Guest Editorial Special Issue on Prediction and Perception in Humans and Robots

I. Scope of This Special Issue

THIS special issue addresses perceptual optimization processes related to attentional and predictive mechanisms. Optimal interaction with the environment requires that agents learn to anticipate and evaluate which sensory information is relevant for a task in a specific context so as to prioritize its processing. It has been suggested that during perception, the selection of sensory information depends on predictive and attentional mechanisms that have modulatory effects, enhancing/facilitating, or attenuating/canceling sensory signals. These mechanisms that modulate the way bottom-up incoming sensory information and top-down predictions based on previous experience are prioritized and integrated during perception, are commonly referred to as perceptual optimization processes.

Scientific production on topics, such as internal models, predictive processing, prediction error minimization, sensory attenuation/cancellation, and sensory enhancement/facilitation, has been increasing in the cognitive science literature during the last decade. Still, there is no consensus on the role of perceptual optimization processes during perception, action, and learning, and on how they influence motor and cognitive development. Furthermore, the influence of the context where the task is being executed is seldom addressed. Several empirical contradictions and theoretical controversies can be found in the cognitive science literature regarding these processes. Moreover, the modeling of perceptual optimization processes in artificial agents and cognitive architectures, especially considering relevant contextual information when performing a task is very limited. In this issue, some of the submissions addressed the potential of applying these processes to robotics.

II. Contributions to the Special Issue

The manuscripts submitted to this special issue address important topics on different optimization processes. As predicting the sensory consequences of one own action is at the core of perceptual optimization processes, Dehban et al. [A1] used deep neural networks to propose a hybrid model to learn visual representations of tasks. The model includes a forward model and a backward model. The forward model goal was to predict the visual consequences of the agent's actions on the final position of a number of objects. This was achieved by predicting the value of important latent variables or features in the image. The backward model was trained considering the most probable previous appearances of the objects, given the current state. After training, the model is capable of learning the features in terms of action possibilities for object recognition, capturing the functional statistics of action and perception, and planning. A remarkable feature of the proposed model is the use of hybrid data for training, exploiting the advantages of both worlds, the image space and the features space.

Perceptual optimization processes are closely related to the ability to perform context-dependent sequential actions. In [A3], the importance of predictive motion optimization is highlighted. The experiments presented here were implemented on a simulated robot arm and using a state-of-the-art deep reinforcement learning algorithm. The results show that, in a sequential reaching task, giving an agent predictive information about the next target identity during action execution greatly improves its performance toward target locations. The findings highlight the potential of utilizing predictive information to optimize the execution of context-dependent sequential actions. By incorporating predictive cues into the decision-making process, agents can enhance their ability to perform sequential reaching tasks more effectively.

Learning actions from task demonstration is central to agents' development of new skills. In this case, perceptual optimization processes are required to select the most relevant task parameters to learn and reproduce actions taking into account contextual changes between learned skills. Xu et al. [A2] implemented a model using deep neural networks for an artificial agent to segment subactions from human demonstration videos and execute the learned actions. The central aspect of the model is a self-attention mechanism that accomplishes action segmentation by extracting contextual features and capturing the spatiotemporal dynamics of the videos, while also generating action commands simultaneously. Moreover, introducing the self-attention mechanism into symbolic task learning enables the prediction of action labels for novel video segments. Experiments on a robotic arm provided empirical evidence of the viability of the model in accurately generating motor commands to reproduce actions based on video demonstrations.

During perception, biological agents face uncertainty by integrating incoming sensory information with inferences based on previous context-dependent experiences. However, it is still an open question if individual differences in previous experiences cause a perceptual misalignment during social

Digital Object Identifier 10.1109/TCDS.2023.3293669

interactions. Mazzola et al. [A4] investigated and modeled the perceptual errors associated with the phenomenon of context-dependency, particularly, in a social scenario. Participants performed a perceptual task in which they have to reproduce the length of a stimulus in three conditions. The findings suggest that participants' perception was more strongly affected by prior knowledge of the individual condition, resulting in an overestimation of shorter stimuli and an underestimation of larger stimuli. In contrast, the influence of prior knowledge was reduced when the line to be reproduced was created by a humanoid robot mimicking a mechanical arm, and even further reduced when the same robot exhibited social behavior. Their results point to the fact that during social interactions the effects of inferential processes in perception are attenuated in favor of higher reliance on incoming sensory information. It is important to note that a perceptual optimization process takes place, wherein individuals modulate the main source of information they rely on to a greater extent in a context-dependent manner.

