Homeostatically Motivated Intelligence for Feeling Machines

Kingson Man[1] and Antonio Damasio[1]

[1] University of Southern California, Los Angeles CA 90089 USA
{kman,damasio}@usc.edu

Abstract. We present a position paper of work-in-progress on homeostatically motivated artificial intelligence.

Keywords: Homeostasis, Feeling, Consciousness, Soft Robots, Multisensory Integration.

1 Introduction

We propose the design and construction of a new class of machines organized according to the principles of life regulation. These machines have physical constructions—bodies—that must be maintained by homeostasis within a narrow range of viability states and thus share some essential traits with all living systems. The fundamental innovation of these machines is the introduction of risk to self. Rather than up-armoring or adding raw processing power to achieve resilience, we begin the design of these robots by, paradoxically, introducing vulnerability. Drawing on recent developments in soft robotics and algorithms for multisensory abstraction, we outline a way forward to more biologically authentic implementations of feeling in machines.

The main motivation for this project is a set of theoretical contributions to the understanding of biological systems endowed with feeling. Damasio has provided a rationale for the emergence of feelings as a means to optimize survival, a process that is closely associated with the biological basis of selfhood (Damasio 2018). Feelings are intrinsically about something: the state of life in the organism, the likelihood of persistence with well-being or, instead, failure, disease, and death. Ultimately, feelings help achieve homeostasis, i.e., survival with an energy surplus capable of ensuring a future. Organisms do not take a neutral or value-free stance towards information processing. What they do, and think, matters to them.

Machines that implement a process resembling homeostasis might acquire a new source of motivation and evaluation of behavior, akin to that of feelings in living organisms. The world affords risks and opportunities, not in relation to an arbitrary reward or loss function, but in relation to the continued existence of the robot itself. Rewards aren't rewarding and losses don't hurt unless they are rooted in the ground of life and death. True agency arises when the machine picks a side in this dichotomy, when it acts with a preference for existence over dissolution. A robot engineered to participate in its own homeostasis might finally become its own locus of concern. This elementary

concern infuses meaning into information processing. A robot operating on intrinsically meaningful representations might seek broadly intelligent solutions to the human relevant tasks set before it, natural and social.

2 Background

Our approach diverges from traditional conceptions of intelligence that emphasize outward-directed perception and abstract problem solving. We regard high-level cognition as an outgrowth of resources that originated to solve the ancient biological problem of homeostasis. A physical body is subject to perennial risk and decay. But to a disembodied algorithm—or to a robot whose physical persistence is guaranteed—nothing can happen that carries any consequences. Sense-data become meaningful when the data are connected to the maintenance and integrity of the sensing agent and to its homeostasis. Sensory processing that is not attached to a vulnerable body "makes no sense".

The presence of a body serving as an aid or scaffold to problem-solving does not suffice to generate meaning – otherwise all physical robots would be trivially eligible. Preliminary steps have been taken to add "emotional circuits" to influence robotic goal selection (Parisi & Petrosino 2010). Others have constructed robots capable of emotional expressions (typically through facial movements) to facilitate human-robot interaction (Brezeal 2003). But these motivation schedules and emotional performances have not been rooted in the machine's own "life-state" and "well-being". The emotions and "values" underlying them were not relevant to the continuance of the system itself. Such systems ultimately lack a viability constraint. All robots of this class appear to be biologically indifferent (see Jonas 1966), and, in reality, affect-less.

Di Paolo (2003) has criticized the program of embodied robotics as still missing an organism-level logic. Previous work has produced agents that can regulate their own susceptibility to environmental cues based on abstract internal variables (Parisi 2004; Doya & Uchibe 2005). In simulation experiments making explicit reference to homeostasis, phototactic robots used neural plasticity to restore adaptive behavior following visual inversion (Di Paolo 2000).

We advocate a transition from embodied artificial intelligence to homeostatically motivated artificial intelligence. Intelligence has been defined as "an agent's ability to achieve goals in a wide range of environments." (Legg & Hutter 2007) This immediately prompts the follow-up question: "Whose goals?" An agent that myopically follows orders, to the extent that it endangers itself and its ability to follow future orders, is no great intellect.

3 Soft Power

The past decade has witnessed the birth of the sub-discipline of soft robotics. This was enabled by new discoveries in the design and construction of soft "tissues" embedded with electronics, sensors, and actuators. These artificial tissues are flexible, stretchable, compressible, and bounce back resiliently. In short, they are naturally compliant with

their environments. Soft materials potentially provide a richer source of information on body and environment than do traditional rigid materials.

Living organisms are composed of tissues that participate in their own self-maintenance. Living tissues sense and signal the state of their life process. In contrast, consider the "life" of a piece of sheet metal bent into a boxy robot. In general, metal is so durable in comparison to its niche – our human environment – that its security and viability are afterthoughts. Metals and hard plastics are ubiquitous in robotics with the consequence that, in most cases, material integrity can be safely ignored. But durability comes at a cost. An invulnerable material has not much to say about its well-being. If we imagine strain gauges embedded throughout a hard surface, they would spend most of their time reporting "no change", until a catastrophic failure occurs and all the sensors cry out for the first and last time.

