A top-down approach for a synthetic autobiographical memory system

Andreas Damianou[1,2], Carl Henrik Ek[3], Luke Boorman[1], Neil D. Lawrence[2], and Tony J. Prescott[1]

[1]Sheffield Centre for Robotics (SCentRo), Univ. of Sheffield, Sheffield, S10 2TN, UK
[2]Dept. of Computer Science, Univ. of Sheffield, Sheffield, S1 4DP, UK
[3] CVAP Lab, KTH, Stockholm, Sweden
andreas.damianou@shef.ac.uk

Abstract. Autobiographical memory (AM) refers to the organisation of one's experience into a coherent narrative. The exact neural mechanisms responsible for the manifestation of AM in humans are unknown. On the other hand, the field of psychology has provided us with useful understanding about the *functionality* of a bio-inspired synthetic AM (SAM) system, in a higher level of description. This paper is concerned with a top-down approach to SAM, where known components and organisation guide the architecture but the unknown details of each module are abstracted. By using Bayesian latent variable models we obtain a transparent SAM system with which we can interact in a structured way. This allows us to reveal the properties of specific sub-modules and map them to functionality observed in biological systems. The top-down approach can cope well with the high performance requirements of a bio-inspired cognitive system. This is demonstrated in experiments using faces data.

Keywords: Synthetic Autobiographical Memory, Hippocampus, Robotics, Deep Gaussian Process, MRD

1 Introduction and Motivation

Autobiographical memory (AM) refers to the ability to recollect episodes from one's experience, relying on organising events and context (semantics) into a narrative. A key task for the intersection of cognitive robotics and biomimetics is to create *Synthetic* Autobiographical Memory (SAM) systems inspired by the current understanding of brain physiology. However, our current knowledge of neural connection formation and activity does not go as far as to enable understanding of how high-level structures, such as semantics, emerge. On the other hand, experimental psychology has provided us with useful understanding about how the *functionality* of AM is organised in "modules" and upon which requirements. Consequently, for practical purposes *top-down* approaches to SAM are developed. These approaches ensure that the known AM requirements are respected and focus on implementing the functional (rather than physiological) AM modules and their interconnections. Any known physiological information is sought to be incorporated in the low-level components of the approach, which implement specific tasks (e.g. pattern completion).

Machine learning (ML) has been used in the past to implement the higher levels of top-down approaches to synthetic physiological systems. However, if ML is used as a "black-box" and purely out of necessity to improve functionality, then consistency in the overall framework is lost. In other words, it is no longer clear whether certain properties of the artificial system emerge due to the low-level bio-inspired components or due to the high-level ML methods. This hinders subsequent evaluation of hypotheses about the system. This paper studies the requirements for enabling top-down approaches to SAM, in a way that functionality is improved (making it usable in a real robotic system) without sacrificing transparency. Subsequently, an existing ML approach, referred to as deep Gaussian processes (deep GPs) [1], is studied here and linked to the SAM framework as a key ingredient of the top-down SAM approach we present. Finally, the results section demonstrates how deep GPs differ from many ML "black-box" approaches (in the context of biomimetics) by enabling uncertainty quantification and intuitive interaction with the model. It is shown that by using a deep GP, not only do we obtain high-level SAM functionality, but we can also recognise individual low-level components of the model as proxies of known sub-functions of AM, such as compression or pattern completion.

2 Requirements for a Top-down SAM System

In [2] the authors recognise the following requirements enabling a biologically inspired SAM system to match the functionality of AM:

– **Compression** of perceived signals in a way that *information* emerges from raw data (recognition of patterns).
– **Pattern separation** to encode different contexts separately and ensure that weak but important signals are not overwhelmed by stronger ones.
– **Pattern completion** for reconstructing events from partial information.

