Transhuman intelligences will be hyperemotional

Claudius Gros[1]

For real-world biological intelligences increasing cognitive capabilities go hand in hand with rising complexity and predominance of emotional states. We argue that this correlation between cognitive and emotional complexity is not a coincidence, but an indispensable feature of both biological and artificial advanced intelligences. Transhuman superintelligences will hence be hyperemotional in the sense that they will possess a hyper-complex emotional control system.

High intelligence comes with an exponentially increasing set of options for actions. How can an agent then make optimal use of its capability? Given a well defined goal, e.g. to play and to win a chess game, the agent may decide to pursue this objective efficiently. On medium to long time scale the agent however faces the problem to select a sequence of life-defining goals.

Traditional artificial intelligence research proposes two strategies for the selection of overarching goals, viz for solving the motivational problem [1]. The first one conjectures that AIs [2], even if of transhuman intelligence, could be programmed to wait patiently for a human operator to provide a task. In the second scenario an AI would perform a search strategy for the optimal course to reach a life-long overarching goal [3]. Life-long optimization seems however unlikely in an environment where chaotic phenomena, such as the weather [4], abound and in which the actions of other intelligences need to be taken into account. Two transhuman intelligences cannot out-compute each other.

Biological intelligences solve the motivational problem, viz goal selection, through a diffusive emotional control system acting indirectly through the mostly non-specific release of neuromodulators like Dopamine, Serotonin, Norepinephrine and Acetylcholine [5]. Neuromodulators do not interfere directly, by definition, with cognitive information processing, i.e. with neural firing, modulating however the processing pathways of the cell, synaptic transmission and learning [6]. The mixture of active neuromodulators defines the working regime of the targeted cognitive systems.

Emotions are sensed only when a concurrent activation of the neuromodulatory system has occurred [7, 8]. The vast majority of neuromodulators flowing through our brain is however neutral in the sense that we do not directly note their effects. Emotions arise only when neuromodulation has a secondary effect on the body, the James-Lange conjecture, with the propio-sensory feedback from muscles and viscera causing in turn the defining introspective awareness of emotional states [9]. The fact that the right anterior insula (rAI), which plays a central role in the neural circuitry mediating emotional propioception, has evolved rapidly in the recent evolutionary history of the human brain [10], indicates that the sensation of emotions, especially that of higher social emotions like guilt, empathy and exclusion, did become progressively more vivid over the course of our phylogenetic development.

[1] Goethe University Frankfurt, Germany, email: gros07–at–itp.uni-frankfurt.de

Apart from a relative small fraction of emotional states connected to basic drives like hunger and survival, emotions do not cause behavior directly, acting instead indirectly through anticipation, reflection and feedback [11]. Comparative evaluation of behavioral options is consequently a core functionality of emotional states, which hence play a central role in goal selection. In this view it is not a surprise that a substantial part of human cultural production, such as novels and movies, deals with people confronted with motivational problems in situations where our emotional weighting mechanisms offer conflicting solutions. Advanced artificial intelligences will need likewise a net of emotional weighting qualia matching the complexity of their behavioral options.

Even animals capable of only limited sets of behaviors dispose of non-trivial neuromodulatory systems selecting active neural circuits among the ones with are anatomically available [12]. On phylogenetic higher levels evolution has developed the James-Lange feedback mechanism, which we experience as emotions, allowing the brain to become indirectly aware of the activation of a certain fraction of its own neuromodulatory control system. The power of passion is therefore, in contrast to unconscious neuromodulation, that neuromodulatory regulation may now be the subject of cognitive deliberations and hence also of cognitive control.

REFERENCES

[1] Claudius Gros. Emotions, diffusive emotional control and the motivational problem for autonomous cognitive systems. In Jordi Vallverdú, editor, *Handbook of Research on Synthetic Emotions and Sociable Robotics*, page 119. IGI Global, 2009.

[2] Michael Anderson and Susan Leigh Anderson. *Machine ethics*. Cambridge University Press, 2011.

[3] Marcus Hutter. Open problems in universal induction & intelligence. *Algorithms*, 2(3):879–906, 2009.

[4] Claudius Gros. Pushing the complexity barrier: diminishing returns in the sciences. *Complex Systems*, 21:183, 2012.

[5] Claudius Gros. Cognition and emotion: perspectives of a closing gap. *Cognitive Computation*, 2(2):78–85, 2010.

[6] Eve Marder. Neuromodulation of neuronal circuits: back to the future. *Neuron*, 76(1):1–11, 2012.

[7] Jean-Marc Fellous. Neuromodulatory basis of emotion. *The neuroscientist*, 5(5):283–294, 1999.

[8] Joseph LeDoux. Rethinking the emotional brain. *Neuron*, 73(4):653–676, 2012.

[9] Olga Pollatos, Klaus Gramann, and Rainer Schandry. Neural systems connecting interoceptive awareness and feelings. *Human brain mapping*, 28(1):9–18, 2007.

[10] AD Craig. Human feelings: why are some more aware than others? *Trends in cognitive sciences*, 8(6):239–241, 2004.

[11] Roy F Baumeister, Kathleen D Vohs, C Nathan DeWall, and Liqing Zhang. How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, 11(2):167–203, 2007.

[12] Cornelia I Bargmann. Beyond the connectome: how neuromodulators shape neural circuits. *Bioessays*, 34(6):458–465, 2012.