Emotional Concept Development

Claes Strannegård[1,2(✉)], Simone Cirillo[2], and Johan Wessberg[3]

[1] Department of Philosophy, Linguistics and Theory of Science,
University of Gothenburg, Gothenburg, Sweden
claes.strannegard@gu.se

[2] Department of Applied Information Technology,
Chalmers University of Technology, Göteborg, Sweden
simone.cirillo@alumni.chalmers.se

[3] Institute of Neuroscience and Physiology,
University of Gothenburg, Gothenburg, Sweden
johan.wessberg@gu.se

Abstract. Artificial emotions of different varieties have been used for controlling behavior, e.g. in cognitive architectures and reinforcement learning models. We propose to use artificial emotions for a different purpose: controlling concept development. Dynamic networks with mechanisms for adding and removing nodes are more flexible than networks with a fixed topology, but if memories are added whenever a new situation arises, then these networks will soon grow out of proportion. Therefore there is a need for striking a balance that ideally ensures that only the most useful memories will be formed and preserved in the long run. Humans have a tendency to form and preserve memories of situations that are repeated frequently or experienced as emotionally intense (strongly positive or strongly negative), while removing memories that do not meet these criteria. In this paper we present a simple network model with artificial emotions that imitates these mechanisms.

Keywords: Autonomous agent · Concept development · Emotion

1 Introduction

One strategy toward artificial general intelligence (AGI) uses mathematical methods developed without regard to natural intelligence [23]. A second strategy imitates the mechanisms of human psychology [3,18]. A third tries to simulate the human brain at the neural level – as attempted in the BRAIN Initiative and the Human Brain Project. A fourth tries to imitate computational mechanisms that are present in nervous systems across the animal kingdom [1,6].

Bees have less than a million neurons in their brains, yet they are able to learn new concepts with the help of reward and punishment and adapt to a wide range of environments [9,24]. Bees are arguably more flexible and better at adapting to new environments than present-day AI systems, so it might be possible to create more flexible AI systems by mimicking certain of their computational mechanisms.

J. Bieger (Ed.): AGI 2015, LNAI 9205, pp. 362–372, 2015.
DOI: 10.1007/978-3-319-21365-1_37

In this paper, we present a simple graphical model for network-based computation and an algorithm for developing such networks – one that uses emotional factors to guide their development. Thus we tackle the problem of novelty-driven concept-formation, which easily leads to explosive memory formation [20]. Although our model was inspired by mechanisms described in neuroscience, we have made no attempt to model any particular biological system. Because our research focus is on AGI, we have felt free to mix biologically realistic features with more strictly pragmatically motivated ones.

Bees are capable of forming memories, reflecting capacities that cannot possibly be innate. In one revealing study [9], bees learned to differentiate between vowels and consonants of the Latin alphabet with the help of bowls of water containing or not containing sugar – placed next to the letters. The bees learned the two concepts robustly despite large variations in color, font, size, and mode of presentation. Bees have only about 950,000 neurons in their brains, implying that they can only form a limited number of memories [24]. This raises the obvious question of which memories would be most useful from the perspective of survivability.

Contemporary research has emphasized the importance of emotions to memory formation [13] and the role of emotional systems in decision making [4]. Sensory events can trigger reward signals (e.g., food) or indicate danger (e.g., an approaching predator). The emotional circuits receive sensory information from both lower and higher (i.e., cortical) levels; in mammals, they include the amygdala (for punishment) and the dopaminergic and opioid systems – such as the ventral tegmental area and periaqueductal gray (for reward). Their activation affects memory formation via several mechanisms: e.g., by directing attention towards the stimulus and then activating the brain's arousal systems [12]. Emotions can act directly on memory circuits in the hippocampus to sort more from less relevant memories: so-called emotional tagging [17]. It has recently been shown how repetitive or iterative mechanisms for memory formation – the classical Hebbian view – interact critically with emotion-driven mechanisms in the formation of behaviorally useful long-term memories [10].

