languages are similar (cp. Figure 5 A and Figure 6 A): these are bimodal, "banana-shape" distributions, each exhibiting two dense clusters connected by a curved "neck", surrounded by a diffuse "fringe".

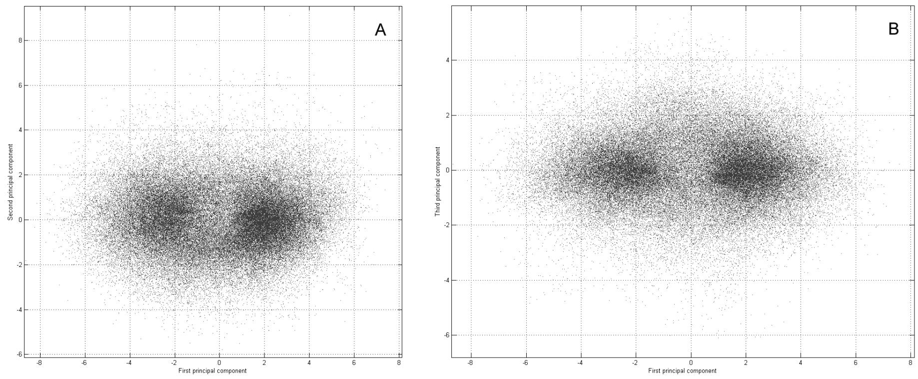

Figure 6. The French core in its "ground state". **A:** PC #1 vs. PC #2. The "banana shape" distribution resembles that of Figure 5 A. **B:** PC #1 vs. PC #3. Like in the case of Figure 5, the distribution is bimodal with two clusters corresponding to positive and negative values (as found by examining the cluster elements).

Another interesting detail is that the PC amplitude is quickly decaying with the PC number (Figure 5 B and Table 1, left column), which may be a consequence of the fact that the matrix W is very sparse – or an indication that there are indeed very few cognitive dimensions in the corpus. Consistently with the first interpretation, the decay is slower, yet also very strong, for the French core (not shown here).

The findings of similarities in geometry extend to the semantics of distributions: Tables 1, 2 present the end-portions of sorted lists of words (words were sorted by their PC scores). Table 1 suggests that there are definite semantics associated with each of the first ten PCs: an intuitive impression that is difficult to quantify. Table 2 was prepared as follows. Each French sorted list was subjected to automated translation into English, after which duplicate words in it were removed, and then both lists within each row were truncated to an equal size. Observed semantic correlations are significant.

Table 2. Top portions of sorted lists for each PC: top three dimensions for two languages. Words that are common across cells within each row are typed in boldface.

PC #	English: 8,236 words total core size	French: translated (87,811 words total core size)

1	**increase**, well, rise, support, **accept**, clear, improve, right, continue, direct, good, make, respect, honor, **happy**, secure, order, understanding, **fix**, power, bright, present, definite…	**happy**, agreement, stable, joined together, delighted, approve, net, some, honest, rich, added, **increased**, pleasant, sincere, union, frank, **fix**, favor, praise, optimist, **accept**, abundance, help…
2	**calm**, easy, **soft**, gentle, relaxed, light, ease, **simple**, quiet, soothe, smooth, empty, mild, weak, gently, peaceful, compliant, lenient, pale…	**calm**, modest, discrete, **simple**, subjected, thin, alleviated, softened, flexible, sober, moderate, **soft**, immobility, measured, silence, humble, reserved, simplicity, obeying
3	start, **open**, fresh, begin, **release**, original, new, reveal, speed up, **free**…	**release**, deliver, freedom, yield, **open**, leave, **free**, disencumbered, discovered, dispersion, broad…

2.1. Main Result: Cross-Language Semantic Consistency of the Cognitive Map Structure

Each of the three top portions of paired lists shown in Table 2 has at least three words in common with its counterpart. All words within each row have similar or closely related semantics. Semantic similarities within and across columns of the table seem to be at the same level of strength; however, an objective measure would be necessary to quantify this impression. How can we estimate the statistical significance of co-occurrence of the same words in top portions of two lists in each row of Table 2? Here is one easy way to estimate p-values from above. Given the size of the English core, and assuming that each French-to-English translation is a "blind shot" into the English core (null-hypothesis), we can estimate the probability to find one and the same word in top-twelve portions of both lists: $p \sim 2*12*12 / 8{,}236 = 0.035$ (we included the factor 2, because there are two possible ways of aligning the lists with respect to each other[4]). Therefore, the p-value of the case of word repetition that we see in Table 2 is smaller than 0.035, at least. In conclusion, we have found significant correlations among sorted lists across languages for each of the three PCs. It is also remarkable that there are no common words shared by any two rows in Table 2.

