A Pilot Investigation into Robotic Self-Awareness

Ali AlQallaf[1] and Gerardo Aragon-Camarasa[1]

***Abstract*— While humans are aware of their body and capabilities, robots are not. To address this, we present here the first step towards an artificial self-aware architecture for dual-arm robots. Our approach is inspired by human self-awareness developmental levels and serves as the underlying building block for a robot to achieve awareness of itself while carrying out tasks in an environment. Our assumption is that a robot has to know itself before interacting with the environment in order to be able to support different robotic tasks. For this, we propose a neural network architecture to enable a robot to know itself by differentiating its limbs from different environments using visual and proprioception sensory inputs. We demonstrate experimentally that a robot can distinguish itself with an accuracy of 89% from four different uncluttered and cluttered environmental settings and under confounding and adversarial input signals.**

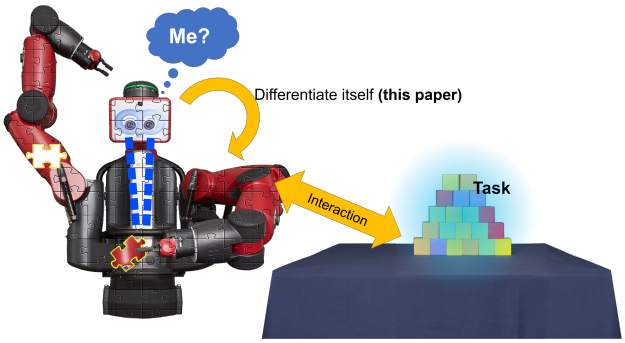

Fig. 1. A robot differentiates, recognises and situates itself first with its body, and then interacts with the environment.

I. INTRODUCTION

While humans are aware of their body and capabilities, robots are not. To address this, we present in this paper the first step towards an artificial self-aware architecture for dual-arm robots. When we become self-aware, we can recognise ourselves in any environment. This is possible because we can distinguish and recognise our body as a separate entity from the world, allowing us to adapt to different situations and scenarios. Robots, however, lack this capability because they are limited to fixed configurations, engineered to work in constrained environments.

Researchers have theorised [1], [2], [3], [4] that an adaptable robot can increase its productivity, and that a self-aware robot can increase its task efficiency over different settings and environments. For this, we propose to ground our approach to robotic self-awareness in Rochat's [5] five levels of self-awareness where each level represents a competence that humans utilise to learn and adapt to its body and then to environments. We, therefore, propose that a robot starts by interacting with itself to construct a *self*, before interacting and dealing with the environment and objects, as shown in Fig. 1. Our approach contrasts to previous and current approaches to construct the self, where the self is built following a top-down approach via the interaction with the environment [4], [6], [7], [8].

In this paper, we investigate the first, basic level of self-awareness which will serve as the building block for enabling a robot to become an adaptable and flexible autonomous machine. For this, we frame the basic level of self-awareness as a binary classification task in which we let the robot to answer whether it can distinguish itself as an entity in an environment with a certain degree of certainty (i.e. certainty is the accuracy of the classification).

[1] Computer Vision and Autonomous group, School of Computing Science, University of Glasgow. Email: a.alqallaf.1@research.gla.ac.uk, gerardo.aragoncamarasa@glasgow.ac.uk

II. RELATED WORK

Rochat [5] has classified self-awareness into five levels, starting from sensing self as a separate entity in the world (Level 1) to self-consciousness (Level 5). Later, Rochat [9] proposed that *self-unity* (Level 0) is the primary phase of newborns which comprises the initial experience of sensory during the first hours of life, and concluded that *self-unity* is equipped in robots in terms of its kinematics, sensors, software, and physical capabilities (e.g. working volume, reach, etc.). The ordering of the five levels of self-awareness is based on their relative complexity and are further divided into implicit (from zero to two) and explicit (from three to five) levels [5], [10]. That is, Legrain et al. [10] have formulated that the implicit self-awareness levels are related to correlating the internal states with the body based on the experience of the self within an environment. The explicit self-awareness levels are those that link the environment to how the environment influences the person. In this paper, we focus on the first self-awareness level and, for completeness, we summarise the implicit levels of self-awareness according to [5] as follows:

- Level Zero – "Self-unity". An individual is born with basic multi-sensory and motor control capabilities which they use to learn about itself.
- Level One – "Differentiation". The individual gets a sense that there is something unique in the experience between what is out there and the felt movements to initiate the sense of self.
- Level Two – "Situation". An individual situates within its body by experiencing the relationship between seen

movements and body stimulation over time.