Automatic-concept-formation techniques have been used for categorization [16,21], clustering [11], and automatic theorem proving [8]. Blum and colleagues [7] survey several concept-formation techniques for machine-learning. Concept formation is a central component of such cognitive architectures as Sigma [18] and MicroPsi [2].

Concept formation finds a close statistical analogue in learning the structure of graphical models [19]: e.g., variable-order Markov models (VMMs: see [5]), which can be used for sequential prediction. The main difficulty with learning such models is discovering which parts of the past are useful for predicting the future. VMMs make predictions based on variable-length history windows; they are very efficient to learn, given that they can be described in terms of non-parametric tree distributions. Consequently, VMMs – and other tree models – have been used in reinforcement learning for some time. One of the first successful models was the U-tree [15], which adds leaf nodes to a VMM tree only when the new nodes' utility predictions are statistically different from the current ones. This and similar models are not limited to sequential partitions of observations: it is possible to generate trees using an arbitrary metric, to compare histories [22] within a fully Bayesian framework.

Marsella and colleagues [14] survey computational models of emotion, including models based on appraisal theory; while Bach [2] offers a framework for modeling emotions.

Section 2 presents our network model and Section 3 describes computations in such models. Section 4 offers an algorithm for developing these networks automatically. Section 5 presents results. Section 6 draws some preliminary conclusions.

2 Transparent Networks

Definition 1 (Network). *A (transparent) network is a finite, labeled, directed, and acyclic graph* (V, E) *where nodes* $a \in V$ *may be labeled:*

- $SENSOR_i$*, where* $i \in \omega$ *(fan-in 0)*
- $MOTOR$ *(fan-in 1, fan-out 0)*
- AND *(fan-in 2)*
- OR *(fan-in 2)*
- $DELAY$ *(fan-in 1)*
- $REVERB$ *(fan-in 1)*

The fan-in and fan-out conditions in parentheses are restrictions on E*. Each* $(a, b) \in E$ *has an associated weight* $w(a, b) \in [0, 1]$*.*

Nodes labeled $SENSOR_i$ model sensors of modality i. $SENSOR_i$ could e.g. model a receptor cell with ion channels sensitive to cold temperature, mechanical pressure, or acidity. Nodes labeled $MOTOR$ model muscle-controlling motor neurons. Nodes labeled AND and OR model nerve cells with high and low thresholds respectively. Nodes labeled $DELAY$ model nerve cells that retransmit action potentials with a delay. Nodes labeled $REVERB$ model nerve cells or nerve-cell clusters that stay active (i.e., reverberate) for some time after they have been excited. Figure 1 provides example networks. Note that some nodes that appear in figures throughout this paper have labels that do not appear in Definition 1. They represent sensors or more complex networks computing the concept indicated by the label.

3 Network Computation

Definition 2 (Stimulus). *Let* $G = (V, E)$ *be a network and let* $S(V)$ *consist of the sensors of* V*, i.e. those nodes that are labeled* $SENSOR_i$*, for some* i*. A stimulus for* G *is a function* $\sigma : S(V) \rightarrow \{0, 1\}$*.*

Stimuli model the presence or absence of action potentials on receptors.

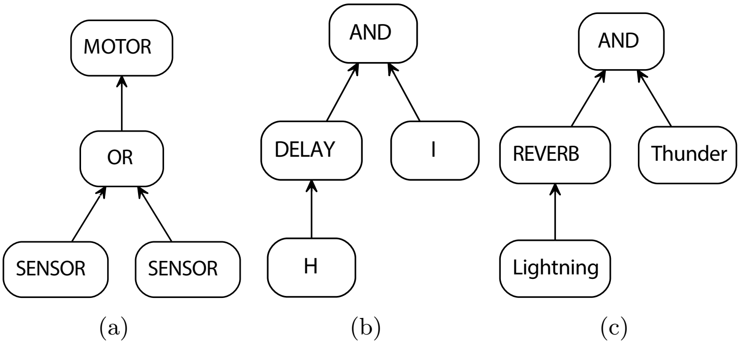

Fig. 1. Examples of transparent networks. (a) The tentacle of an anemone that retracts upon being touched. (b) The letter *H* immediately followed by the letter *I*. (c) Lightning followed by thunder (within ten time steps of the system).

