



**HAL**  
open science

**Learning Action Hierarchies Comment on "An active inference model of hierarchical action understanding, learning and imitation" by Riccardo Proietti, Giovanni Pezzulo, Alessia Tessari**

Elisabeth Pacherie

► **To cite this version:**

Elisabeth Pacherie. Learning Action Hierarchies Comment on "An active inference model of hierarchical action understanding, learning and imitation" by Riccardo Proietti, Giovanni Pezzulo, Alessia Tessari. *Physics of Life Reviews*, 2023, 46, pp.292-294. 10.1016/j.plrev.2023.08.002 . hal-04285279

**HAL Id: hal-04285279**

**<https://ens.hal.science/hal-04285279>**

Submitted on 14 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Physics of Life Reviews

## Learning Action Hierarchies

--Manuscript Draft--

<b>Manuscript Number:</b>	
<b>Article Type:</b>	Comment
<b>Keywords:</b>	action representation hierarchy; action learning; skill
<b>Corresponding Author:</b>	Elisabeth Pacherie, Ph.D Institute Jean Nicod Paris, FRANCE
<b>First Author:</b>	Elisabeth Pacherie, Ph.D
<b>Order of Authors:</b>	Elisabeth Pacherie, Ph.D

## Learning Action Hierarchies

Comment on “An active inference model of hierarchical action understanding, learning and imitation” by Riccardo Proietti, Giovanni Pezzulo, Alessia Tessari

Elisabeth Pacherie

Institut Jean-Nicod, École normale supérieure, Université PSL, EHESS, CNRS, Paris, France

Orcid ID : [0000-0003-2337-581X](https://orcid.org/0000-0003-2337-581X)

Email : [elisabeth.pacherie@ens.fr](mailto:elisabeth.pacherie@ens.fr)

In their fascinating article, Proietti, Pezzulo and Tessari propose a unified computational account of the cognitive processing underlying action observation, understanding, learning and imitation abilities, conceptualized as hierarchical active inference [1]. The authors justly recognize that neither the idea of a hierarchy of action representations nor the idea that hierarchical action representations and internal models are shared across action execution and perception are exactly novel [2,3,4,5]. Rather, the novelty of their proposal lies in the computational implementation they propose and test in four different simulations and the specific, quantitative predictions their model yields, regarding, for instance, the distinctive patterns of attention deployed when observing familiar as opposed to novel actions or the neuronal dynamics we should observe during action observation and imitation. My focus will be on the hierarchy of action representations that support observation and execution and the way it is learned. I wish to highlight three main points.

First, the authors distinguish four levels in this hierarchy, an observation level where the kinematics of a body part is observed and three levels of hidden variables that are inferred (movement segments, proximal goals and distal goals). Learning this representational hierarchy is for them the process that allows a novel, meaningless action to be associated with semantic meaning by establishing mappings between action representations across hierarchical levels. This, however, should not be taken to mean that learning action representations at a given level requires one to have fully mastered action representations at lower levels of the hierarchy. For example, an indifferent observer won't take more than a few minutes to learn the difference between a forehand shot and a backhand shot and start reliably classifying observed shots accordingly. In contrast, even a dedicated learner will need hours if not weeks of practice before they can reliably execute backhand and forehand shots. This suggests that salience, both in the technical sense in which the authors use the term, where it is determined by prior beliefs about where uncertainty-reducing information is to be found, and in the sense of the physical salience of a stimulus, plays an important role in the early phases of learning. On the one hand, even people not versed in tennis typically know that it is a racket sport that involves hitting a ball across a court and, on the other, collisions and sudden changes of trajectory are physically salient events. Both types of saliences would thus initially orient the observer's attention to the location of impact between racket and ball and the positions of arm and hand at the time of impact, giving them enough information about the core characteristics backhand and forehand to enable them to distinguish between the two strokes. Yet, the functional relations between the different phases of a stroke (e.g., type of grip, footwork, stance, backswing, forward swing, contact and follow through) are at first much less evident and understanding them will demand the right diet of repetitions and variations. All of this is consistent with the active

inference model proposed by the authors, as well as with empirical and philosophical work on skilled action learning [6, 7, 8, 9]. For instance, there is abundant evidence from sport psychology that experts differ from non-experts on sport specific measures of attention allocation and information pick-up and that, e.g., expert tennis and squash players are able to use earlier perceptual cues than non-experts in order to anticipate the actions of their opponents [10, 11, 12]

Second, the authors distinguish between two types of uncertainty reduction processes respectively involved when observing or executing familiar action (driven by prior-determined salience) as opposed to novel actions (driven by novelty-seeking). Yet, actions are rarely if ever entirely novel or entirely identical to past actions. As emphasized by Rafael Nadal, every single shot is different [13]. Even for (relatively) novel actions, pure novelty seeking behavior may not be the most efficient form of learning. Many recent philosophical and psychological theories of skill learning insist that effective improvement of performance is best achieved through deliberate practice where the learner focuses on specific goals and systematically explores and varies means or techniques to improve one's success at these goals [6, 7, 14]. It is through such deliberate and regimented variations, rather than novelty-seeking behavior simpliciter that the dependencies between component elements of task are best learned. In this respect, performers may be better placed to learn than mere observers as they have a form of control on the variations they execute that observers typically lack on the variations they watch. Indeed, large differences have been found in the ability to anticipate events at soccer between players and keen spectators [15].