In [2], unitary coherent perception was also added to the requirements of the generic SAM system, although it was highlighted that this selection is suboptimal in practice and choices stemming from Bayesian brain hypotheses [3] are an alternative. In practice, the unitary coherent perception is usually used as a means of avoiding the costly computational requirements associated with full Bayesian inference. However, in a top-down approach that targets improved functionality and transparency, the Bayesian component is vital. This is because the model components are only functional approximations to the true underlying physiological system and, therefore, the Bayesian quantification of uncertainty in our (unavoidably imperfect) representation is important. Deterministic inference is also desired to achieve transparency. In [2] this requirement was fulfilled indirectly through a deterministic Gibbs sampler variant within the unitary coherent perception framework. Here, the intractable full Bayesian inference is approximated with a deterministic *variational* approximation. Based on the above, two more requirements are recognised particularly for the top-down SAM approach:

- **Deterministic inference:** The same starting conditions and parameters should always result in the same outcome.
- **Encoding consistency:** Supervised, semi-supervised and unsupervised learning should be handled using the same representations of memory

3 A Top-down Approach to SAM

3.1 Properties

The top-down SAM approach needs to be robust for embedding in a real robotic system while, at the same time, fulfilling the requirements specified in the previous section, so that connections with expert knowledge from the domain of psychology and neuroscience can be established. For this reason, the approach proposed in this paper comes from a family of models which is:

– **Bayesian probabilistic:** Random variables encode the observables (signal perceived by the robotic agent) and unknowns/latents (internal representation of memories). In a Bayesian framework, prior knowledge (e.g. mammals have legs) can be combined with observations (e.g. past memories) to define *posterior distributions* (e.g. probability that an observed animal is a mammal). The posterior uncertainty is important in practice. For example, an agent operating in a dangerous environment can avoid actions associated with high uncertainty.

– **Latent variable method:** Latent variables correspond to the unknowns in the modelling scenario, and can be inferred from the data. In the approach proposed in this paper, the latent variables are taken to be much simpler and compact (low-dimensional) compared to the high-dimensional observables. By further associating the simple latent variables with the complex, noisy observations we can implement the compression and pattern separation requirements.

– **Generative:** The latent variables are associated with the observables via a *generative mapping function*. This encodes our assumption that the highly compressed latent variables (encoding memory events) should be able to generate fantasy data in the observable domain.

– **Non-parametric:** A non-parametric approach allows to define the generative mapping without having to make crude assumptions about its nature. Memory models built upon artificial neural networks [4] often assume parametric activation functions. Instead, the approach suggested in this paper is based on Gaussian processes (GPs) [5], which learn functional relationships from data with minimal assumptions. For example, fig. 1 shows two functions (posterior processes) learned with the same GP in the presence of different data.

3.2 Gaussian processes

To formalise the above, let us denote the noisy observables as $\mathbf{y}$ and the latent variables as $\mathbf{x}$. A mapping function f relates latent points to observables, so that $\mathbf{y} = f(\mathbf{x}) + \boldsymbol{\epsilon}$; here, $\boldsymbol{\epsilon}$ denotes Gaussian noise, which leads to the Gaussian likelihood $p(\mathbf{y}|\mathbf{f})$, where $\mathbf{f} \triangleq f(\mathbf{x})$ is the collection of mapping function values (noise-free versions of $\mathbf{y}$). To obtain a non-parametric mapping, we place a Gaussian process prior distribution on the mapping function f. Given finite inputs $\mathbf{x}$, this leads to a prior $p(\mathbf{f}|\mathbf{x})$ which is also Gaussian. Thanks to the analytic Bayesian framework, the mapping function values $\mathbf{f}$ can actually be integrated out, to obtain the marginal likelihood (which is again Gaussian):

$$p(\mathbf{y}|\mathbf{x}) = \int_{\mathbf{f}} p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{x}) = \int_{\mathbf{f}} \mathcal{N}\left(\mathbf{y}|\mathbf{f}, \sigma_\epsilon^2\right) \mathcal{N}\left(\mathbf{f}|\mathbf{0}, \mathbf{K}\right). \quad (1)$$

Contrast this with the parametric Bayesian regression approach, which assumes that the mapping function has a fixed form $\mathbf{w}\phi(\mathbf{x})$ parametrised by $\mathbf{w}$, and marginal likelihood $p(\mathbf{y}|\mathbf{x}) = \int_{\mathbf{w}} \mathcal{N}\left(\mathbf{y}|\mathbf{w}\phi(\mathbf{x}), \sigma_\epsilon^2\right) \mathcal{N}\left(\mathbf{w}|\mathbf{0}, \sigma_w^2\right)$. As can be seen, rather than assuming a fixed parametric form and placing a prior on the parameters, in the Gaussian process framework we place the prior directly on the mapping function. More details on GPs can be found in [5].

