Take another little piece of my heart: a note on bridging cognition and emotions

Giuseppe Boccignone

Science urges philosophy to be more empirical and philosophy urges science to be more reflective. This markedly occurred along the "discovery of the artificial" (CORDESCHI 2002): in the early days of Cybernetics and Artificial Intelligence (AI) researchers aimed at making machines more cognizant while setting up a framework to better understand human intelligence.

By and large, those genuine goals still hold today, whereas AI has become more concerned with specific aspects of intelligence, such as (machine) learning, reasoning, vision, and action. As a matter of fact, the field suffers from a chasm between two formerly integrated aspects. One is the engineering endeavour involving the development of tools, e.g., autonomous systems for driving cars as well as software for semantic information retrieval. The other is the philosophical debate that tries to answer questions concerning the nature of intelligence. Bridging these two levels can indeed be crucial in developing a deeper understanding of minds.

An opportunity might be offered by the cogent theme of emotions. Traditionally, computer science, psychological and philosophical research have been compelled to investigate mental processes that do not involve mood, emotions and feelings, in spite of Simon's early caveat (SIMON 1967) that a general theory of cognition must incorporate the influences of emotion.

Given recent neurobiological findings and technological advances, the time is ripe to seriously weigh this promising, albeit controversial, opportunity.

In the heart of cognition

Affective neuroscience (DALGLEISH *et al.* 2009, pp. 355-368) is helping us understand the neural circuitry that underlies emotional experience. It integrates functional neuroimaging, behavioural experiments, electrophysiological recordings, animal and human lesion studies and behavioural experiments striving to understand emotion at the neurobiological and psychological levels. Conflicting explanations at the psychological level, e.g. basic emotions vs. appraisal theories, find a novel synthesis at the neurobiological level. One outstanding example is Damasio's work (DAMASIO 1994, 1999).

Taking stock of such results, affective computing (PICARD 2000) is dealing with artificial agents that aim at instantiating the ability to 1) recognize emotion, 2) express emotion, 3) "have emotions", the latter being the hardest task. So far, most current research focuses on 1) and 2), whereby machine learning based affect detection plays a prominent role (CALVO, D'MELLO 2010). In order to give a thorough discussion of these aspects, we make clear the modelling strategy we adopt from now on.

Given a system (human observer) and its behaviour (e.g. facial expression, gaze shifts), together with available knowledge (psychological/neurobiological theories and descriptions, experiments and measurements), we set up a computational theory (MARR 1982) formalized in terms of Bayesian theory (BOCCIGNONE, CORDESCHI 2007). More precisely, we draw on the Bayesian framework of Probabilistic Graphical Models (PGM) (BISHOP 2006). In brief, by exploiting knowledge and constraints available both at the psychological and at the neurobiological description levels: 1) we identify the essential random variables (RVs) that ground the probabilistic model; 2) we encode the statistical dependencies between RVs in the PGM structure.