In robotics, Torras [3] and Chatila et al. [4] have stated that there is a need for robots to be capable of handling different environments while showing high adaptability to any environment. However, Agostini et al. [11] have argued that robots cannot accommodate all human environments, and hard-coding all possible situations is a challenging task. To mitigate this, researchers [12], [13] have proposed to learn an awareness model inspired by the free-energy principle [14] (or a variation of it) in robotics which states that the interactions with the environment are aimed at reducing the internal entropy (i.e. maximising the robot's self-certainty) of an agent. For example, [12], [13] has shown that a robot or its environment might change, and the capability of the robot to adapt to different environments is predicated on the assumption that a robot learns continuously using an active inference model. They thus enabled a robot to adjust its control to the task at hand by minimising the distance between the robot's hand and the target object [13] or where the robot's hand is to its internal belief [12]. However, the authors constrained the robot to have reduced visual perception capabilities in order to simplify the inference task, relying on an observed action within an uncluttered, simple operating environment.

Similarly, Amos et al. [15] have demonstrated that by framing awareness on predictive control models allow a robot to create a link between itself and the environment. Haber et al. [16] has developed an intrinsically motivated agent by using world-model predictions via a supervised learning strategy to model agent awareness in order to generate different behaviours in complex environments. The above robotic agents have learned to deal with the environment while carrying out a task. However, we argue that a robotic system must have the capability to recognise itself before performing actions for a task within an environment (as shown in Fig. 1). Kwiatkowski et al. [2] have shown that a robot can model itself without prior knowledge of its structure, and constructs a self-model that can adapt to mechanical changes that occur to the robot. Their work has demonstrated that self-modelling is the conduit to adaptable and resilient robotic systems. However, the proposed self-model architecture learns about the robot's internal mechanical structure, and it is not able to make a distinction of itself as an entity in the environment without being explicitly defined. The basic robot's existence as an entity reflects the first level of self-awareness, and Kwiatkowski's self-model is not aware of the distinction between itself and the environment.

In this paper, we, therefore, propose that a robot learns how to distinguish itself from the environment before acting on it. For this, we investigate and develop the first level of self-awareness [5] and demonstrate that a robot can experience the *self* by simplifying the learning task to distinguishing itself in different contexts.

III. MATERIALS & METHODS

Our approach to artificial self-awareness focuses on building an initial sense of self in the robot by enabling it to differentiate itself (i.e. Level One in Rochat's self-awareness levels, Section II) from the environment using proprioception and vision. For this, we design a Deep Neural Network architecture to support and understand the self in the robot. The levels of implicit self-awareness (Section II) inspire our architecture design, and we, therefore, propose that these implicit levels can be mapped to robots as follows:

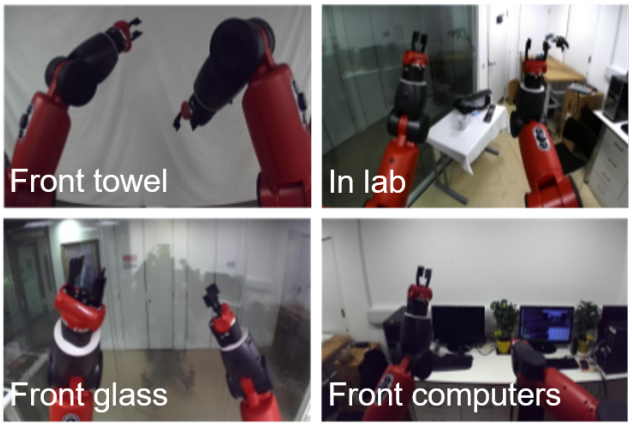

Fig. 2. Sample images from captured scenes, ref. Table I

- Level 0 – "Self-Unity": This level corresponds to the robot's physical, mechanical and sensory capabilities, and its manufacturer's structure configurations, e.g. robot's kinematics, dynamics, sensor definition and configuration, motion planning, etc. These capabilities are interfaced via software APIs and software drivers (e.g. the Robot Operating System, ROS).
- Level 1 – "Differentiation": This level is the initial self, and we propose that this level is about learning how to differentiate itself by seeing its arms and gripers in association with its proprioception without temporal connection between observations. The assumption at this level is that the robot has a description of its limbs via forward and inverse kinematics, and can move its arms via motion planning. The objective is then to confirm if the observed arms and grippers belong to the robot.