Finally, while the authors rightfully insist that both action execution and action perception involve a hierarchy of action representation, one should note that hierarchical structure is also present within levels. The movement segments level involves a chunking of the elements accessible at the kinematic level. The empirical evidence regarding motor chunking suggests that it is the result of two opposing, hierarchical processes: a concatenation process through which elements in an action sequence are bound together into larger units, and a segmentation process where multiple contiguous elements are broken down into shorter action sets [16]. While concatenation is best thought of as an associative learning process, parsing appears to be an executive or control process rather than an associative one [17]. Parsing is sensitive to the internal logic of the action sequence and, by carving action at its joints, counteracts rigid automatization and allows for flexibility and manipulability [6]. While internal hierarchical organization may be inchoative at the movement segmentation level, it is a key feature of robust semantic representations at the levels of proximal goals and intentions. For instance, it was found that in experts the semantic representation of a tennis serve was characterized by an organization in a distinctive hierarchical tree-like structure, well matched with the functional and biomechanical demands of the action and with very little variation from player to player. In contrast, in novices, hierarchical organization was almost absent, representations were less well matched with the functional phases of the action, and inter-individual variability was high [18]. Thus, while it may take only a few minutes to acquire a coarse-grained semantic representation of a forehand stroke, sufficient for telling forehand shots from backhand shots, the rich semantic representations needed by players to insure high-level performance, by coaches to give sound advice and by sports commentators for intelligent analyses of play may take years to acquire and require them to learn, as the authors insist, accurate likelihood mappings between all levels of the action hierarchy.

I have little doubt that the model proposed by the authors is powerful enough to accommodate the features of action hierarchy learning I emphasized. Yet, attention to these features suggests that while it may be true that hierarchical action representations once learned support both action production and action production, learning by acting rather than by observation alone allows for the acquisition of more robust representations and mappings between representations at different levels.

## Acknowledgement

This work was supported by grants ANR-17-EURE-0017 and ANR-10-IDEX-0001-02 from the French Agence Nationale de la Recherche.

## References

- [1] R. Proietti, G. Pezzulo and A. Tessari (2023). An active inference model of hierarchical action understanding, learning and imitation, *Physics of Life Reviews*, 46, 92-118, this issue.
- [2] Jeannerod, M., 2006. *Motor cognition: What actions tell the self*, *Motor cognition: What actions tell the self*. New York, NY: Oxford University Press.
- [3] Grafton, S. T., & Hamilton, A. F. D. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, 26(4), 590-616.
- [4] Kilner, J.M., 2011. More than one pathway to action understanding. *Trends in Cognitive Sciences*, 15, 352–357.
- [5] Mylopoulos, M., & Pacherie, E. (2019). Intentions: The dynamic hierarchical model revisited. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(2), e1481.
- [6] Fridland, E. (2019). Longer, smaller, faster, stronger: On skills and intelligence. *Philosophical psychology*, 32(5), 759-783.
- [7] Pacherie, E., & Mylopoulos, M. (2021). Beyond automaticity: The psychological complexity of skill. *Topoi*, 40(3), 649-662.
- [8] Christensen, W., Sutton, J., & McIlwain, D. J. (2016). Cognition in skilled action: Meshed control and the varieties of skill experience. *Mind & Language*, 31(1), 37-66.
- [9] Schack, T., & Frank, C. (2021). Mental representation and the cognitive architecture of skilled action. *Review of Philosophy and Psychology*, 12, 527-546.
- [10] Jones, C., & Miles, J. (1978). Use of advance cues in predicting the flight of a lawn tennis ball. *Journal of Human Movement Studies*, 4, 231-235.
- [11] Abernethy, B. (1990). Expertise, visual search, and information pick-up in squash. *Perception*, 19, 63-77.
- [12] Mann, D. T., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of sport and exercise psychology*, 29(4), 457-478.
- [13] Nadal, R. & Carlin, J. (2011) *Rafa: My Story*. New-York: Hyperion.
- [14] Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. *The Cambridge Handbook of Expertise and Expert Performance*, 38 (685-705), 2-2.
- [15] Williams, M., & Davids, K. (1995). Declarative knowledge in sport: A by-product of experience or a characteristic of expertise?. *Journal of Sport and Exercise Psychology*, 17(3), 259-275.
- [16] Wymbs, N. F, Bassett, D. S, Mucha, P. J, Porter, M. A, & Grafton, S. T. (2012). Differential recruitment of the sensorimotor putamen and frontoparietal cortex during motor chunking in humans. *Neuron*, 74(5), 936-946.
- [17] Verwey, W. B. (2010). Diminished motor skill development in elderly: Indications for limited motor chunk use. *Acta Psychologica*, 134, 206–214.

- [18] Schack, T., & Mechsner, F. (2006). Representation of motor skills in human long-term memory. *Neuroscience Letters*, 391(3), 77-81.

**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: