You, Robot
Empathy in a Hybrid World

Jesús M. Siqueiros[1]

[1]IIMAS, Unidad Yucatán, UNAM
jmario.siqueiros@iimas.unam.mx

Introduction

Since its first inception in the 1960's the idea of a Heideggerian AI & robotics has led an effort towards an embodied, situated and embedded based-approach. Hubert Dreyfus, its principal philosophical champion, argued that a true Heideggerian AI needs to simulate how humans *are-in-the-world* as *Dasein* or *being-there* in the world, which, among other things, means to be in a co-inhabited social world. The premise of this abstract is that the design of AI and robots, including that from a Heideggerian perspective, do not consider that the *being-in-the-world* for an AI or robot is most likely of a shared world with humans.

According to Heidegger, tools and technology, in general, may be described as equipment and the way humans engage with them is in pragmatic terms. In other words, equipment is not so much about *what is* but *what is for*. Our skilful engagement with the world mediated by the equipment makes it transparent to us. When I use a hammer to hit a nail, my attention is on the task of hammering and not on the hammer itself.

This discussion becomes especially relevant when thinking about the *being-there* with *us* of hybrid robots. If hybridity is in the continuum between heteronomy and autonomy, then robots are not mere equipment for human ends either they are some *subjective other*. The question is, what are they and what would be the ground for our hybrid robot-human interactions? I argue that hybridity may bring a new area for empathy resulting from the paradoxical situation of having to deal simultaneously with their transparency as tools and the potential stubbornness of autonomous entities to disclose themselves by themselves.

Kant 2.0

Hybridity can be seen through different lenses: for example, a hybrid system may be defined by the nature of its components – the system may be composed of mechanical, digital and bio. Instead, I see hybridity in terms of the interplay of autonomy and heteronomy.

In the Critique of Judgement, Kant made the distinction between machines and living things. The first ones are defined by an external agent, imposing purpose and law to which the entity must abide: machines are always a mean to somebody's goals. By contrast, living things are self-organising, self-generated and self-ruled, hence, autonomous systems.

Kant's ideas echo with the modern autopoietic, autonomous agents and enactivist perspectives (Di Paolo, 2005; Weber and Varela, 2002). For these approaches, an autonomous system is a self-organising network of processes that emerge as a self-sustained unity and identity with its own normativity which defines its own space of viability. Moreover, these are adaptive systems because they actively control their interactions with the environment to sustain their organisation. Hence, the environment appears as relevant and meaningful as contributing to or threatening the system continuity.

In summary, autonomous systems breakout from the background as self-ruled and self-maintained entities that define their own unity and identity regardless of the presence of an external observer.[1]

Hybridity

Hybridity may include the intervention of an external designer, mechanical, electrical components plus some form of bio components and processes. These two sets of components, may contribute differently to the final end for which the system was designed. But in their interplay there are different possibilities for an instantiation of the system as a point in the spectrum of heteronomy and autonomy.

Hybridity becomes challenging if the system defines at least one domain of robot-environment interaction relevant to itself as the components take over the system's organisation. The domain of relevance for the robot may or may not coincide with the one of the designer or user.

An instance of such a hybrid system might be the xenobots. Xenobots are made of frog skin and heart cells but designed using a genetic algorithm (Kriegman et al., 2020).

[1]Biological Autonomous Systems perspective has discussed for a long time the nature of robots as heteronomous systems, for example (Ziemke, 2016).

Xenobots are intended to execute certain tasks and goals that have been imposed by their designers (e.g. targeted drug delivery), however due to their biological constitution, they have inherited a domain of existence that can maintain as far as they stay alive. If this is so, then there is at least one domain in which in Kant's sense, the xenobots have made their normativity prevail, despite their end goal have been settled by an external agency. Moreover, the fascinating fact is that designers are counting on the capacity of the robot to maintain its domain of relevance for it to execute the programmed task.