The rationale behind our approach is to define a neural network architecture that provides a way to learn the first level of self-awareness and to understand the internal mechanisms of a self-aware robot. The predicted output of the neural network is, therefore, a supervised binary classification task that predicts the sense of self of the robot.

To achieve Level 1, the robot uses its visual sense to discriminate its limbs together with proprioception. For this, we used the robot's vision and proprioception capabilities as the sensory inputs for our approach. Vision comprises RGB images captured using a stereo ZED camera from Stereolabs configured to output images at 720p resolution. Captured images contain a representation of the robot's arms or environment. Proprioception consists of the robot's joint states; being velocity, angular position, and motor torque.

Our architecture for Level 1 of self-awareness consists of

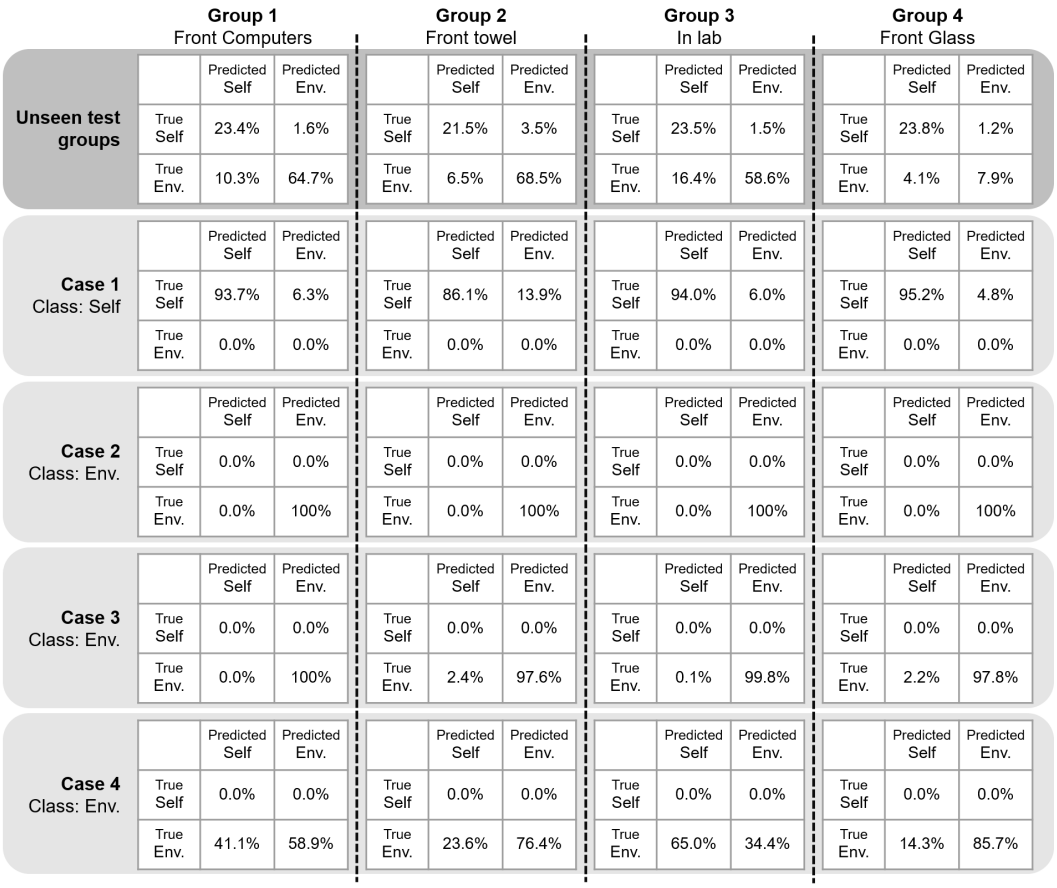

Fig. 3. Unseen test group confusion matrices each with its four confounding cases

TABLE I

EXPERIMENTAL GROUPS AND UNSEEN TEST GROUP DATASETS

	Experimental Groups Sets	**Unseen Test Group**
Group-1	{ In lab, Front glass, Front towel }	Front computers
Group-2	{ Front computers, In lab, Front glass }	Front towel
Group-3	{ Front computers, Front glass, Front towel }	In lab
Group-4	{ Front computers, In lab, Front towel }	Front glass