Notice that so far we have assumed that the latent points $\mathbf{x}$ are known. In the approach taken in this paper, the latent variables are unknown. The Gaussian process latent variable model [6] handles the unknown latent variables by placing a prior on them and optimising the new objective $p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ also w.r.t $\mathbf{x}$.

Inducing Point Representations. In a non-parametric model, the learned quantities (posteriors over the mapping function process and over the latent points) are conditioned on the training data. Being a generative model, this already achieves compression; indeed, the posterior Gaussian process is able to interpolate between the training points for every point in the input domain. This is demonstrated in fig. 1, where the blue solid line gives an estimate for the function in the whole line of real numbers. To achieve further compression via a fixed set of points, one can use inducing point representations. In this case, all the information in the set of pairs $\{\mathbf{x}, \mathbf{f}\}$ is compressed through a smaller set of pairs $\{\mathbf{z}, \mathbf{u}\}$ that remains constant in size as the training dataset grows. This is achieved by replacing the original Gaussian prior $p(\mathbf{f}|\mathbf{x})$ with a *sparse* prior $p(\mathbf{f}|\mathbf{u}, \mathbf{x}, \mathbf{z})$ which depends on the inducing points. This is demonstrated in fig. 1.

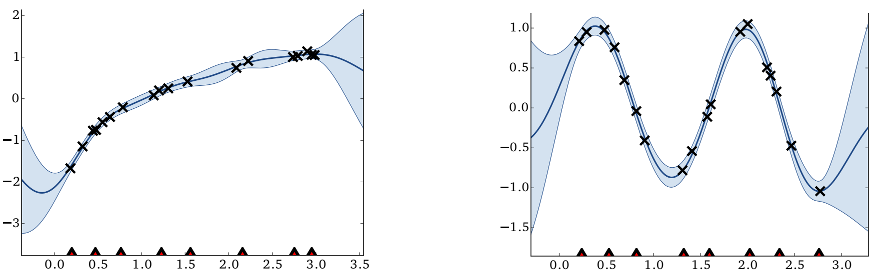


Fig. 1. The same GP prior combined with two different sets of observations (black x's) to obtain the posterior processes (blue solid line) and posterior uncertainty (shaded area). Triangles along the x axis indicate the position of the inducing inputs, $\mathbf{z}$.

3.3 Top-down SAM Architecture.

In this paper the focus is on developing a top-down SAM approach the functionality of which is inspired by physiology. In particular, deep probabilistic approaches have been linked in the past with functionality that can be observed in the human brain [7]. For example, human vision involves a hierarchy of visual cortices (V1 – V5) which process visual signals in a progressive manner [8]. Evidence that this kind of hierarchical learning is performed in the brain areas associated with memory is not yet existent; the exact localisation and functionality of the billions of neural interconnections associated with memory is yet unknown and, therefore, making a neural simulation is currently impossible. Instead, the top-down approach to SAM works on a higher level and seeks to simulate the organisation of *functionality* (rather than the organisation of neurons) into hierarchically structured (sub)modules. As such, the designed architecture involves high level modules grouped into a core and a set which extends beyond the core of the SAM, as can be seen in fig. 2. These modules are explained below.

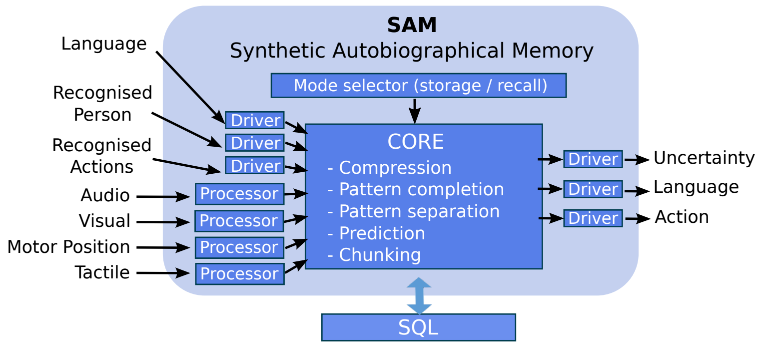


Fig. 2. The developed Synthetic Autobiographical Memory (SAM) system. This paper focuses on the core module, which is implemented using deep Gaussian processes [1].