2.2.Testing the Constructed Cognitive Map: Synonym Selection with a Bias

Our analysis so far was focused on extremities of the emergent cognitive maps. It is natural to ask whether words are organized consistently in the middle of a cognitive map. To address this question, a simple experiment was conducted. The algorithm was the following. (1) Select a word from the core: e.g., "*problem*". (2) List all its synonyms. (3) Sort them along a given dimension: e.g., PC #1. (4) Take one word from the top of the list and one word from the bottom. Below is an example of an outcome.

From the top: Problem → exercise.

From the bottom: Problem → obstacle.

Intuitively, the reader would probably agree that "exercise" is a more positive choice than "obstacle": when I am in a positive mood, I am more likely to take a new

[4] The "top end" of each English list was in effect selected at random, but the "top end" of the counterpart French list was selected in order to match the semantics of the English list: here we had two possibilities and selected one of them, for each PC.

encountered problem as an exercise rather than an obstacle. This observation indicates that the cognitive map is consistently organized within itself. How can one quantify the consistency of its internal organization? We address this topic immediately below.

2.3.Analysis by Comparison with Psychometric Data

In order to further validate our findings, we compared our principal dimensions found in the English core against the dimensions of the ANEW dataset: *pleasure*[5], *arousal* and *dominance*. The ANEW list contains 1,034 words, 479 of which were found in the English core. The scatter plot of our PC #1 versus the first dimension of ANEW, which is the mean value of *pleasure*, is represented in Figure 7. The plot shows strong correlation, with similar bimodal distributions in both PC #1 and the ANEW-*pleasure* dimensions. Pearson correlation coefficient $r = 0.70$.

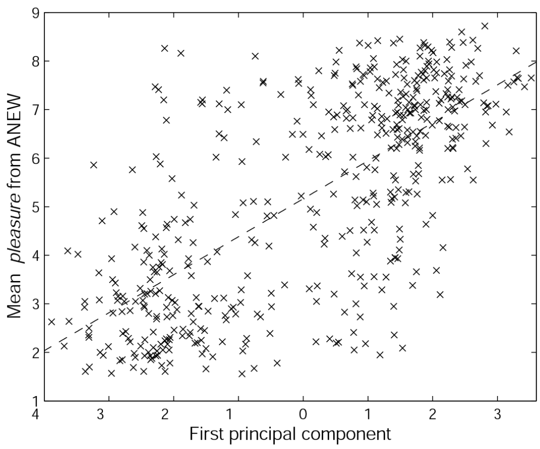

Figure 7. Scatter plot demonstrating strong correlation of PC #1 with the first dimension of ANEW: pleasure. The dashed line is a linear fit. The two clusters ("positive" and "negative") are separated in each dimension.

How can we match PCs with ANEW dimensions? Our correlation analysis shows that PC #1 is the best match (i.e., most highly correlated among all PCs) for ANEW-pleasure, and vice versa ($r = 0.70$, $p = 10^{-70}$). PC #2 is the best match for ANEW-arousal ($r = 0.32$, $p = 10^{-12}$). Finally, ANEW-dominance among all ANEW dimensions is the best match for our PC #3 ($r = 0.25$, $p = 10^{-7}$); however, PC #1 and ANEW-dominance are correlated stronger ($r = 0.64$).

Why do arousal and dominance have so low (albeit significant) correlations with the matching PCs? One possible answer is that semantics of ANEW-arousal and ANEW-dominance is different from the semantics of our PC #2 and PC #3. Indeed, in the given subset of 479 words that are common between the English core and ANEW,

[5] In the ANEW dataset, the first dimension is called "pleasure", while in some studies based on ANEW it is called "valence", consistently with the names of Osgood's dimensions [17].

ANEW dimensions #1 ("pleasure") and #3 ("dominance") are strongly correlated (r = 0.86). On the other hand, our PC #1 and PC #3 are not strongly correlated (r = 0.13), because PCs were calculated using principal component analysis (hence they could be expected to be independent). Debates continue in the literature as to whether *dominance* should be considered an independent dimension [22].

3. Discussion

In the present work, by applying a simple statistical-mechanic model to English and French dictionaries, we constructed multidimensional cognitive maps that capture the semantics of words. We calculated the principal dimensions of the resultant cognitive maps (the PCs) and found their semantics consistent across two languages. In this context it would be interesting to analyze other languages. Our preliminary results of a similar study conducted with a Spanish dictionary of synonyms and antonyms (not reported here) support all our key findings described above, including semantics of the first three PCs.