TABLE II

CONFOUNDING EXPERIMENTAL CASES

	Class	**Description**
Case-1	Self	Vision and propioception correspond to the robot's arms being in the field of view
Case-2	Environment	Vision and propioception correspond to the robot's arms not being in the field of view
Case-3	Environment	The robot's arms are in the field of view but the propioception matches the environment class
Case-4	Environment	Propioception corresponds to the self class but the robot's arms are not in the field of view

a Resnet18 network [17] to process the visual state of the robot. Similarly, for proprioception, we used a single, fully connected network layer to process the internal state of the robot. The output from Resnet18 is a tensor size of 19 that is concatenated with the output of the proprioception tensor of size 76. The concatenated tensor is of size 95 and is passed to a fully connected layer - $FC1$. The output of $FC1$ is a tensor of 32 that goes into the fully connected layer, $FC2$, that predicts *self* or *environment*.

IV. EXPERIMENTS

To understand Level 1, we adopted a leave-one-out cross-validation strategy to test each trained experimental group. By having an unseen experimental group, we are then able to verify the validity of our hypothesis that *Level 1 for artificial self-awareness in the robot increases its self-certainty in an unseen environment*. Accordingly, confusion matrices for each unseen test group in Table I are shown in Fig. 3. The classification accuracy for each unseen test group is: Group-1 is 88.1%, Group-2, 90%, Group-3, 82%, and Group-4, 94.7%. We can, therefore, state that our architecture enables the robot to differentiate itself from the environment with an average certainty of 86.2%.

To further test our hypothesis, we devised an experiment comprising four confounding experimental cases (Table II) that compare the unseen experimental groups against confounding scenarios the robot may encounter. The objective is to confirm that the robot can differentiate itself with a certain

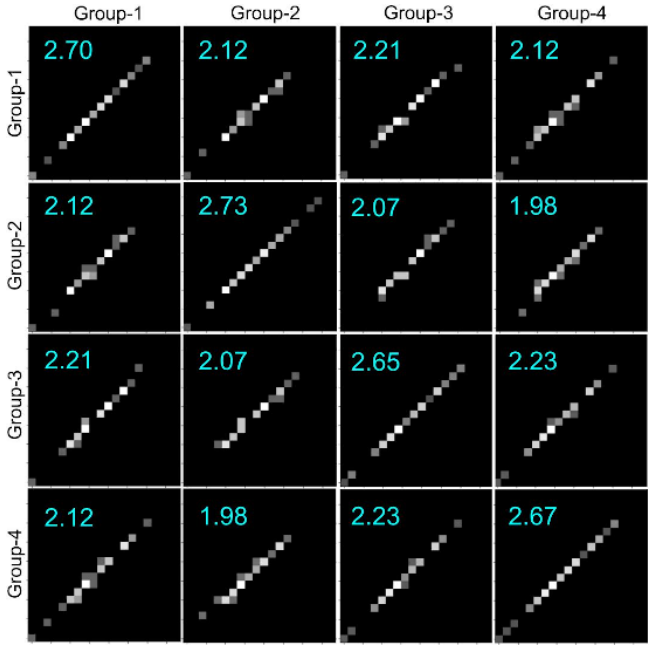

Fig. 4. Mutual information and joint 2D histograms of the trained weights for four Level 1 architectures. The mutual information is noted at the top left corner on each joint histogram plot.

degree of certainty while presented with confounding sensor signals.

To further understand whether our Level 1 architecture learns to differentiate the robot from the environment, we computed the Mutual Information [18] for each group's train dataset (Table I). Our objective is to measure and compare if four Level 1 trained architectures have a degree of similar knowledge that it is invariant to the training set. Mutual information allows us to compare multimodal sources and measure how well two sources are matched by mutual dependence between two variables. That is, different sources of information means more distributed points in the joint histogram and, consequently, low mutual information metric.