Deep Gaussian Processes. To start with, the *SAM core* relies on a set of latent variables which encode memory events in a compressed and noise-free space. The latent variables are part of a deep Gaussian process model (deep GP) [1]. A deep GP is the hierarchical extension to a standard GP. Instead of having a single set of latent points, $\mathbf{x}$, we now have a hierarchy: $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_L$, where L denotes the number of layers. Every layer $\mathbf{x}_\ell$ is linked to its previous layer through a mapping function f_ℓ with a GP prior, so that $\mathbf{x}_\ell = f_\ell(\mathbf{x}_{\ell-1})$. In other words, the observed layer is successively processed by L non-parametric functions, so that each function operates on the already processed output of the previous in the hierarchy. Importantly, the intermediate latent spaces are available for inspection, revealing intuitive features. Fig. 3 demonstrates this process for the task of recognising handwritten digits; samples can be drawn from the latent spaces in each layer, to reveal features that successively encode more abstract information due to the successive processing.

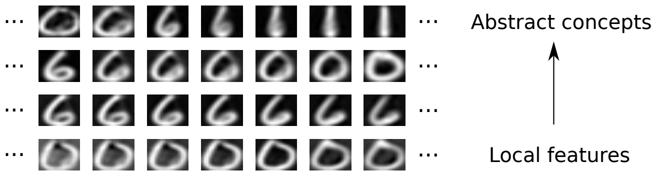


Fig. 3. Samples from the hierarchy of the latent spaces discovered for a collection of handwritten digit images. The lowest layer encodes very local features (e.g. if the circle in a zero is closed or not), but successive processing allows the top layer to encode abstract information, such as general characteristics of different digits.

Inference in deep Gaussian processes is not analytically tractable straightforwardly. This is because the model is required to marginalise over the latent representation $\mathbf{x}$ so as to obtain a posterior over it through the Bayes rule: $p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}$. The intractability in the denominator requires approximate solutions. In deep GPs, inference proceeds through a variational framework. In contrast to stochastic inference approaches like sampling and MCMC, a variational inference approach is deterministic [9]. This means that the same starting conditions (e.g. initialisation of parameters) will always result in the same approximation of the quantities of interest (posterior distributions, inducing points, latent representation). Therefore, the requirement for deterministic inference is fulfilled when deep GPs are used within the SAM core.

Multiple Modalities. In the SAM framework, multiple representations of the same event must be taken into account consistently. Consider e.g. the separate signals (visual, audio) associated with a memory from watching a theatrical play. However, there is some commonality (specific scenes are associated with specific sounds). Formally, assume that $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(M)}$ represent the segmentation of the observables into M different modalities. These can be accounted for in the latent variable framework by maintaining for all modalities a single $Q-$dimensional latent space (i.e. a single representation compressed in Q features). Subsequently, learning which parts of the whole latent space are relevant for which modality is achieved by optimising a set of *relevance weights* $\mathbf{w}^{(m)} \in \Re^Q$ for each modality. If, for example, $w_5^{(2)}, w_5^{(3)} \neq 0$, this means that the latent dimension 5 encodes information for views 2 and 3, thus avoiding redundancy and achieving compression. This idea was developed in [10]. This approach can be embedded in the deep GP framework of the SAM core, and is demonstrated in the next section.

Modules Outside of the SAM Core. As can be seen in fig. 2, the top-down SAM architecture links inputs to the SAM core through drivers and processors. A processor is a module that allows raw stimuli to be pre-processed. This is not a compulsory requirement, since the SAM core achieves compression through the deep GP. However, this allows the usage of sophisticated feature extraction methods (e.g. SURF features [11] from raw images) as a pre-processing step. On the other hand, already processed information is incorporated into the SAM through drivers. A driver is a module which is specific to the input/output it is responsible for, and is employed to "translate" the highly structured information into a language understandable by the SAM. For example, language/actions in the context of social interaction with the robotic agent [12, 13] can be represented as a set of frequencies of terms from a pre-built dictionary. On the other hand, recalling an action involves another driver which translates the memory into a series of motor commands. The current implementation accompanying this paper only contains the appropriate processors and drivers for visual stimuli.