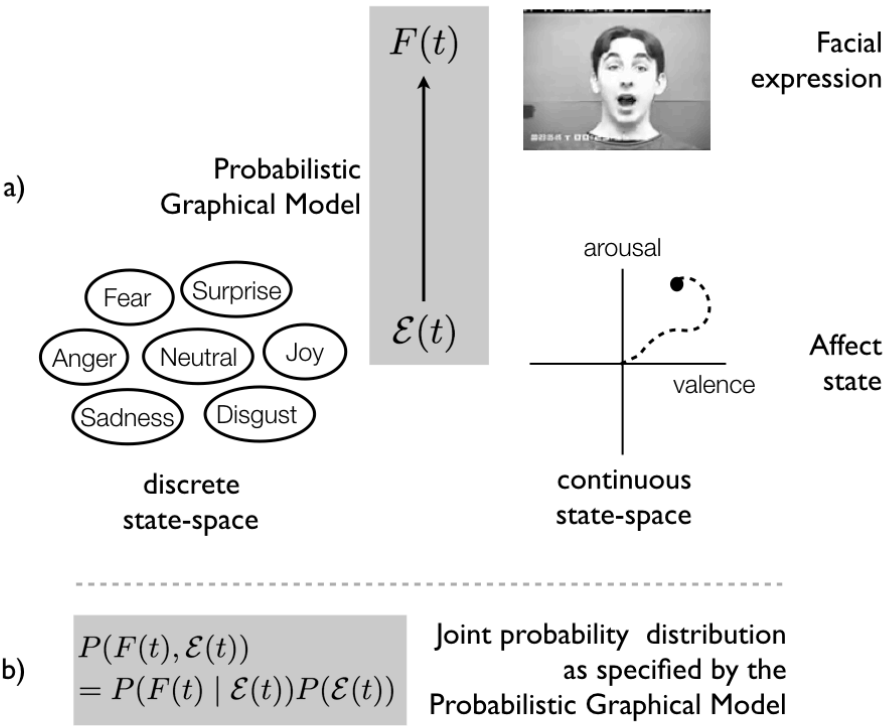

Fig. 1 - Modelling affective facial expression: a) the PGM models the dependency of the observed expression under the current affect state (discrete or continuous); b) the model represents a possible factorization of the joint probability $P(\boldsymbol{F}(t), E(t))$.

Coming back to the issue of dealing with the problem of affective expression generation/detection, one modelling example is presented in Fig. 1. The time-varying RVs, $E(t)$ and $\boldsymbol{F}(t)$ stand for the latent affective state of an agent and the facial expression induced by such state at time t, respectively; such RV's are represented by graph nodes. The structural dependency (arrow) $E(t) \rightarrow \boldsymbol{F}(t)$ captures the statistical dependency of $\boldsymbol{F}(t)$ on $E(t)$, quantified via the conditional probability $P(\boldsymbol{F}(t) \mid E(t))$. The model is *generative*: $P(\boldsymbol{F}(t) \mid E(t))$ specifies the likelihood of generating an expression (by sampling) under a given affective state. The recognition problem (inferring the most plausible affective state given an observed facial expression), boils down to "inverting the arrows" by computing the posterior probability $P(E(t) \mid \boldsymbol{F}(t))$ via Bayes' rule. If $E(t)$ spans a discrete affective state-space, the model can account for a large number of computational models based on discrete theories *á la* Ekman (EKMAN 1993, p. 384); if the space is continuous (e.g., specified via valence/arousal dimensions), it is suitable to cope with Russell's core affect theory (RUSSELL 2003, p. 145).

Yet, the actual challenge is designing artificial agents that "have emotion" and use it for making decisions. Indeed, neurological studies indicate that decision-making without emotion can be impaired. Damasio's findings point to such an essential role (DAMASIO 1994).

It goes without saying, modelling emotion at the most general level is a mind-blowing endeavour for current research. Thus, we will focus on the integration of emotion with cognitive behaviour (PESSOA 2008) by drawing on the minimalist case of active sensing.

Where to look next?

Among the variety of active sensing behaviours, oculomotor behaviour (saccades, pursuit, fixational movements) is the least energy process. Though minimal, from a theoretical standpoint, a gaze shift action can be considered as the result of a decision-making process (conscious or unconscious) (YANG *et al.* 2016).

Such process can be modelled through the perception-action loop as a dynamic PGM (unfolded in time, top of Fig. 2): $I(t)$ denotes the stimulus, e.g. a time varying scene, and $\boldsymbol{r}_F(t)$ is the point of gaze (center of the fovea) at time t; $A(t)$ is the ensemble of RVs defining the oculomotor action setting (e.g., maintain current fixation or saccade in a certain direction); $W(t)$ stands for the ensemble of RVs (e.g., features, objects) characterising the scene as actively perceived by gazing the stimulus at point $\boldsymbol{r}_F$; G summarizes the given goal (e.g, search for a kid). The action setting dynamics $A(t) \rightarrow A(t+1)$ and the scene perception dynamics $W(t) \rightarrow W(t+1)$ are intertwined with one another through the gaze shift $\boldsymbol{r}_F(t) \rightarrow \boldsymbol{r}_F(t+1)$. The actual shift is recovered as the statistical decision of selecting a particular gaze location with probability $P(\boldsymbol{r}_F(t+1) \mid A(t), W(t), \boldsymbol{r}_F(t))$, so to maximize the expected payoff under G, the current goal.[1]