Definition 3 (Input Stream). *Let $G = (V, E)$ be a network. An input stream for G is a sequence $\sigma_1, \sigma_2, \ldots$, where each σ_i is a stimulus for G.*

Input streams give rise to two types of activity that propagate through the networks: perception and imagination. We chose to model perception and imagination separately, thus distinguishing clearly between exogenous perception and endogenous imagination.

Definition 4 (Time). *Let T be the set of natural numbers, modeling time.*

Input streams give rise to two types of activity that propagate through the networks: perception and imagination. We chose to model perception and imagination separately, thus distinguishing clearly between exogenous perception and endogenous imagination.

Definition 5 (Perception). *Let $G = (V, E)$ be a network and let $L(a)$ be the label of node $a \in V$. The perception $p_G : V \times T \to \{0, 1\}$ generated by the input stream $\sigma_1, \sigma_2, \ldots$ is defined as follows. Let $p_G(a, 0) = 0$ for all $a \in V$. Let*

$$p_G(a, n+1) = \begin{cases} \sigma_{n+1}(a) \textit{ if } L(a) = SENSOR_i \\ p_G(a', n+1) \textit{ if } L(a) = MOTOR, (a', a) \in E \\ \min\{p_G(a', n+1) : (a', a) \in E\} \textit{ if } L(a) = AND \\ \max\{p_G(a', n+1) : (a', a) \in E\} \textit{ if } L(a) = OR \\ p_G(a', n) \textit{ if } L(a) = DELAY, (a', a) \in E \\ 1 \textit{ if } L(a) = REVERB, (a', a) \in E, \exists n' \in [n-10, n] p_G(a', n') = 1 \\ 0 \textit{ if } L(a) = REVERB, (a', a) \in E, \nexists n' \in [n-10, n] p_G(a', n') = 1 \end{cases}$$

Given a certain input sequence, node a is *active* at step n in G if $p_G(a, n) = 1$. A DELAY node is active at n *iff* its parent node was active at $n - 1$. A REVERB node is active at n *iff* its parent node was active at some point during the last ten time steps. Figure 2 offers examples of perception, where perceptual activity is indicated by boldface node borders.

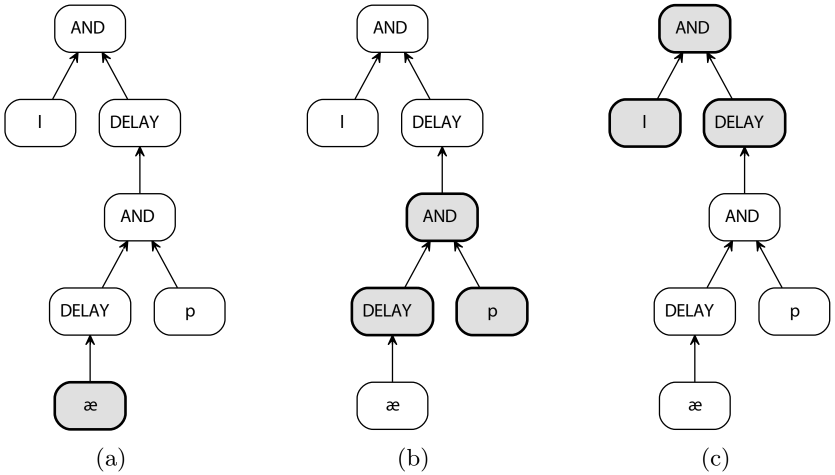

Fig. 2. Propagation of perception. The phonetic sequence [æpl] is perceived in three consecutive steps.

Definition 6 (Imagination). *Imagination $i : V \times T \rightarrow [0,1]$ is defined as follows. Let $i(a,n) = max\{p(a',n) \cdot w(a',b,n) : E(a',b)$ and $E(a,b)\}$, where $w(a',b,n)$ is the label on edge $(a',b) \in E$ at time n.*

Figure 3 offers examples of imagination, where imagination is indicated by dashed-line node borders. The darker the interior of the node, the more intense the imagination.

