Goal-free sensory encoding and learning

A commentary on Molinaro & Collins (2023)

Sophie Smit[1], Genevieve L. Quek[1], Manuel Varlet[1,2], Tijl Grootswagers[1,3]

[1] The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

[2] School of Psychology, Western Sydney University, Australia

[3] School of Computer, Data and Mathematical Sciences, Western Sydney University, Australia

*Correspondence: s.smit@westernsydney.edu.au (Sophie Smit)

Keywords: sensory encoding, representations, goals, learning, time-resolved decoding, visual perception

Investigating how goals impact the way we explore, represent, and interact with the world is vital for understanding human cognition. In their insightful review, Molinaro & Collins (2023) redefine the conventional role of goals in computational theories of learning and decision-making, arguing that in reinforcement learning frameworks, traditionally 'fixed' elements (e.g., states, actions, and rewards) are in fact intricately linked to and influenced by an agent's current goals. In support of their claim that goals are dynamic elements that actively shape information processing altering an agent's state, they draw on fMRI work showing that neural representations in prefrontal cortex vary systematically when participants imagine using the same object to achieve different goals (Castegnetti et al., 2021). These and other findings suggest that goals do not only influence high-level cognitive processes, but can also **modulate the encoding of sensory information**, including representations in early sensory areas (Schaffner et al., 2023). Here we provide nuance to this perspective by highlighting the obligatory and largely automatic nature of early sensory processing, wherein evoked responses to complex stimuli (e.g., faces, objects) encode visual input in a manner that is largely independent of an agent's goal state. This caveat arises out of the time-resolved neural decoding literature that suggests that while the observer's task undoubtedly guides attention and goal states, its influence on the early stages of visual processing is comparatively subtle.

A large body of M/EEG decoding work indicates that prioritisation of task-relevant information typically manifests as a relatively late modulation of visual processing, with (rather modest) effects arising at least 150-200 ms after stimulus onset (Chen et al., 2023; Grootswagers et al., 2021; Moerel et al., 2022; Shatek et al., 2022). For example, when observers view superimposed gratings with distinct orientations, both orientations are represented equally in early EEG responses, even though participants must selectively report the orientation of just one (colour-indicated) grating, while ignoring the other (Moerel et al., 2022). Attentional enhancement of orientation decoding is only evident some 230 ms after stimulus onset, suggesting initial encoding of low-level visual features is robust to the observer's specific goal.

Further along the visual hierarchy, encoding of semantic features of objects also appears consistent under varying goal-states (Shatek et al., 2022). Observers in one EEG study passively viewed images of stationary living entities (e.g., plants) or moving inanimate phenomena (e.g., waves), or else actively categorised them according to either aliveness or capacity for movement. Remarkably, early neural representations of both dimensions, as well as stimulus category, were highly similar across the different tasks, indicating that participants' goals minimally impacted the encoding of high-level stimulus features. Importantly, the observation that early neural representations of both low- and high-level object features appear generally robust to task context does not conflict with fMRI findings that tasks can modulate information in early sensory areas (which receive feedback inputs), but rather underscores the robustness of feedforward visual processing in the first 150 ms (Robinson et al., 2023).

Along a separate line, there is also evidence that agents can passively acquire meaningful sensory representations in the absence of goals, merely via exposure to statistical regularities in the environment. For example, untrained deep neural network models of the human visual hierarchy naturally form units that respond to faces, simply due to random variations in initial parameters, suggesting the potential for cognitive functions like face recognition to emerge organically (Baek et al., 2021). Similarly, unsupervised networks that encode visual similarities between objects, automatically represent attributes such animacy and size, producing sensory representations remarkably similar to those in high-level human vision (Doshi and Konkle, 2023). It is plausible that the brain can similarly employ unsupervised learning to optimise sensory representations. For example, infants actively seek novel information (Kidd and Hayden, 2015) – a wonderful example of goal-directed behaviour guiding visual attention and thus determining which sensory input arrives in visual cortex. In this sense, goals indeed 'shape' the representation of the environment (Molinaro and Collins, 2023). Nevertheless, the subsequent initial *processing* of this incoming sensory information remains largely automatic, and thus indicative of goal-independent learning. Similarly, while repeated exposure or practice can improve sensory processing efficiency, such learning is likely unaffected by immediate goals due to the need for gradual adaptation (Goldstone, 1998).