Perceptual decision-making calls for the notions of value and reward (YANG *et al.* 2016) that, in turn, pave the way to bringing emotions into the loop (BOCCIGNONE 2016). At the neurobiological level, it has been made clear that, crucially, cognitive (perceptual) and emotional contributions cannot be separated (PESSOA 2008), as outlined in Fig. 2 (centre).

Indeed, there is a large body of evidence that responses from visual cortex reflecting stimulus significance are the result of simultaneous top-down modulation from fronto-parietal attentional regions and emotional modulation from the amygdala and the

[1] For different instantiations of the model, see CLAVELLI *et al.* 2014, and NAPOLETANO *et al.* 2015.

posterior orbitofrontal cortex (OFC). The affective value attributed to a stimulus – either consciously or unconsciously - drives attention and enhances the processing of emotionally modulated information (much like the physical salience of the stimulus), while exogenously driven attention influences the outcome of affectively significant stimuli (PESSOA 2008). At the same time, the cognitive control system (lateral prefrontal cortex, LPFC, anterior cingulate cortex, ACC) guides behaviour while handling goal-related information; action strategies incorporate value through the mediation of the nucleus accumbens, the amygdala, and the OFC. Basal forebrain cholinergic neurons provide regulation of arousal and attention while dopamine neurons located in the ventral tegmental area (vTA) modulate the prediction and expectation of future rewards (PESSOA 2008).

It is worth noting that neurobiological evidence is relevant for modelling purposes, if we surmise a certain degree of association between the neurobiological and the behavioral levels. We will further comment on this point in the final section of this note, but briefly, we assume that processes that support behavior are implemented by the interaction of multiple areas, (networks), which are dynamically recruited into multi-region assemblies (no "necessary and sufficient" brain regions). In this perspective, we draw on Damasio's cleavage between emotions and feelings, which are the first person experience of the corresponding emotion (DAMASIO 1994). An emotion is a neural reaction to a certain stimulus, realised by a complex ensemble of neural activations in the brain (internal emotional state). The latter often are preparations for (muscular, visceral) actions (facial expressions, heart rate increase, etc.), as a consequence the body will be modified into an "observable" emotional body state. Thus, we introduce the RV ***F(t)*** standing for visceral responses (e.g., heart rate, dermal response) that can be gauged via physiological measurement (ECG, skin conductance, etc.). Note in Fig. 2 (centre panel) the central role of the amygdala and the OFC. Their tight interaction provides a suitable ground (SALZMAN, FUSI 2010, p. 173) for representing, at the psychological level, the core affect dimensions (RUSSELL 2003). Core affect can then be functionally modelled as a latent space (VITALE *et al.* 2014) - see Fig. 1 - spanned by *E(t)*. In addition, *C(t)* indexes a higher cognitive level of interest. As a result, the original PGM is modified in the PGM shown at the bottom of the same figure.

Due to limitation of space, we are not entering details about software / hardware implementations. As to algorithms, a viable solution to provide a simulation of the model is that of exploiting the huge number of state-of-the-art machine learning algorithms (BISHOP 2006). Learning and inference on PGMs can then be accomplished either through approximate optimization-based techniques (e.g., Variational Bayes) or stochastic techniques (Monte Carlo). Eventually, notice that, in the last decade, the number of public repositories has grown larger, where behavioral data gathered in realistic, natural setting experiments have been recorded by multiple modalities (CALVO, D'MELLO 2010). Such data can be readily employed for model learning and validation.