The spread in the joint histogram is associated with uncertainty, and in Fig. 4, joint histograms show minor variability in the correlation between the group's models weights. The latter shows that there are no significant differences between the trained models despite the differences in the training datasets, and the misclassification in the confusion matrices results (Table I) are based on the environment noise as other objects within the environment distract the network attention. Since mutual information is computed at the last layer of our architecture, proprioception is taken into account during the classification. Therefore, this demonstrates that our Level 1 network architecture captures a degree of self-awareness and, consequently, certainty. We can, therefore, conclude our experimental hypothesis in Section IV holds for the experiments presented in this paper.

V. CONCLUSIONS

In this paper, we presented an approach to Level 1 of artificial self-awareness in a dual-arm robot. Our approach is inspired by the first level of self-awareness defined by Rochat [5]. By using vision and proprioception, we have demonstrated that a robot can differentiate itself from the environment with an average classification accuracy of 88.7% using unseen test samples and across four different scenes' groups presented in Fig. 3. Future work comprises developing further levels of artificial self-awareness. For level 2, we propose to employ temporal sequences of the robot's arms, and model visual and proprioception experiences.

REFERENCES

[1] J. P. Vasconez, G. A. Kantor, and F. A. A. Cheein, "Human–robot interaction in agriculture: A survey and current challenges," *Biosystems engineering*, vol. 179, pp. 35–48, 2019.

[2] R. Kwiatkowski and H. Lipson, "Task-agnostic self-modeling machines," *Science Robotics*, vol. 4, no. 26, 2019.

[3] C. Torras, "From the turing test to science fiction: The challenges of social robotics," in *Proceedings of the 16th International Conference of the Catalan Association of Artificial Intelligence*, 2013, pp. 5–7.

[4] R. Chatila, E. Renaudo, M. Andries, R.-O. Chavez-Garcia, P. Luce-Vayrac, R. Gottstein, R. Alami, A. Clodic, S. Devin, B. Girard, and M. Khamassi, "Toward self-aware robots," *Frontiers in Robotics and AI*, vol. 5, p. 88, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2018.00088

[5] P. Rochat, "Five levels of self-awareness as they unfold early in life," *Consciousness and cognition*, vol. 12, no. 4, pp. 717–731, 2003.

[6] J. Tani, "An interpretation of the self from the dynamical systems perspective: a constructivist approach," *Journal of Consciousness Studies*, vol. 5, no. 5-6, pp. 516–542, 1998. [Online]. Available: https://www.ingentaconnect.com/content/imp/jcs/1998/00000005/f0020005/880

[7] Y. Nagai, Y. Kawai, and M. Asada, "Emergence of mirror neuron system: Immature vision leads to self-other correspondence," in *2011 IEEE International Conference on Development and Learning (ICDL)*, vol. 2, Aug 2011, pp. 1–6.

[8] P. Lanillos, J. Pages, and G. Cheng, "Robot self/other distinction: active inference meets neural networks learning in a mirror," 04 2020.

[9] P. Rochat, "Self-unity as ground zero of learning and development," *Frontiers in Psychology*, vol. 10, p. 414, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2019.00414

[10] L. Legrain, A. Cleeremans, and A. Destrebecqz, "Distinguishing three levels in explicit self-awareness," *Consciousness and Cognition*, vol. 20, no. 3, pp. 578–585, 2011.

[11] A. G. Agostini, C. Torras, and F. Wörgötter, "Integrating task planning and interactive learning for robots to work in human environments," in *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[12] C. Sancaktar and P. Lanillos, "End-to-end pixel-based deep active inference for body perception and action," *arXiv preprint arXiv:2001.05847*, 2019.

[13] P. Lanillos and G. Cheng, "Active inference with function learning for robot body perception," in *International Workshop on Continual Unsupervised Sensorimotor Learning, IEEE Developmental Learning and Epigenetic Robotics (ICDL-Epirob)*, 2018.

[14] K. Friston, "The free-energy principle: a unified brain theory?" *Nature reviews neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[15] B. Amos, L. Dinh, S. Cabi, T. Rothörl, S. G. Colmenarejo, A. Muldal, T. Erez, Y. Tassa, N. de Freitas, and M. Denil, "Learning awareness models," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[16] N. Haber, D. Mrowca, S. Wang, L. Fei-Fei, and D. L. K. Yamins, "Learning to play with intrinsically-motivated, self-aware agents," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018, p. 8398–8409.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

[18] H. Fang, V. Wang, and M. Yamaguchi, "Dissecting deep learning networks—visualizing mutual information," *Entropy*, vol. 20, no. 11, p. 823, 2018.