In the context of language comprehension, perceptual optimization processes can be seen as a product of prediction, which aids in facilitating the reading of predictable words and leads to smaller neural signals linked to processing difficulty, such as the N400 component. The cloze probability of a word, which measures the proportion of people who fill a gap in a sentence with that specific word, is considered a valuable metric for assessing contextual word predictability, predicting processing difficulty, and determining the magnitude of the elicited N400 response. However, since cloze probability is a language production task, its relevance for language comprehension remains uncertain. Therefore, Michaelov et al. [A5] investigated whether cloze probability is the most suitable operationalization of the predictions that underlie the N400. The authors discuss that language models (LMs) predictions are based on the regularities in the statistics of the preceding linguistic context, highlighting that humans' language predictions may rely on other factors. Thus, given that LMs predictions are based on the statistics of language, the authors argue that they can be used to isolate the specific effect of linguistic input on the amplitude of the N400 component. They tested several LMs and found their measure of surprise is a better predictor for N400 than the traditionally used cloze probability. The findings indicate a strong correlation between LMs predictions and the amplitude of the N400 component, pointing to the fact that LMs' predictions closely resemble the predictions made by humans during language comprehension. This suggests that the mechanism responsible for the N400 response in humans is finely attuned to the statistical regularities in language.

Humans have the ability to directly influence perceptual optimization processes. This can be achieved by consciously selecting stimuli that redirect attention, as well as by influencing the emotional and enjoyable aspects of the experience. For example, listening to music while exercising helps to attenuate the unpleasant proprioceptive sensations that may arise from physical effort. However, the impact of music on attention and distraction remains a topic of debate within the scientific literature. Bao et al. [A6] contributed to the research on the topic by investigating the distractive effects of music across different time scales. Their study aims to identify the most effective strategy for assessing these effects through the measurement of autonomic nervous activity. The authors present a comprehensive study monitoring a number of autonomic responses in participants while they exercised. Specifically, participants rode static bicycles with a constant low-intensity workload for consecutive intervals while listening to music (Jazz). An important aspect to emphasize in this study is the association between autonomic information and exteroceptive acoustic information. This association becomes particularly relevant when the acoustic stimuli, in this case, likable music, are considered pleasant by the participants. The results indicate that during intervals where subjects reported experiencing enjoyable music, their perceived exertion levels, as measured by the Rating of Perceived Exertion, were lower. This suggests that the presence of likable music can contribute to attenuating the perception of physical effort during exercise, potentially making the activity more enjoyable and less demanding.

III. Conclusion

The articles presented in this special issue address important challenges for the study of perceptual optimization processes as well as their implementation on artificial agents. A recurring topic is prediction and attention, as well as their importance and influence on learning and adaptive behavior. The contributions to the ongoing research in perceptual optimization processes presented here highlight the relevance of predicting the sensory consequences of one's own actions, the role of spefic perceptual information for motion optimization, the ability to attend relevant task parameters for learning and reproducing actions taking into account contextual changes, the way social interactions modulate inferential processes during perception, the predictions made by humans during language comprehension, and the effects of exteroceptive stimuli to actively influence perception and attention.

Appendix: Related Articles

[A1] A. Dehban, S. Zhang, N. Cauli, L. Jamone, and J. Santos-Victor, "Learning deep features for robotic inference from physical interactions," *IEEE Trans. Cogn. Devel. Syst.*, vol. 15, no. 3, pp. 985–999, Sep. 2023.

[A2] X. Xu, K. Qian, X. Jing, and W. Song, "Learning robot manipulation skills from human demonstration videos using two-stream 2D/3D residual networks with self-attention," *IEEE Trans. Cogn. Devel. Syst.*, vol. 15, no. 3, pp. 1000–1011, Sep. 2023.

[A3] D. Sen, R. de Kleijn, and G. Kachergis, "Behavioral optimization in a robotic serial reaching task using predictive information," *IEEE Trans. Cogn. Devel. Syst.*, vol. 15, no. 3, pp. 1012–1019, Sep. 2023.

[A4] C. Mazzola, F. Rea, and A. Sciutti, "Shared perception is different from individual perception: A new look on context dependency," *IEEE Trans. Cogn. Devel. Syst.*, vol. 15, no. 3, pp. 1020–1032, Sep. 2023.

[A5] J. A. Michaelov, S. Coulson, and B. K. Bergen, "So cloze yet so far: N400 amplitude is better predicted by distributional information than human predictability judgements," *IEEE Trans. Cogn. Devel. Syst.*, vol. 15, no. 3, pp. 1033–1042, Sep. 2023.

[A6] C. Bao, Z. Ma, Z. Wu, W. Gao, G. Jeon, and T. Kiryu, "Evaluation of distraction effect of music stimuli during cycling exercise with low intensity in terms of multiple time scale," *IEEE Trans. Cogn. Devel. Syst.*, vol. 15, no. 3, pp. 1043–1050, Sep. 2023.

ALEJANDRA CIRIA, *Guest Editor*
Universidad Nacional Autónoma de México
Mexico City 04510, Mexico

EMILY S. CROSS, *Guest Editor*
Eidgenössische Technische Hochschule
8092 Zürich, Switzerland

GUIDO SCHILLACI, *Guest Editor*
Independent Researcher
50100 Florence, Italy

BRUNO LARA, *Guest Editor*
Universidad Autónoma del Estado de Morelos
Cuernavaca 62209, Mexico

ANGELO CANGELOSI, *Guest Editor*
The University of Manchester
M13 9PL Manchester, U.K.

Alejandra Ciria received the Ph.D. degree in experimental psychology from the Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico, in 2019.

She worked as a Postdoctoral Fellow with the Adaptive Systems Group, Humboldt-Universität zu Berlin, Berlin, Germany, funded by the Alexander von Humboldt Foundation. Since 2020, she has been an Associate Professor with the Department of Cognitive and Behavioral Sciences, Faculty of Psychology, UNAM, where she is also the Head of the Embodied Cognition Lab. Her research is primarily focused on the study of perceptual, predictive, attentional, and emotional mechanisms within the framework of embodied cognition.

Dr. Ciria is a Diverse Intelligences Summer Institute Fellow.

Guido Schillaci received the Ph.D. degree in sensorimotor learning and simulation of experience as a basis for the development of cognition in robotics from Humboldt-Universität zu Berlin, Berlin, Germany.

He is a Roboticist and an AI Researcher currently working in the energy industry. He has worked for different academic and intergovernmental research institutions, including the Joint Research Center of the European Commission, Ispra, Italy; Scuola Superiore Sant'Anna, Pisa, Italy; the Humboldt-Universität zu Berlin; and the Università di Palermo, Palermo, Italy. He is a former EU-FP7 and H2020 Marie Skłodowska Curie Fellow. His research interest lies primarily in the fields of cognitive and developmental robotics, machine learning, computer vision, optimization, and robotic inspection.

Bruno Lara received the Ph.D. degree in mechatronics from King's College London, London, U.K.

He was a Postdoctoral Fellow with the TheoriLabor, University of Jena, Jena, Germany, working on evolutionary robotics, and then with the Max Planck Institute for Psychology Research, Munich, Germany, focusing on research on cognitive robotics. He is a Professor with the Universidad Autónoma del Estado de Morelos, Cuernavaca, Mexico, where he has been the Head of the Cognitive Robotics Lab, Science Research Centre since 2005. In 2011 and then again in 2019, he did long research stays in the Adaptive Systems Group, Humboldt-Universität zu Berlin, Berlin, Germany. His research interests include internal models, prediction, and predictive processing in the framework of cognitive robotics.

Prof. Lara is an Alexander von Humboldt Fellow.

Emily S. Cross received the Ph.D. degree in cognitive neuroscience from Dartmouth College, Hanover, NH, USA.

Recently, she joined ETH Zürich to establish the Professorship for Social Brain Sciences in Zurich, Switzerland. Prior to this, she was based jointly with the MARCS Institute, Western Sydney University, Sydney, NSW, Australia; the Department of Cognitive Science, Macquarie University, Sydney; and the School of Psychology and Neuroscience, University of Glasgow, Glasgow, U.K. She leads a vibrant and diverse research team that uses brain imaging techniques, robots, and complex action training paradigms to explore how experience-dependent plasticity and expertise is manifest across brain and behavior, with a particular focus on how our experiences and expectations about artificial agents influence our interactions with these agents. Her work has been supported by a range of funders worldwide, including the European Research Council, the Australian Research Council, Research Councils UK, the Dutch Science Foundation, the Alexander von Humboldt Stiftung, the National Institutes of Health, and the Fulbright Commission.

Angelo Cangelosi received the Ph.D. degree in psychology and artificial intelligence from Universitá di Genova, Genoa, Italy.

He is a Professor of Machine Learning and Robotics with the University of Manchester, Manchester, U.K., and the Co-Director and the Founder of the Manchester Centre for Robotics and AI. He holds a European Research Council (ERC) Advanced Grant. He is also a Turing Fellow with The Alan Turing Institute, London, U.K. Overall, he has secured over £38M of research grants as a Coordinator/PI, including the ERC Advanced eTALK, the UKRI TAS Trust Node and CRADLE Prosperity, the U.S. AFRL Project THRIVE++, and numerous Horizon and MSCAs grants. He has produced more than 300 scientific publications. His book *Developmental Robotics: From Babies to Robots* (MIT Press) was published in January 2015, and translated in Chinese and Japanese. His latest book *Cognitive Robotics* (MIT Press), coedited with M. Asada, was recently published in 2022. His research interests are in cognitive and developmental robotics, neural networks, language grounding, human–robot interaction and trust, and robot companions for health and social care.

Dr. Cangelosi is the Editor-in-Chief of *Interaction Studies* and *Cognitive Computation and Systems* (IET), and in 2015, he was the Editor-in-Chief of IEEE TRANSACTIONS ON AUTONOMOUS DEVELOPMENT. He has chaired numerous international conferences, including ICANN2022 Bristol and ICDL2021 Beijing.