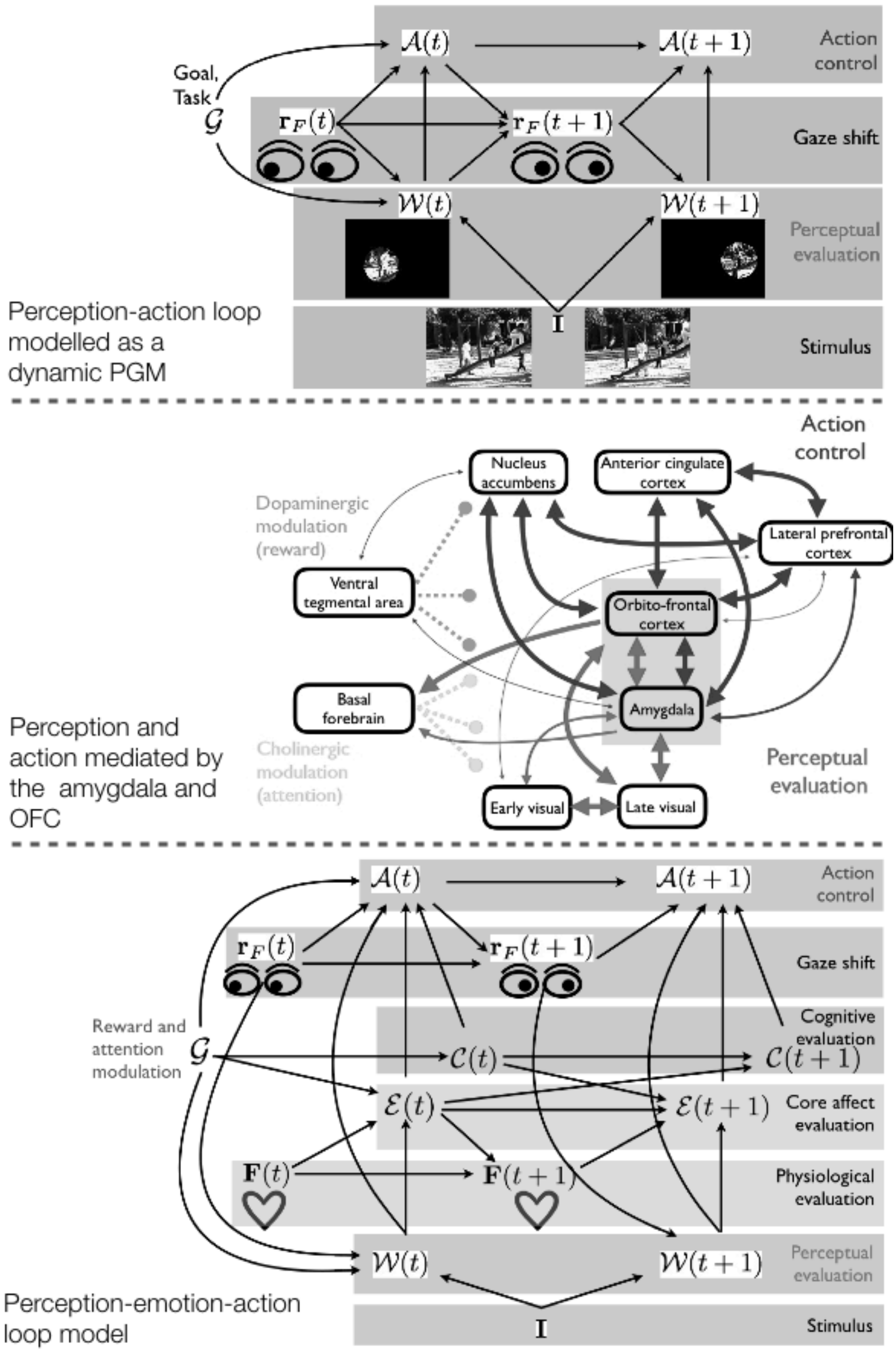

Fig. 2 - *Top*: the dynamic PGM of an active sensing loop. *Centre*: structural/functional constraints implied by circuits for visual processing and executive control. *Bottom*: the perception-emotion-action PGM.