Tasks and Consistency. The SAM model operates in a supervised as well as unsupervised scenario. In the supervised scenario, the latent representation learned by the internal deep GP model is guided through additional input information, $\mathbf{t}$, expressed through a prior, that is, $p(\mathbf{x}|\mathbf{t})$. For example, $\mathbf{t}$ might be the time-stamp of a particular frame associated with a visual stimulus and would force the latent representation to form a smooth time-series . Unsupervised learning corresponds to the scenario where the latent representation is learned in an unconstrained manner, only from the data. In this case, the latent representation is assigned a fairly uninformative prior $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mathbf{0}, \mathbf{I})$. Semi-supervised learning can also be handled by following a data-imputation approach [14]. Overall, the suggested SAM approach satisfies the encoding consistency requirement.

Related Work. Related work involves methods inspired by low-level neural structures (e.g. temporal codes [17]) and "traditional" bio-inspired but high-level deep learning approaches, e.g. convolutional neural networks [18]. In particular, the latter have achieved remarkable results in many vision tasks, but do not always provide transparent/manipulable representations as required for a SAM.

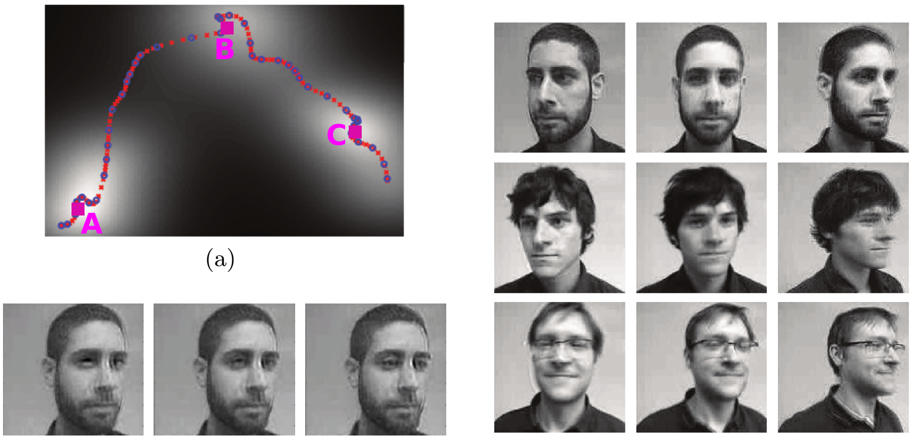
4 Demonstration Using Human Faces Data

This section demonstrates a selection of representative results from the top-down SAM system. Additional results can be seen at `https://youtu.be/rIPX3CIOhKY`. Software for reproducing the results is available at: `http://git.io/vTBMt`.

The employed machine learning approach has been previously demonstrated quantitatively in classification tasks [10]. In contrast, the focus of this paper is to demonstrate the emergence of ABM functionality through our interpretation of the model components, in particular the latent and the inducing points which compress the perceptual information from multiple event modalities.

4.1 Face Rotations Experiment

For the first demonstration, images of 3 subjects were captured using a standard mobile phone's low-resolution camera (140×140 pixels per image). For each subject, multiple images were recorded under different rotations of the face with respect to the camera. To demonstrate the method in imperfect data, the images were collected while the camera was held by hand and the subjects were rotating on their own; no further processing was made on the data (e.g. no cropping). 250 images of each subject were stored in three matrices, $(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{Y}^{(3)})$, so that each triplet of rows, $(\mathbf{y}_n^{(1)}, \mathbf{y}_n^{(2)}, \mathbf{y}_n^{(3)})$, corresponds to the three faces under a similar rotation. These data were presented to a top-down SAM system. The Gaussian processes used 45 inducing inputs (see end of section 3.2) and a $Q = 20-$dimensional latent space, much smaller than the original output dimensionality $(140 \times 140 = 19,600)$. The inducing points and latent space together achieve strong compression and chunking of the original signal.



(a) (b) (c)

Fig. 4. Results from the rotating faces experiment. Fig. (a) depicts the projection of the internal SAM representation on the two dimensions shared for all modalities. Background intensities correspond to the variance of the distribution for predicting $\mathbf{y}$ from $\mathbf{x}$. Fig. (c) shows the corresponding outputs generated by conditioning on the selected locations shown as A, B, C in fig. (a). Fig. (b) shows outputs generated by conditioning on latent locations which encode weak but highly descriptive signal.