4 Network Development

Next, we define the network-development mechanism that generates a sequence of networks $G_0, G_1, \ldots$ from input stream $\sigma_0, \sigma_1, \ldots$ and initial network G_0. The initial graph G_0 is called the *genotype*; all graphs G_{n+1} are *phenotypes*. For each n, G_{n+1} is obtained either by extending G_n or trimming G_n. As in natural nervous systems, activity continues to flow in the networks while they are being modified. The definitions of activity propagation can be taken directly from fixed graphs and applied to graph sequences. First, we must introduce some basic concepts pertaining to networks.

Definition 7 (Reward Signal). *A reward signal is a function $r : T \rightarrow [-1,1]$, where $[-1,1]$ is the real interval between -1 and 1.*

Positive reward signals model reward; negative reward signals model punishment.

Definition 8 (Arousal). *Let $arousal(n) = abs(r(n))$, where abs means absolute value.*

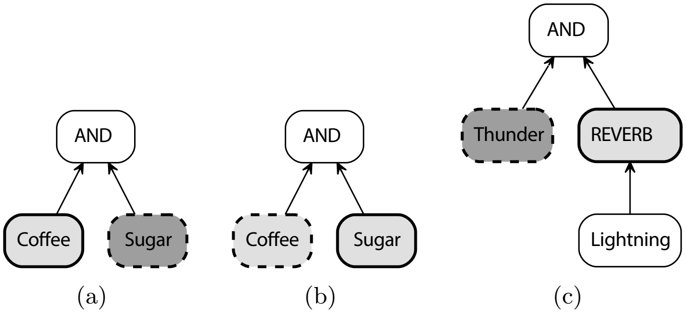

Fig. 3. Propagation of imagination. (a) Perceiving coffee, while imagining sugar strongly. (b) Perceiving sugar, while imagining coffee weakly. (c) Expecting thunder after lightning.

Definition 9 (Birth). *Let $G_0, G_1, \ldots$ be a sequence of networks. Suppose node a appears in some G_i. Then $birth(a)$ is the smallest n such that $a \in G_n$.*

Definition 10 (Relative Frequency). *Let $RF(a,n) = card\{m \in [birth(a), n] : p(a,m) = 1\}/(n - birth(a))$, where card is the cardinality function.*

Definition 11 (Closure). *Let E^* be the reflexive and transitive closure of E.*

Definition 12 (Learning Parameters). *The following parameters regulate the network development process:*

- $p_0 \in \omega$ *(size parameter)*
- $p_1 \in [0,1]$ *(construction parameter)*
- $p_2 \in [0,1]$ *(viability parameter)*
- $p_3 \in [0,1]$ *(destruction parameter)*
- $p_4 \in [0,1]$ *(multimodality parameter)*

Next, we will introduce a number of notions that trigger extensions (14-17) or trimming (18-19) of the network. We begin with a local notion of emotionality.

Definition 13 (Emotionality). *Let $emo(a,n) = avg\{r(n') : p_{G_{n'}}(a,n') = 1 : n' \in [birth(a), n]\}$, where avg means average.*

Example 1. Here are examples of how emotionality might be computed:

- $emo(cake, t) = avg\{0.7, 0.8, 0.3\} = 0.6$
- $emo(snake, t) = avg\{-0.5, -0.7, -0.9\} = -0.7$

Definition 14 (Top Active Node). *Suppose $G = (V, E)$ is a graph and $\sigma_0, \sigma_1, \ldots$ a sequence of stimuli: $a \in V$ is top active in G at n if $p_G(a,n) = 1$ and there is no $b \neq a$ such that $(a,b) \in E^*$ and $p_G(b,n) = 1$.*

Definition 15 (Surprise). *Let $surprise(n) = min\{abs(r(n) - emo(a,n)) : a$ is top active at $n\}$.*