Soft robots, on the other hand, more readily enter into a graceful and sensitive coupling with the environment (see reviews by Rogers et al 2010; Pfeifer et al 2014; Rus & Tolley 2015). Under stress, they deform without breaking, then enter a gradual decline instead of sudden failure. The same strain gauges embedded in the volume of a soft material can localize forces and signal graded disruptions in body surface continuity, such as caused by punctures and tears. Soft materials have continuously varying morphology, admitting of more points of contact, control, and force dispersion. They densely sample the environment across multiple sensing modalities, returning rich information about evolving interaction. In short, soft matter is more likely to create the kind of relationship that admits of feeling.

4 Computing Crossmodal Associations

The effort to build a robot with a homeostatic self representation would present a complex exercise in machine learning. It could draw upon a neuroarchitectural framework originally proposed in 1989 (Damasio 1989 a,b). According to this framework, sensory inputs coalesce into abstract concepts by being progressively re-mapped in a neural hierarchical fashion, with each higher level registering more complex features. Nodes in each level can also reinstantiate their lower-level constituent features by top-down projections. This convergence-divergence architecture can form representations that bridge across the sensory modalities (Meyer & Damasio 2009; Man et al 2013).

Crucially, crossmodal processing can accommodate the interoceptive modalities. A proposal has been made that the feeling of existence itself, or of conscious presence, may be due to predictive coding of internal sensations (Seth et al 2012). In theory, external objects can be bound together with internal homeostatic parameters. In practice, however, this invariant mapping is still elusive in the literatures on human brains and machine brains.

There is an intriguing correspondence between the biological architecture of sensory convergence and some variants of deep neural networks. We suggest that artificial neural networks are poised to tackle the next great challenge of building correspondences between inner space and outer space, between internal homeostatic data and external sense data.

A machine in charge of its own self-regulation, and constructed of soft and sensitive tissues, will have a wealth of internal data upon which to draw to inform its plans and perceptions. The homeostatic robot will process information with the aid of something akin to feeling. How does the color, taste, and texture, say, of an apple, systematically associate with changes to the ongoing management of life? All of which is to say that the question "How does this make you feel?" might soon be asked of machines.

References

1. Breazeal, C. 2003. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies* 59(1-2): 119-155.
2. Damasio, A.R. 1989. The brain binds entities and events by multiregional activation from convergence zones. *Neural computation* 1(1): 123-132.
3. Damasio, A.R. 1989. Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition* 33(1-2): 25-62.
4. Damasio, A. 2018. *The Strange Order of Things: Life, feeling, and the making of cultures*. Pantheon.
5. Di Paolo, E. A. 2000. Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., & Wilson, S. (Eds.), *From Animals to Animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior*. Cambridge MA: MIT Press.
6. Di Paolo, E., 2003. Organismically-Inspired Robotics: Homeostatic Adaptation and Teleology Beyond the Closed Sensorimotor Loop. *Dynamical systems approach to embodiment and sociality*, 19–42.
7. Doya, K. & Uchibe, E. 2005. The Cyber Rodent Project: Exploration of Adaptive Mechanisms for Self-Preservation and Self-Reproduction. *Adaptive Behavior* 13(2): 149–160.
8. Legg, S. and Hutter, M., 2007. Universal intelligence: A definition of machine intelligence. *Minds and Machines* 17(4):391-444.
9. Man, K.; Kaplan, J.; Damasio, H.; Damasio, A. 2013. Neural convergence and divergence in the mammalian cerebral cortex: from experimental neuroanatomy to functional neuroimaging. *Journal of Comparative Neurology* 521(18): 4097-4111.
10. Meyer, K. & Damasio, A. 2009. Convergence and divergence in a neural architecture for recognition and memory. *Trends in neurosciences* 32(7): 376–82.
11. Parisi, D. 2004. Internal robotics. *Connection science* 16(4): 325-338.
12. Parisi, D.; Petrosino, G. 2010. Robots that have emotions. *Adaptive Behavior* 18(6): 453–469.
13. Pfeifer, R.; Iida, F.; Lungarella, M. 2014. Cognition from the bottom up: On biological inspiration, body morphology, and soft materials. *Trends in Cognitive Sciences*, 18(8): 404–413.
14. Jonas, H. 1966. *The Phenomenon of Life.* Northwestern University Press.
15. Rogers, J.A.; Someya, T.; Huang, Y. 2010. Materials and Mechanics for Stretchable Electronics. *Science* 327(5973): 1603–1607.
16. Rus, D. & Tolley, M.T. 2015. Design, fabrication and control of soft robots. *Nature* 521(7553): 467–475.
17. Seth, A.K.; Suzuki, K.; and Critchley, H.D. 2012. An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology* 2:395.