Caveats on methodology

Though minimal, would the implemented version of the model meet the requirement of making decisions by virtue of "having emotions"? Such question entails a number of hindrances.

First, we have assumed, based on Damasio's distinction between emotions and feelings, to rule out the latter. Under such disentanglement, emotions are likely to be amenable to third person description (and thus modelled), whilst feelings would necessarily involve first person experience (opening to the conundrum of consciousness) (TRAUTTEUR 2016).

Second, we have set up a computational theory (MARR 1982) in the Bayesian framework of Probabilistic Graphical Models, where the RVs capturing essential behavioral properties are shaped in the PGM structure by using structural constraints suggested at the neurobiological level. Once implemented, the model is in principle suitable to simulate attentive behavior conditioned by emotion. How things stand, putting the simulation of the model into work[2] is nothing but an instance of the *synthetic method* (CORDESCHI 2002), i.e. the building of artefacts as explanatory models of living organisms. The synthetic method, *per se*, entails a variety of problems (CORDESCHI 2008).

In particular, the "underdetermination" problem involves the choice of "the right grain of analysis for models". To handle the computational explanation at different grains/levels (figure 3), we have adopted a revised form of Marr's framework (MARR 1982). In a Bayesian formalism (BOCCIGNONE, CORDESCHI 2007), Marr's three-fold hierarchy can be re-organized into two levels (KNILL *et al.* 1996): the *computational theory level*, which can be formalized precisely in terms of Bayesian theory, and the *implementation theory level* embedding both algorithmic and realization levels.

[2] For a nice discussion of simulation in cognition, see SANTUCCI *et al.* 2016.

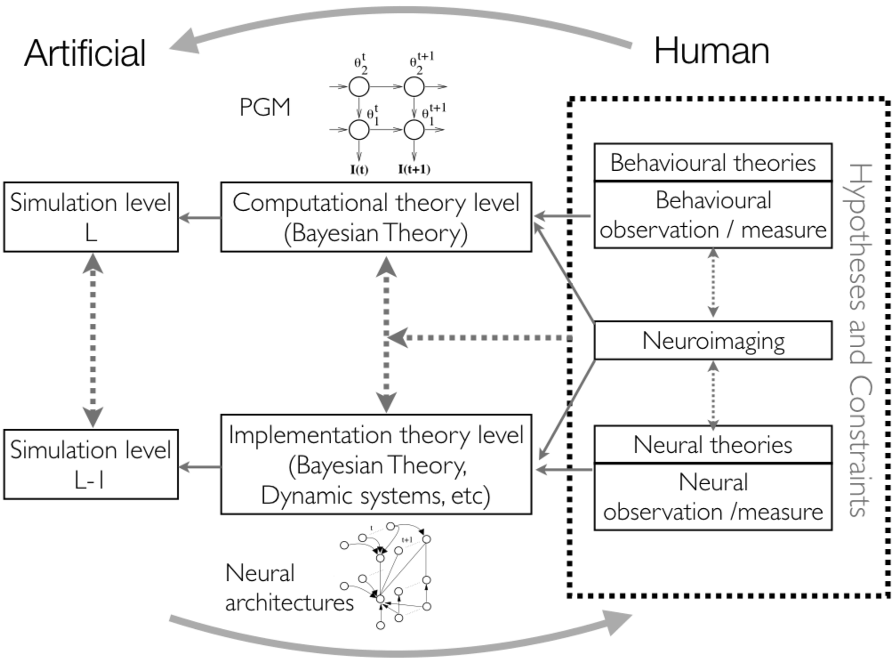

Fig. 3 - The synthetic method loop embedding different levels of explanation.

Differently from Marr (BOCCIGNONE, CORDESCHI 2007), the notion of architecture becomes crucial: the PGM embodies constraints assumed by the scientist for his own purpose at the chosen level of explanation. The algorithmic level does not so far provide an autonomous level, rather one encompassing simulations of different grains (a coarse-grained simulation of Bayesian inference at the behavioural level, a fine-grained simulation at the neural level, see Fig. 3).