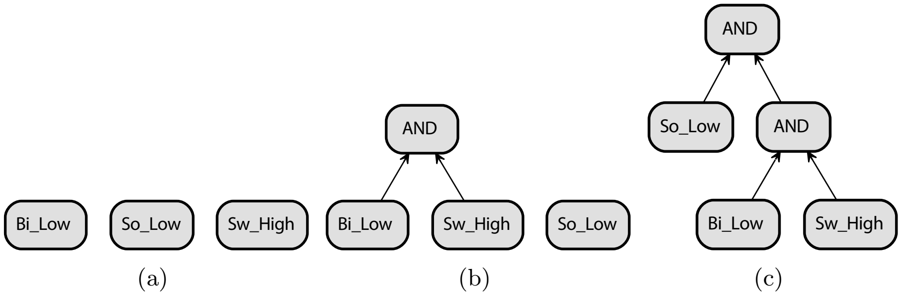

Fig. 4. Unimodal spatial construction: formation of a memory structure for the taste of a certain apple. (a) The sensors for low bitterness, low sourness, and high sweetness are activated. (b) Two of the top active nodes are randomly selected and joined. (c) The only top active nodes are joined.

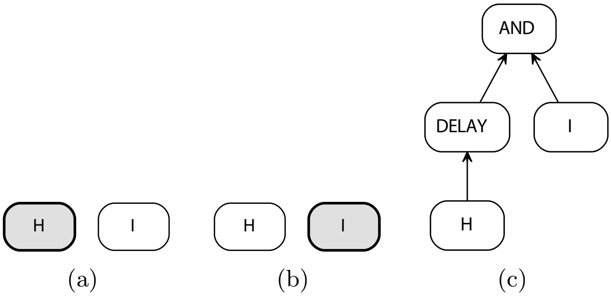

Fig. 5. Unimodal temporal construction: formation of a memory structure for the written word "HI" takes place in three steps.

Definition 16 (Learning Rate). *Let $LR(n) = p_1 \cdot surprise(n) + (1 - p_1) \cdot arousal(n)$.*

Definition 17 (Modality). *Suppose $G = (V, E)$ is a network and $b \in V$. Modality $mod(b, G)$ is defined as $\{i : E^*(a, b)$ and $L(a) = SENSOR_i\}$.*

Definition 18 (Emotional Importance). *Let $EI(a, n) = max\{abs(emo(b, n)) : E^*(a, b)\}$.*

Definition 19 (Viability). *Let $via(a, n) = p_2 \cdot EI(a, n) + (1 - p_2) \cdot RF(a, n)$.*

Finally we are ready to introduce our operations: Two that extend the networks and one that trims them.

Definition 20 (Spatial Construction). *Suppose $G = (V, E)$ is a graph and $a, b \in V$. Then $spatial(G, a, b)$ is the graph (V', E'), where $V' = V \cup \{c\}$, c is a new node labeled AND, and $E' = E \cup \{(a, c), (b, c)\}$, with the weights of both new edges set to 1.*

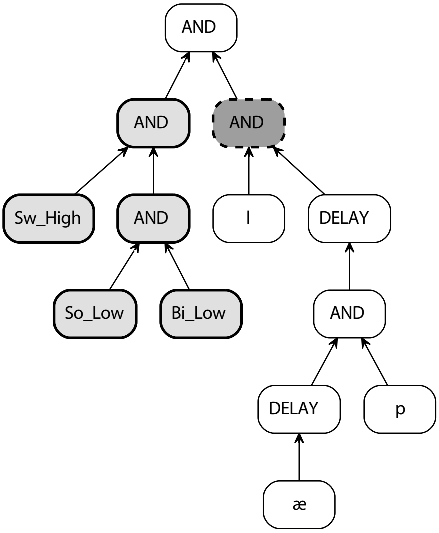

Fig. 6. Multimodal spatial construction: when the top node was formed, the two nodes representing apple taste and the phonetic sequence [æpl] were active and the level of arousal was sufficiently high. At present, only apple taste is active, giving rise to imagination in the form of the word [æpl].