To summarise, while it is clear that goals shape the way we interpret and learn about the world (Molinaro and Collins, 2023), evidence from both neural decoding and unsupervised learning shows that goals are not a necessary prerequisite for agents to construct meaningful sensory representations from environmental patterns to aid learning. This points to a dual capability of cognitive systems in learning: 1) the flexibility to adaptively seek and prioritise information guided by goals, and 2) the capacity to exploit robust and largely automatic sensory representations. Integrating these dual capabilities into computational models could advance current theoretical frameworks and practical applications in learning and decision-making.

Acknowledgements

This submission was supported by Australian Research Council grants DP220103047 and DE230100380.

References

Baek, S., Song, M., Jang, J., Kim, G., Paik, S.-B., 2021. Face detection in untrained deep neural networks. Nature Communications 12, 7328. https://doi.org/10.1038/s41467-021-27606-9

Castegnetti, G., Zurita, M., De Martino, B., 2021. How usefulness shapes neural representations during goal-directed behavior. Science Advances 7, eabd5363. https://doi.org/10.1126/sciadv.abd5363

Chen, C.-Y., Leys, G., Bracci, S., Op de Beeck, H., 2023. The representational dynamics of the animal appearance bias in human visual cortex are indicative of fast feedforward processing. Imaging Neuroscience 1, 1–26. https://doi.org/10.1162/imag_a_00006

Doshi, F.R., Konkle, T., 2023. Cortical topographic motifs emerge in a self-organized map of object space. Science Advances 9, eade8187. https://doi.org/10.1126/sciadv.ade8187

Goldstone, R.L., 1998. Perceptual Learning. Annual Review of Psychology 49, 585–612. https://doi.org/10.1146/annurev.psych.49.1.585

Grootswagers, T., Robinson, A.K., Shatek, S.M., Carlson, T.A., 2021. The neural dynamics underlying prioritisation of task-relevant information. Neurons, Behavior, Data analysis, and Theory 5. https://doi.org/10.51628/001c.21174

Kidd, C., Hayden, B.Y., 2015. The Psychology and Neuroscience of Curiosity. Neuron 88, 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Moerel, D., Grootswagers, T., Robinson, A.K., Shatek, S.M., Woolgar, A., Carlson, T.A., Rich, A.N., 2022. The time-course of feature-based attention effects dissociated from temporal expectation and target-related processes. Scientific Reports 12, 6968. https://doi.org/10.1038/s41598-022-10687-x

Molinaro, G., Collins, A.G., 2023. A goal-centric outlook on learning. Trends in Cognitive Sciences 27, 1150–1164. https://doi.org/10.1016/j.tics.2023.08.011
Robinson, A.K., Quek, G.L., Carlson, T.A., 2023. Visual Representations: Insights from Neural Decoding. Annual Review of Vision Science 9, 313–335. https://doi.org/10.1146/annurev-vision-100120-025301
Schaffner, J., Bao, S.D., Tobler, P.N., Hare, T.A., Polania, R., 2023. Sensory perception relies on fitness-maximizing codes. Nature Human Behaviour 7, 1135–1151. https://doi.org/10.1038/s41562-023-01584-y
Shatek, S.M., Robinson, A.K., Grootswagers, T., Carlson, T.A., 2022. Capacity for movement is an organisational principle in object representations. NeuroImage 261, 119517. https://doi.org/10.1016/j.neuroimage.2022.119517