Sharing a world

The notion of autonomy used here is not in the sense of autonomous vehicles. Either it is in a more philosophical loaded sense of responsibility attribution and personhood. But as I mentioned already, autonomous systems are not observer-dependent as they can disclose themselves from the background of things.

As for potential hybrid autonomous robots, self-disclosure must happen as well. However, different from other living things that we may or may not be aware of, hybrid robots are inevitably linked to us, embedded in our life-world in virtue of their constrained domain of existence defined by the designer/user. This situation places us in a position in which we are attached to hybrid systems by design but separated from them by their autonomy. Yet, the separation may not be absolute if the emergent domain of relevance (for self-maintenance) overlaps with at least one relevant domain for us. Then, there could emerge a second-order domain shared between two autonomous agents. Such a domain would be a shared world by co-participation that would require both autonomies in interaction to rise.

You, Robot

Breaking out of the background is what autonomous systems do. Such behaviour contrasts with that of ordinary things that become salient only because of our pragmatic engagement with the world. Hybrid robots instead may self-disclose regardless of our presence and regardless of our pragmatic interest in them. Yet, there may be different ways in which hybrid robots could become present. As with xenobots, robots might irrupt into a domain available to us, but our interaction with them will mostly be from a third-person perspective.[2] As a designer, users or scientists, we will describe their behaviour as an external observer. Notice that this might be different from us describing free-living bacteria or an animal in the wild. Describing hybrid systems will be constrained to a space of action previously designed for it to perform a particular task.

The second way of presentation would be that of hybrid systems aimed at social interaction. Social interaction may require a different set of considerations if breaking out from the background coincides with the social interaction domain. For this particular situation, we should at least consider the plausibility of intersubjective interactions, not simulated but enacted (Dumouchel and Damiano, 2017). For example, if we interact with a chatbot we expect a reply, but the bot has no real concern for us to engage back with it. However, mutual expectancy would be produced if a social domain of relevance emerges in the human-robot dyadic connection. (Zahavi, 2019). A hybrid system would be incomplete without the engagement from the side of the human. This means that a hybrid system that self-disclose in a social milieu will demand to be approached as a *you* from the part of the user, as a potential social interactor and not as a mere *it*.

[2]With drug delivery xenobot-like robots we may interact with them at the level of our immune system, for example.

Final comments

Design and autonomy may be the existential tension for our interaction with hybrid robots, escaping the simple dichotomy between machines vs living things. Transparency –as in equipment- is something that such robots will continuously defy. No matter what, as much as we want to keep them in the background, their self-disclosure capacity will break through into our lives. Such a tension will set the ground for our hybrid robot-human interactions, representing an opportunity for bringing out a new form of empathy, a new way of engaging with a new form of existence, an original feel for the *other*. Further work is needed. There is much to learn and inspiration from the work on Artificial Empathy and Developmental Robotics (Asada, 2014). Also, Ii would be important to move from hybrid constitution to hybridity in participatory sense-making. One final consideration is that hybrid robots are different from domesticated animals and plants. These were already there, and we acted on them to make them work for us. Hybrid robots would be autonomous but created domesticated –in principle.

Acknowledgements

PAPIIT, UNAM grant No. IT300220.

References

Asada, M. (2014). How can we design the development of artifcial empathy ? *Proc. of the 9th ACM/IEEE International Conference on Human–Robot Interaction*.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4(4):429–452.

Dumouchel, P. and Damiano, L. (2017). *Living with Robots*. Harvard University Press.

Kriegman, S., Blackiston, D., Levin, M., and Bongard, J. (2020). A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences*, 117(4):1853–1859.

Weber, A. and Varela, F. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the cognitive sciences*, 1(2):97–125.

Zahavi, D. (2019). Second-Person Engagement, Self-Alienation, and Group-Identification. *Topoi*, 38(1):251–260.

Ziemke, T. (2016). The body of knowledge: On the role of the living body in grounding embodied cognition. *BioSystems*, 148:4–11.