Definition 21 (Temporal Construction). *Suppose $G = (V, E)$ is a graph and $a, b \in V$. Then $temporal(G, a, b)$ is the graph (V', E'), where $V' = V \cup \{c, d\}$, c is a new node labeled $DELAY$, d is a new node labeled AND, and $E' = E \cup \{(a, c), (c, d), (b, d)\}$, with the weights of the three new edges set to 1.*

Definition 22 (Destruction). *Suppose $G = (V, E)$ is a graph and $a \in V$. Then $forget(G, a)$ is the graph (V', E'), where $V' = V - V''$, $V'' = \{b \in V : E^*(a, b)\}$ and $E' = E - \{(b, c) \in E : b \in V'' \text{ or } c \in V''\}$.*

Definition 23 (Admissibility). *Let $G_0, G_1, \ldots$ be a sequence of networks and $\sigma_0, \sigma_1, \ldots$ a sequence of stimuli. $Spatial(G_n, a, b)$ is admissible at n if both a and b are top active in G_n at n. $Temporal(G_n, a, b)$ is admissible at n if a is top active in G_{n-1} at $n - 1$ and b is top active in G_n at n.*

With the terminology in place, we are ready to define the network development algorithm: see Algorithm 1, where *flip*(p) is the result of flipping a weighted coin that produces outcome 1 with probability p.

Figures 4 and 6 offer examples of network development processes generated by Algorithm 1. Figure 4 shows the formation of a memory of apple taste. Figure 5 shows the formation of a memory of the written word "HI". A memory of the spoken word [æpl], shown in Figure 2 (a), can be formed analogously, but it requires one repetition of the sequence [æpl]. Figure 6, finally, shows how the apple taste and apple word networks are joined.

Algorithm 1. Network development algorithm

```
loop
  if $card(V_n) < p_0$ and $flip(LR(n)) = 1$ then
      if there are preferred $a, b$ s.t. $spatial(G_n, a, b)$ is admissible at $n$
      and $mod(a, G_n) = mod(b, G_n) = \{i\}$, for some $i$ then
        Let $G_{n+1} = spatial(G_n, a, b)$.
      else if there are preferred $a, b$ s.t. $temporal(G_n, a, b)$ is admissible at $n$
      and $mod(a, G_n) = mod(b, G_n) = \{i\}$, for some $i$ then
        Let $G_{n+1} = temporal(G_n, a, b)$.
      else if $arousal(G_n) > p_4$
        if there are preferred $a, b$ s.t. $spatial(G_n, a, b)$ is admissible at $n$
          Let $G_{n+1} = spatial(G_n, a, b)$.
        else if there are preferred $a, b$ s.t. $temporal(G_n, a, b)$ is admissible at $n$
          Let $G_{n+1} = temporal(G_n, a, b)$.
        end if
      end if
  else if $via(a, n) < p_3$ for some $a \in V_n$ then
    Let $G_{n+1} = forget(G_n, a)$, where $via(a, n)$ is minimal.
  end if
  Compute the edge weights $w(a, b, n + 1)$ reflecting $Pr(b|a)$.
  Compute the learning rate $LR(n + 1)$.
  Compute the viabilities $via(a, n + 1)$.
end loop
```

5 Results

Algorithm 1 was implemented in Python 2.7 using the graphic package Graphviz for visualization. All of the development processes described in this paper were obtained using this program and straightforward input streams.

Figures 1–6 illustrate how networks are formed by the algorithm. In this case the algorithm develops exactly the desired memory structures with no undesirable structures as side effects. The algorithm gravitates toward memories that are emotionally intense, frequently repeated, or both.

6 Conclusion

Our study indicates that artificial emotions are well suited for guiding the development of dynamic networks by regulating the quality and quantity of memories formed and removed. The presented network model and network development mechanism are relatively simple and were mainly devised for presenting the idea of emotional concept development. Both can clearly be improved and elaborated in several directions. We conclude that artificial emotions can be fruitful, not only for guiding behavior, but also for controlling concept development.

Acknowledgments. This research was supported by The Swedish Research Council, grants 2012-1000 and 2013-4873. We would like to thank Christos Dimitrakakis for many helpful suggestions.

References

1. Abbeel, P., Coates, A., Quigley, M., Ng, A.Y.: An application of reinforcement learning to aerobatic helicopter flight. Advances in Neural Information Processing Systems **19**, 1 (2007)
2. Bach, J.: A framework for emergent emotions, based on motivation and cognitive modulators. International Journal of Synthetic Emotions (IJSE) **3**(1), 43–63 (2012)
3. Bach, J.: MicroPsi 2: The Next Generation of the MicroPsi Framework. In: Bach, J., Goertzel, B., Iklé, M. (eds.) AGI 2012. LNCS, vol. 7716, pp. 11–20. Springer, Heidelberg (2012)
4. Bechara, A., Damasio, H., Damasio, A.R.: Role of the amygdala in decision-making. Annals of the New York Academy of Sciences **985**(1), 356–369 (2003)
5. Begleiter, R., El-Yaniv, R., Yona, G.: On prediction using variable order markov models. Journal of Artificial Intelligence Research, 385–421 (2004)
6. Bengio, Y.: Learning deep architectures for ai. Foundations and trends in Machine Learning **2**(1), 1–127 (2009)
7. Blum, A.L., Langley, P.: Selection of relevant features and examples in machine learning. Artificial Intelligence **97**(1), 245–271 (1997)
8. Colton, S., Bundy, A., Walsh, T.: Automatic concept formation in pure mathematics (1999)
9. Gould, J.L., Gould, C.G., et al.: The honey bee. Scientific American Library (1988)
10. Johansen, J.P., Diaz-Mataix, L., Hamanaka, H., Ozawa, T., Ycu, E., Koivumaa, J., Kumar, A., Hou, M., Deisseroth, K., Boyden, E.S., et al.: Hebbian and neuromodulatory mechanisms interact to trigger associative memory formation. Proceedings of the National Academy of Sciences **111**(51), E5584–E5592 (2014)
11. Lebovitz, M.: Experiments with incremental concept formation. Machine Learning **2**, 103–138 (1987)
12. LeDoux, J.: Emotion circuits in the brain (2003)
13. LeDoux, J.E.: Emotional memory systems in the brain. Behavioural Brain Research **58**(1), 69–79 (1993)
14. Marsella, S., Gratch, J., Petta, P.: Computational models of emotion. A Blueprint for Affective Computing-A sourcebook and Manual, 21–46 (2010)
15. McCallum, R.A.: Instance-based utile distinctions for reinforcement learning with hidden state. In: ICML, pp. 387–395 (1995)
16. Pickett, M., Oates, T.: The Cruncher: Automatic Concept Formation Using Minimum Description Length. In: Zucker, J.-D., Saitta, L. (eds.) SARA 2005. LNCS (LNAI), vol. 3607, pp. 282–289. Springer, Heidelberg (2005)
17. Richter-Levin, G., Akirav, I.: Emotional tagging of memory formationØl' the search for neural mechanisms. Brain Research Reviews **43**(3), 247–256 (2003)
18. Rosenbloom, P.S.: The sigma cognitive architecture and system. AISB Quarterly **136**, 4–13 (2013)
19. Schmidt, M., Niculescu-Mizil, A., Murphy, K., et al.: Learning graphical model structure using l1-regularization paths. In: AAAI. vol. 7, pp. 1278–1283 (2007)

20. Strannegård, C., von Haugwitz, R., Wessberg, J., Balkenius, C.: A Cognitive Architecture Based on Dual Process Theory. In: Kühnberger, K.-U., Rudolph, S., Wang, P. (eds.) AGI 2013. LNCS, vol. 7999, pp. 140–149. Springer, Heidelberg (2013)
21. Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D.: How to grow a mind: Statistics, structure, and abstraction. Science **331**(6022), 1279–1285 (2011)
22. Tziortziotis, N., Dimitrakakis, C., Blekas, K.: Cover tree bayesian reinforcement learning. The Journal of Machine Learning Research **15**(1), 2313–2335 (2014)
23. Veness, J., Ng, K.S., Hutter, M., Uther, W., Silver, D.: A monte-carlo aixi approximation. Journal of Artificial Intelligence Research **40**(1), 95–142 (2011)
24. Witthöft, W.: Absolute anzahl und verteilung der zellen im him der honigbiene. Zeitschrift für Morphologie der Tiere **61**(1), 160–184 (1967)