The principal dimensions that we found appear to be approximately semantically consistent with the previously known dimensions ("affective dimensions" or "Osgood's dimensions") determined psychometrically, in experimental studies with human subjects (e.g. [17, 18] and followers): those studies have revealed that *valence*, *arousal* and *dominance* are the primary dimensions characterizing human values. In this context, it may not be a surprise to learn that modern theoretical models of emotions also use a similar or closely related set of principal dimensions [16, 23]. In the foregoing experimental studies, in contrast with our study, the initial choice of the set of dimensions and respective scoring concepts, was based on subjective reports or psychophysiological measures (SCR, HR, EEG, etc.). In the present work we show that the same or similar dimensions can be found, and a cognitive map of human values can be constructed, without testing human subjects, but using linguistic corpora instead.

At the same time, the match between our PCs and ANEW dimensions labeled by words "arousal" and "dominance" is not perfect. We surmise that semantics of our extracted dimensions cannot be characterized completely by a single word (e.g., "pleasure", "arousal", "dominance") or a pair of antonyms. One may need the entire cognitive map to define semantics of dimensions of this map precisely; however, an approximate definition could be based on a small set of antonym pairs, selected based on their map coordinates, degrees and frequencies. This will be done elsewhere.

Nevertheless, results of comparison with ANEW are surprising. We had no a priori reason to believe that ANEW-dimension #1 (pleasure) and ANEW-dimension #2 (arousal) should correspond to our PC #1 and PC #2. The fact that they do show significant semantic correlations and make best matches with those counterparts is in and by itself intriguing and worth attention. It suggests that the linguistically derived cognitive map dimensions found in this study are not only language-invariant. They appear to be invariant at a broad scale of methodologies, across fields of science, and therefore they are likely to reflect fundamental aspects of human cognition. Extending this logic, one may expect similar results when the same method is applied to other representation systems: corpora, ontologies, databases and indices of various nature, as long as the notions of similarity and contrast can be defined for their elements. The question of whether the same semantics will hold for the first two or three principal dimensions in those cases remains open.

The possibility to construct a complete cognitive map of natural language based on semantic differences and similarities among words opens many important questions. What is the number of independent dimensions of the conceptual value map (cf. [24])? Which of the previous intuitively defined dimensions are orthogonal, and which would be represented as linear combinations of others? What is a canonical choice of a coordinate system, if this notion makes sense, and what would be the semantics of those special coordinates? Would all answers to the above questions be consistent across individuals, social groups, languages – or be individual-specific? Would they be extensible beyond human cognition: to robots and other forms of intelligence? We believe that answering these questions would greatly benefit modern science and technology, because understanding the human value system and the degree of its logical necessity is a key to understanding the laws of cognitive explosion on Earth.

Viewing the findings of the present work in the context of cognitive architecture design, we can imagine an artificial cognitive system that learns or generates new concepts and assigns values to them "on the fly": i.e., in the process of (unsupervised) learning. This may be possible, as soon as the system "knows" how to identify synonyms and antonyms of a new concept among familiar concepts that are already allocated on the cognitive map. Given this strategy, and assuming that the system is capable of cognitive growth[6], we can imagine as the system will gradually develop a personal system of higher values and ideals starting from primitive notions of reward and punishment. This capability could be vital for cognitive systems growing up in social embedding and intended to become human partners.

What else could cognitive maps be used for in a cognitive architecture? Here is an abridged list of their possible functions: filtering of search trees, finding analogies, satisfying semantic constraints, building systems of values for new domains of knowledge, bootstrapping development of higher values and goals, suggesting a reasonable commonsense initiative, classification of memories and strategic retrieval of episodic memories, guidance in imagery and decision making, etc. In addition, linguistic cognitive maps may find applications elsewhere: intuitive Internet search, etc.

In summary, the main finding of this work is the possibility to extract language-invariant dimensions of the human value system from linguistic corpora, using a statistical-mechanic approach. Similar results are expected with other representation systems and databases, including AGI. The principles of map construction, as well as the map itself extracted from the human language, can be used in a general-purpose self-aware cognitive architecture in order to enable its autonomous cognitive growth.

Acknowledgments

We are grateful to Dr. Kenneth A. De Jong for valuable discussions and interest to this work. This work is supported by the DARPA IPTO BICA Grant "Integrated Self-Aware Cognitive Architecture".

[6] Cognitive growth is understood in modern artificial intelligence as multi-level bootstrapped learning, starting from primitive knowledge and gradually developing higher abstract concepts and goals based on previously learned concepts (e.g., [25]).