The internal SAM representation of the raw visual signal was obtained after a training phase, required for tuning the parameters of the core's model. Next, this internal representation was investigated, to understand the way in which weak and strong signals are chunked and how low-level (subtle, e.g. blinking) and high-level (e.g. face characteristics) concepts emerge automatically. Out of the 20 features used to compress the observed signal, fig. 4(a) depicts two (one plotted versus the other) which were deemed important by all three modalities. Red crosses correspond to latent points $\mathbf{x}_n$ which, in turn, correspond to observations in each of the three modalities, $(\mathbf{y}_n^{(1)}, \mathbf{y}_n^{(2)}, \mathbf{y}_n^{(3)})$. The SAM system successfully recovers a semi-circular shape, corresponding to the rotation of the faces from 1 to 180 degrees. Importantly, this information is encoded once for all three modalities and is learned automatically from the given dataset. Blue circles represent inducing points which further compress the internal representation. The optimisation procedure spreads the inducing points nicely along the latent path. Notice that the compressed discrete representation obtained with the latent and inducing points can be conditioned upon to perform inference for *any* possible area in the latent space. This demonstrates the real power of the model in terms of performing **compression** and **pattern completion**. Specifically, the background intensity of fig. 4(a) represents the variance associated with the distribution, where bright intensities correspond to areas where the SAM model is confident in its predictions. To demonstrate this, predictions were made for the depicted points A, B, C, obtaining the outputs in column 1, 2 and 3 in fig. 4(c) (each row corresponds to one modality). In other words, given specific areas in the compressed representation, the SAM model generated outputs in the original space of the signal. As can be seen, all three modalities are consistent in the rotation, that is, the memory associated with the concept "rotation" was recognised (**chunking**) and compressed for all three faces into the two-dimensional space of fig. 4(a). Finally, fig. 4(b) depicts outputs obtained by conditioning on latent space dimensions that were a) deemed important for only the first of the modalities and b) encoded signal which was very weak in the original output space (images). This signal evidently corresponds to blinking, and the fact that this weak signal is not overwhelmed by the stronger signal is a demonstration of successful **pattern separation** achieved by the SAM system.

4.2 Light Angle and Morphing Experiment

The second experiment used slightly larger face images, 269×186 pixels each. One image was recorded for each of the 6 considered subjects and then processed to simulate illumination under one out of 42 different light source positions around the face, similarly to [15]. The total set of 6×42 images was then split into two groups (modalities) $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ where: a) each row $\mathbf{y}_n^{(m)}$ corresponds to an image in modality m; b) group $\mathbf{Y}^{(1)}$ contains images only from subjects 1,2,3 and $\mathbf{Y}^{(2)}$ only from 4,5,6; c) the rows of the two matrices were aligned, so that $\mathbf{y}_n^{(1)}$ and $\mathbf{y}_n^{(2)}$ are matched in the angle of the light source (but the ordering of the subjects is arbitrary, i.e. not matched). In other words, the two modalities were created such that the illumination condition is a common signal and the face identity is signal private (specific) to each of the two modalities. The challenge is for the SAM system to compress the data by also encapsulating this information.

Fig. 5 depicts the results. Fig. 5(a) depicts a bar graph of the relevance weights corresponding to each of the two modalities and are of the same dimensionality as the latent space ($Q = 14$). Thick/blue bars correspond to modality 1 and red/thinner bars to modality 2. Dimensions 1,2,5 encode information for both modalities. To verify this, fig. 5(b) plots dimension 1 versus 2. As can be seen, the SAM system successfully mapped the information for the light source