As to the realization level, it is a current practice to choose some formal neuronal model to the end (e.g., integrate-and-fire neurons, stochastic differential equations. simple binary, on/off models). This justifies the term "implementation theory": the realization level is but another kind of theoretical model (ABBOTT, KEPLER 1990). Yet, further levels of reduction, and further theoretical models too, could be achieved (*ibidem*) going down in the hierarchy.

In summary, the Bayesian approach provides a sound formalization of Marr's functionalist intuition of a computational theory level. However, a deceptively simple question arises (SPREVAK 2016, p. 92): how should we interpret Bayesian models? One option is *instrumentalism*, where Bayesian machinery should be understood as a formal device (*ibidem*) to describe human behavioural patterns concisely and to make predictions. The alternative is *realism*: models pick out real entities and processes in the human brain ("Bayesian brain hypothesis". See KNILL, POUGET 2004).

The Bayesian approach is advocated for handling uncertainty, stemming from lack of knowledge and from randomness. Going down in the explanation hierarchy, basic sources of randomness are classical dynamics unpredictability and quantum processes, which in living systems are likely to take place simultaneously and affect each other.

Further, different levels of organization make things worse: multi-level interactions induce subsequent forms of randomness (BUIATTI, LONGO 2013). If one assumes a realistic stance, these “living matter” effects pose serious challenges to the functionalism captured by the computationalist account (CORDESCHI, FRIXIONE 2007). Indeed, there is severe criticism (LONGO 2009) in the ability of digital computation to fully reproduce (not just mimicking) this dynamics even in simple cases (deterministic unpredictability).

A viable shortcut (CORDESCHI, FRIXIONE 2007) is the “encapsulation” of any critical level dealing with a non-Turing computable function, in an embedded subsystem so to consider only the computable outputs that might be relevant for higher embedding levels. However, even the “encapsulation” practice is not, at least in principle, unquestionable. Since minor changes in one level might be amplified by the exchanges with another level, such approach might rule out underpinning properties at the biological level, crucial for the overall behaviour of the system, especially in the case of emotions. Even discarding the conundrum of feelings, yet emotions use both neural and humoral routes, so that the resulting emotional state involves continuous, analogue changes within the body proper, e.g., viscera, internal milieu, etc.

All the above issues let us surmise that there is much work left for scientists. And even more for philosophers.

References

ABBOTT L.F., KEPLER T.B 1990, *Model Neurons: from Hodgkin-Huxley to Hopfield*, in *Statistical mechanics of neural networks - Lecture Notes in Physics*, 368, ed. by L. Garrido, Springer-Verlag, Berlin-Heidelberg, pp. 5-18.

BISHOP C.M. 2006, *Pattern Recognition and Machine Learning*, Springer, New York.

BOCCIGNONE G. 2016, *A Probabilistic Tour of Visual Attention and Gaze Shift Computational Models*, in *International Workshop “Vision over Vision: Man, Monkey, Machines, and Network Models”*, Osaka (Available at http://arxiv.org/abs/1607.01232).

BOCCIGNONE G., CORDESCHI R. 2007, *Bayesian Models and Simulations in Cognitive Science*, in *Models and Simulations 2*, Tilburg, PhilSci-Archive (Available at http://philsci-archive.pitt.edu/3556/).

BUIATTI M., LONGO G. 2013, *Randomness and Multilevel Interactions in Biology*, in «Theory in Biosciences», 132(3), pp. 139-158.

CALVO R.A., D’MELLO S. 2010, *Affect Detection: An Interdisciplinary Review of Models, Methods, and their Applications*, in «IEEE Transactions on Affective Computing», 1(1), pp. 18-37.

CLAVELLI A., KARATZAS D., LLADOS J., FERRARO M., BOCCIGNONE G. 2014, *Modelling Task-Dependent Eye Guidance to Objects in Pictures*, in «Cognitive Computation», 6(3), pp. 558-584.