References

[1] Samsonovich, A. V., & De Jong, K. A. (2003). Meta-cognitive architecture for team agents. In Alterman, R., & Kirsh, D. (Eds.). Proceedings of the 25th Annual Meeting of the Cognitive Science Society (CogSci2003), pp. 1029-1034. Boston, MA: Cognitive Science Society.

[2] Samsonovich, A. V., & De Jong, K. A. (2005). Designing a self-aware neuromorphic hybrid. In Thorisson, K.R., Vilhjalmsson, H., & Marsela, S. (Eds.). AAAI-05 Workshop on Modular Construction of Human-Like Intelligence, Pittsburg, PA, July 10. AAAI Technical Report, vol. WS-05-08, pp. 71-78. Menlo Park, CA: AAAI Press.

[3] Samsonovich, A. V., Ascoli, G. A., De Jong, K. A., & Coletti, M. A. (2006). Integrated hybrid cognitive architecture for a virtual roboscout. In Beetz, M., Rajan, K, Thielscher, M., & Rusu, R.B. (Eds.). Cognitive Robotics: Papers from the AAAI Workshop. AAAI Technical Report, vol. WS-06-03, pp. 129-134. Menlo Park, CA: AAAI Press.

[4] Samsonovich, A. V. (2006). Biologically inspired cognitive architecture for socially competent agents. In Upal, M. A., & Sun, R. (Eds.). Cognitive Modeling and Agent-Based Social Simulation: Papers from the AAAI Workshop. AAAI Technical Report, vol. WS-06-02, pp. 36-48. Menlo Park, CA: AAAI Press.

[5] Tolman, E. C. (1948). Cognitive maps in rats and man. Psychological Review 55 (4): 189-208.

[6] Downs, R. M., & Stea, D. (1973). Cognitive maps and spatial behavior: Process and products. In Downs, R. M., & Stea, D. (Eds.). Image and Environments, pp. 8-26. Chicago: Aldine.

[7] OKeefe, J., & Nadel, L. (1978). The Hippocampus as a Cognitive Map. New York, NY: Clarendon.

[8] O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. Brain Research 34: 171-175.

[9] Samsonovich, A., & McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. Journal of Neuroscience 17 (15): 5900-5920.

[10] Nadel, L. & Hardt, O. (2004). The spatial brain. Neuropsychology 18 (3): 473-476.

[11] Samsonovich, A. V. & Ascoli, G. A. (2005). A simple neural network model of the hippocampus suggesting its pathfinding role in episodic memory retrieval. Learning & Memory 12 (2): 193-208.

[12] Manns, J. R. & Eichenbaum, H. (2006). Evolution of declarative memory. Hippocampus 16 (9): 795-808.

[13] Palmer, S. E. (1999). Color, consciousness, and the isomorphism constraint. Behavioral and Brain Sciences 22 (6): 923.

[14] Gärdenfors, P. (2004). Conceptual Spaces: The Geometry of Thought. Cambridge, MA: MIT Press.

[15] Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. Cognition and Emotion 1: 29-50.

[16] Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. Science 221 (4616): 1208-1210

[17] Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). The Measurement of Meaning. Chicago: University of Illinois Press.

[18] Rubin, D. C. (1980). 51 properties of 125 words: a unit analysis of verbal behavior. Journal of Verbal Learning and Verbal Behavior 19: 736-755.

[19] Barabasi, A.-L., & Bonabeau, E. (2003) Scale-free networks. Scientific American, May 2003: 60-69.

[20] Steyvers, M., & Tenenbaum, J. B. (2005) The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. Cognitive Science 29 (1): 41-78.

[21] Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report C-1. Gainesville, FL: University of Florida.

[22] Stamps, A. E. (2005). In search of dominance: The case of the missing dimension. Perceptual and Motor Skills 100 (2) 559-566.

[23] Rolls, E. T. (2006). Emotion Explained. New York: Oxford University Press.

[24] Powers, M. J. (2006) The structure of emotion. Cognition and Emotion 20 (5): 694-713.

[25] Samsonovich, A. V., Ascoli, G. A., & DeJong, K. A. (2006). Human-level psychometrics for cognitive architectures. In Smith L, Sporns O, Yu C., Gasser M., Breazeal C., Deak G., & Weng J. (Eds.). *Proceedings of the Fifth International Conference on Development and Learning ICDL 2006, CD-ROM,* ISBN 0-9786456-0-X. Bloomington, IN: Dept. of Psychological and Brain Sciences, Indiana University.