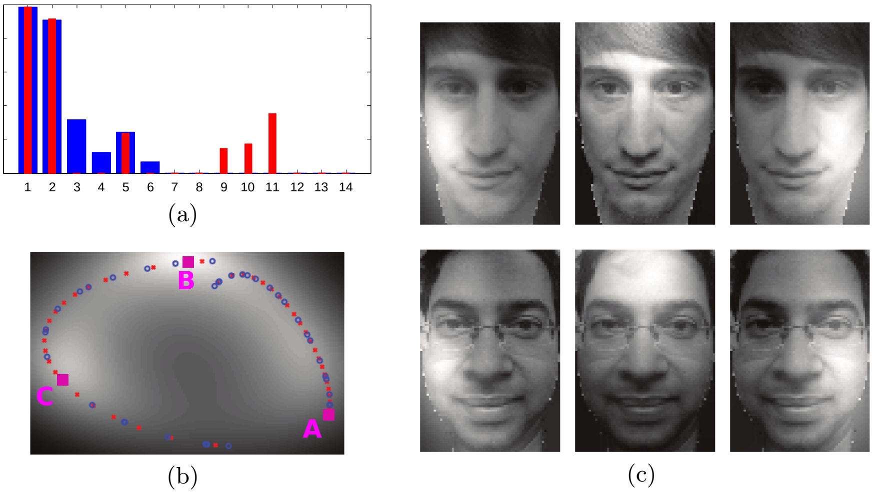
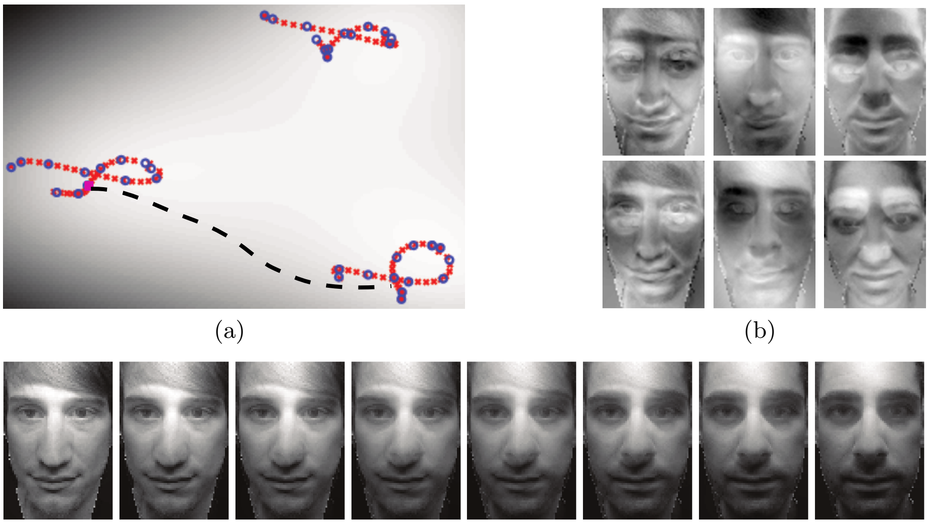


Fig. 5. Second experiment: (a) depicts the relevance weights optimised for the two modalities; (b) shows the compressed representation of the common signal; (c) depicts the outputs obtained by conditioning on the locations depicted as A,B,C in (b).

position from the original $50,034$ dimensional space to only two dimensions. Indeed, by conditioning on the latent space locations indicated by A, B and C we obtain the outputs in columns 1,2,3 respectively of fig. 5(c), which depict faces under the same illumination condition with no other features changed.

On the other hand, one can also perform the same procedure for dimensions taken from the sets $(3, 4, 6)$ or $(9, 10, 11)$ which, from fig. 5(a), is obvious that they are relevant to only one of the modalities. In particular, fig. 6(a) depicts the corresponding internal representation of dimensions $3, 4$, which encode signal relevant only to the first modality. This plot demonstrates the successful **chunking** achieved by the SAM system, since the three clusters which were automatically discovered correspond to each of the three faces contained in modality 1. Again, the inducing points (blue circles) nicely cover each cluster and do not fall in between clusters, thereby using the full compressing capacity of the model. To verify these intuitions, 8 locations were selected from this space (the rest of the dimensions were kept fixed) along the path depicted as a black dotted line in fig. 6(a). Notice that this scenario is different than the procedure followed so far in the experiments, in that the selected latent locations are interpolations between those corresponding to training points. This procedure results in the images depicted in fig. 6(c). The morphing effect verifies the intuition that this part of the compressed space is responsible for encoding face characteristics, and manifests **pattern completion** by producing *novel outputs*. Finally, fig. 6(b) depicts the "fantasy" memories used as a compressed basis, computed as eigenfaces [16] from the inducing output posterior. For example, large variance is observed around the eye area (reflecting the changes in the images of fig. 6(c)) for the two male faces, and around the eyebrows and nose for the female face.



(a) (b)

(c)

Fig. 6. Morphing effect obtained by sampling outside of the training compressed representations' region (black line, fig. (a)) to obtain novel outputs (c). The top row of fig. (b) depicts three of the "fantasy" memories used as a compressed basis (inducing outputs' eigenfaces). Bottom row is just the color-inverted version of the top row. This reveals that the plotted eigenvectors define the tangent direction for interpolating between faces.