CORDESCHI R. 2002, *The Discovery of the Artificial: Behavior, Mind and Machines Before and Beyond Cybernetics*, Kluwer Academic Publishers, Dordrecht.

CORDESCHI R. 2008, *Steps toward the Synthetic Method*, in *The Mechanical Mind in History*, MIT Press, Cambridge, pp. 219-258.

R. CORDESCHI, M. FRIXIONE, *Computationalism Under Attack*, in *Cartographies of the Mind: Philosophy and Psychology in Intersection*, M. Marraffa, M. De Caro, F. Ferretti eds., Springer Netherlands, Dordrecht 2007, pp 37–49.
DALGLEISH T., DUNN B.D., MOBBS D. 2009, *Affective Neuroscience: Past, Present, and Future*, in «Emotion Review», 1(4), pp. 355-368.
DAMASIO A. 1994, *Descartes' Error: Emotion, Reason, and the Human Brain*, Putnam, New York.
DAMASIO A. 1999, *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, Harcourt Brace & Company, New York.
EKMAN P. 1993, *Facial Expression and Emotion*, in «American Psychologist», 48(4), pp. 384-392.
KNILL D.C., POUGET A. 2004, *The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation*, in «Trends in Neurosciences», 27(12), pp. 712-719.
KNILL D.C., KERSTEN D., YUILLE A. 1996, *Introduction: A Bayesian Formulation of Visual Perception*, in *Perception as Bayesian Inference*, ed. by D.C Knill, W. Richards, Cambridge University Press, Cambridge (UK), pp. 1-21.
LONGO G. 2009, *Critique of Computational Reason in the Natural Sciences*, in *Fundamental Concepts in Computer Science*, ed. by E. Gelenbe, J-P. Kahane, Imperial College Press, London, pp. 43-70.
MARR D. 1982, *Vision*, W.H. Freeman, New York.
NAPOLETANO P., BOCCIGNONE G., TISATO F. 2015, *Attentive Monitoring of Multiple Video Streams driven by a Bayesian Foraging Strategy*, in «IEEE Transactions on Image Processing», 24(11), pp. 3266-3281.
PESSOA L. 2008, *On the Relationship between Emotion and Cognition,* in «Nature Reviews Neuroscience» 9(2), pp. 148-158.
PICARD R.W. 2000, *Affective Computing*, MIT Press, Cambridge.
RUSSELL J.A. 2003, *Core Affect and the Psychological Construction of Emotion*, in «Psychological review», 110(1), pp. 145-172.
SALZMAN C.D., FUSI S. 2010, *Emotion, Cognition, and Mental State Representation in Amygdala and Prefrontal Cortex*, in «Annual Review of Neuroscience», 33, pp. 173-202.
SANTUCCI V.G., CILIA N.D., PEZZULO G. 2016, *The Status of the Simulative Method in Cognitive Science: Current Debates and Future Prospects*, in «Paradigmi», 3, pp. 47-66.
SIMON H.A. 1967, *Motivational and Emotional Controls of Cognition*, in «Psychological review», 74(1), pp. 1193-1216.
SPREVAK M. 2016, *Philosophy of the Psychological and Cognitive Sciences*, in *The Oxford Handbook of Philosophy of Science*, ed. by P. Humphreys. Oxford University Press, Oxford.
TRAUTTEUR G. 2016, *Consciousness, Illusory Freedom, Double Feel*, in «Paradigmi», 3, pp. 12-22.
VITALE J., WILLIAMS M-A., JOHNSTON B., BOCCIGNONE G. 2014, *Affective Facial Expression Processing via Simulation: A Probabilistic Model*, in «Biologically Inspired Cognitive Architectures», 10, pp. 30-41.
YANG S.CH., WOLPERT D.M., LENGYEL M. 2016, *Theoretical Perspectives on Active Sensing*, in «Current Opinion in Behavioral Sciences», 11, pp. 100-108.