5 Discussion and Future Work

Previous work in psychology and bio-inspired robotics has resulted in extracting high-level descriptions of the organisation of modules in a SAM system. This paper discussed a top-down approach to SAM where the aforementioned high-level descriptions are guiding the system architecture while the specific (unknown) details of each component are abstracted. This is made possible through a flexible representation of memories, based on Bayesian latent variable models [1, 10] which filter all functionality through a smaller set of learned variables (inducing points). Experiments on "noisy", real-world faces data revealed the robustness of the method in learning powerful representations of the data (simulating *memory formation*), while structured interaction with the framework allowed for examining its properties with respect to requirements for a biologically inspired SAM.

Future work will aim at integrating the SAM system into the cognitive component of a robot, such as the iCub. Although preliminary experiments with regards to handling auditory streams have been performed, a more complete solution which handles heterogeneous sensory data is planned for the future. Finally, we will work towards achieving stronger connections with biology, by incorporating more detailed bio-inspired structure in the lowest levels of the top-down architecture (e.g. through priors and constraints on the inducing points).

Acknowledgements. This research was funded by the European research project EU FP7-ICT (Project Ref 612139 "WYSIWY"). We thank the following for participating in the creation of the faces dataset: F. Yousefi, A. Saul, J. Gonzalez, M. Zwiessele, M. A. Rahman, Z. Dai, R. A. Pacheco. We also thank our colleagues at Sheffield Robotics and, in particular, Uriel Martinez-Hernandez.

References

1. Damianou, A., Lawrence, N.: Deep Gaussian Processes. Proceedings of the 16th International Workshop on A.I. and Statistics (AISTATS), pp. 207–215 (2013)
2. Evans, M.H., Fox, C.W., Prescott, T.J.:Machines learning - towards a new synthetic autobiographical memory. In Proceedings Living Machines (Milan) (2014)
3. Pouget, A., Beck, J.M., Ma, W.J., Latham, P.E.: Probabilistic brains: knowns and unknowns. Nature neuroscience 16(9), 11701178 (2013)
4. Rojas, R.: Neural networks: a systematic introduction. Springer Science & Business Media (1996)
5. Rasmussen, C. E., Williams, C. K. I.: Gaussian processes for machine learning. Cambridge, MA: MIT Press, (2006)
6. Lawrence, N. D.: Probabilistic non-linear principal component analysis with Gaussian process latent variable models. Journal of Machine Learning Research, 6, 17831816. (2005)
7. Bengio, Y., LeCun, Y.: Tutorial on Learning Deep Architectures. Videlectures.net `videolectures.net/icml09_bengio_lecun_tldar/`, June 2009
8. Nielsen, M. A.: Neural Networks and Deep Learning. Determination Press (2015)
9. Bishop, C. M.: Pattern Recognition and Machine Learning. Springer-Verlag, ISBN 0387310738. (2006)
10. Damianou, A., Ek, C.H., Titsias, M., Lawrence, N.: Manifold Relevance Determination. Proceedings of the 29th International Conference on Machine Learning (ICML), pp. 145–152. Omnipress, New York (2012)
11. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. Computer Vision–ECCV, pp. 404-417. Springer (2006)
12. Pointeau, G., Petit, M., Dominey, P.F.: Embodied simulation based on autobiographical memory. Biomimetic and Biohybrid Systems, pp. 240-250. Springer (2013)
13. Pointeau, G., Petit, M., Dominey, P.F.: Successive Developmental Levels of Autobiographical Memory for Learning Through Social Interaction. Autonomous Mental Development, IEEE Transactions on 6, 200-212 (2014)
14. Damianou, A., Titsias, M., Lawrence, N.: Variational inference for uncertainty on the inputs of Gaussian process models. arXiv preprint, arXiv:1409.2287 (2014)
15. Georghiades, A.S., Belhumeur, P.N., and Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence, 23(6) (2001)
16. Turk, M., Pentland, A. Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1), 71-86 (1991)
17. Luvizotto, A., Renn-Costa, C., Verschure, P. A Framework for Mobile Robot Navigation Using a Temporal Population Code. Biomimetic & Biohybrid Systems (2012)
18. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological cybernetics, 36(4), 193-202 (1980)